

Isabel Leiria

Léxico, Aquisição e Ensino

do Português Europeu língua não materna

Este livro, com ligeiras alterações, reproduz a dissertação de doutoramento em Linguística Aplicada que, com o mesmo título, defendi em Outubro de 2001 na Universidade de Lisboa. Foi minha orientadora a Professora Doutora Isabel Hub Faria.

Introdução

- 0.1. *Motivação*
- 0.2. *O estado dos conhecimentos em aquisição do léxico em L2*
- 0.3. *Natureza e objectivos do estudo*
- 0.4. *Enquadramento teórico*
- 0.5. *Organização*

Parte I

Léxico, bilinguismo, aquisição e ensino de L2

Capítulo 1 *Léxico*

- 1.1. *Introdução*
- 1.2. *A abordagem generativa*
- 1.3. *A abordagem funcional*
 - 1.3.1. *A abordagem funcional tipológica*
- 1.4. *A abordagem cognitiva*
- 1.5. *Sumário*

Capítulo 2 *Léxico e bilinguismo*

- 2.1. *Introdução*
- 2.2. *Produção de discurso por bilingues: o léxico como mediador*
 - 2.2.1. *A mensagem pré-verbal*
 - 2.2.2. *Entre a conceptualização e a formulação da mensagem*
 - 2.2.2.1. *O léxico bilingue*
 - 2.2.2.2. *O léxico da L1*
 - 2.2.2.3. *O léxico da L2*
 - 2.2.3. *A formulação da mensagem*
 - 2.2.3.1. *Léxico deficitário e estratégias compensatórias*
 - 2.2.4. *A exteriorização da mensagem*
- 2.3. *Sumário*

Capítulo 3 *Léxico e aquisição de L2*

- 3.1. *Introdução*
- 3.2. *Os primeiros estudos*
- 3.3. *A abordagem linguística*
- 3.4. *A abordagem cognitiva*
 - 3.4.1. *Aprender / saber uma palavra*
- 3.5. *Sumário*

Capítulo 4 *Léxico e ensino de L2*

- 4.1. *Introdução*
- 4.2. *Das listas de palavras às listas de actos de fala*
- 4.3. *Investigação em léxico para o ensino de L2*
 - 4.3.1. *Vocabulários fundamentais*
 - 4.3.2. *Estratégias de inferência e de retenção*
- 4.4. *Ordem de aquisição, léxico e ensino de L2*
 - 4.4.1. *A abordagem lexical*
 - 4.4.2. *A abordagem cognitiva*
- 4.5. *Sumário*

Parte II
Léxico: aquisição e ensino de PE L2

Introdução

Capítulo 5 *Recolha e tratamento de dados de aquisição de PE L2*

- 5.1. *Introdução*
- 5.2. *Informantes e materiais*
- 5.3. *Português Europeu Escrito por Estrangeiros (PESTRA)*
 - 5.3.1. *Transcrição*
 - 5.3.2. *Organização*
 - 5.3.3. *Codificação*
- 5.4. *Para a construção da base de dados*
 - 5.4.1. *Para a avaliação da dimensão do vocabulário*
 - 5.4.2. *Para a construção da tipologia dos desvios*
 - 5.4.3. *Anotação do corpus e marcação dos desvios*
 - 5.4.4. *O programa*
- 5.5. *Dados para a análise*
 - 5.5.1. *Vocabulário disponível*
 - 5.5.2. *Tipologia de desvios*

Capítulo 6 *Vocabulário disponível e léxico deficitário*

- 6.1. *Introdução*
- 6.2. *Vocabulário disponível*
 - 6.2.1. *Vocábulo*
 - 6.2.2. *Combinatórias*
 - 6.2.3. *Conclusão*
- 6.3. *Léxico deficitário*
 - 6.3.1. *Desvios formais em vocabulário disponível*
 - 6.3.1.1. *Ortografia e acentuação*
 - 6.3.1.2. *Formas aproximadas*
 - 6.3.1.3. *Desvios na flexão de nomes, adjectivos e verbos*
 - 6.3.1.4. *Desvios na atribuição de género*
 - 6.3.1.5. *Conclusão*
 - 6.3.2. *Vocabulário indisponível e estratégias compensatórias*
 - 6.3.2.1. *Vocábulo*
 - 6.3.2.2. *Combinatórias*
 - 6.3.2.3. *Conclusão*
 - 6.3.3. *Desvios na selecção do léxico*
 - 6.3.3.1. *Desvios nos nomes*
 - 6.3.3.2. *Desvios nos adjectivos*
 - 6.3.3.3. *Desvios nos verbos*
 - 6.3.3.4. *Desvios em preposições*
 - 6.3.3.5. *Desvios em outras categorias*
 - 6.3.3.6. *Recategorização*
 - 6.3.3.7. *Conclusão*
 - 6.3.4. *O léxico para além do léxico*
- 6.4. *Conclusões*

Capítulo 7 *Léxico e ensino de PE L2*

7.1. *Introdução*

7.2. *Uma abordagem cognitiva centrada no léxico*

7.2.1. *Alguns pares de conceitos a ter em consideração no ensino*

7.2.1.1. *Vocabulário e léxico*

7.2.1.2. *Vocabulário nuclear e vocabulário periférico*

7.2.1.3. *Vocabulário e gramática*

7.2.1.4. *Aprendizagem implícita e aprendizagem explícita*

7.2.1.5. *Estratégias de substituição e estratégias de reconceptualização*

7.2.1.6. *Cognatos e não cognatos*

7.2.2. *Releitura de alguns outros pares de conceitos*

7.2.2.1. *Recepção e produção*

7.2.2.2. *Competência e performance*

7.2.2.3. *Fluência e correcção*

7.2.2.4. *Categorias nocionais e categorias funcionais*

7.2.3. *Releitura de materiais disponíveis para o ensino de PE L2*

7.2.3.1. *O Português Fundamental*

7.2.3.2. *O Nível Limiar*

7.3. *Formação de professores e investigação sobre o PE L2*

7.4. *Conclusão*

Referências Bibliográficas

Índice de Autores

Índice de Assuntos

Anexos

Anexo 1	Quadro de temas e documentos
Anexo 2	Facsimile do documento A35
Anexo 3	Transcrição do documento A35
Anexo 4	Transcrição codificada do documento A35
Anexo 5	Lista de itens diferentes do documento A35
Anexo 6	Lista de ocorrências do sub-corpus alemão (extracto)
Anexo 7	Lista lematizada do sub-corpus alemão (extracto)
Anexo 8	Entradas do <i>Português Fundamental</i> no corpus (extracto)
Anexo 9	Entradas não <i>PF</i> mais frequentes no corpus
Anexo 10	Locuções <i>PF</i> usadas pelo grupo de controle e no <i>PESTRA</i>
Anexo 11	Combinatórias no sub-corpus Alemão (extracto)
Anexo 12	Desvios na acentuação
Anexo 13	Verbos mais frequentes no <i>PESTRA</i>
Anexo 14	Ocorrências verbais no <i>PESTRA</i> correctas e desviantes

Introdução

0.1. *Motivação*

A **Linguística Aplicada** é um vasto e muito variado campo de trabalho de que a Linguística Aplicada ao Ensino das Línguas é uma disciplina. Neste âmbito, ela tem por objectivo desenvolver instrumentos que possam, directa ou indirectamente, apoiar o ensino. E porque é hoje um lugar comum dizer que o ensino das línguas tem de ter em consideração aquilo que se sabe sobre o modo como elas são aprendidas (ainda que, por enquanto, não seja claro o modo como esse conhecimento pode informar a prática lectiva], a linguística aplicada tem de ser também “psicologia aplicada”¹.

Na verdade, pelo menos desde o início dos anos 80, vários investigadores têm vindo a chamar a atenção para os benefícios que o ensino das línguas pode retirar da investigação em aquisição de língua não materna (L2)². Alguns vão até mais longe do que considerar só os benefícios para a formação de professores e propõem que os planos de estudo para o ensino sejam guiados pela informação disponível sobre a ordem de aquisição segundo a qual determinados componentes da língua são aprendidos. Afirmam que só assim se evita a acumulação de conhecimento metalinguístico completamente inútil para o aprendiz e que pode mesmo causar distúrbios ao seu processo de aquisição.

Kenneth Hyltenstam afirmava, em 1985, não acreditar que a investigação em aquisição de L2 seja a chave para resolver todos os problemas do ensino das línguas, mas que, de qualquer modo, **não podemos continuar a trabalhar exclusivamente com base na intuição**³. Diz acreditar, no entanto, que o conhecimento que este campo de estudos põe à nossa disposição é necessário para um bom ensino, embora deva ser

¹ Ver Sharwood-Smith 1988 e Rogers 1996.

² Usarei a sigla “L2” como termo genérico para referir qualquer situação de aprendizagem, de ensino ou de uso de língua não materna; usarei a expressão “língua segunda” (LS) para referir as situações em que ela é língua oficial; e usarei a expressão “língua estrangeira” (LE) para referir a situação de aprendizagem, de ensino ou de uso em ambiente exclusivamente formal, ou seja, em que não se verifica a variável imersão. (Ver Leiria 1999).

³ Os sublinhados são, em todo o texto, e salvo aviso em contrário, da minha responsabilidade.

secundado por conhecimentos de natureza política, social e pedagógica (Hyltenstam 1985:117 e 133). Mas, infelizmente, parece que aquilo que deveria ser matéria de apoio é, de facto, o prato forte, e que o envolvimento entre a investigação em aquisição e o ensino, mesmo no caso de línguas que têm sido alvo de tanta investigação como o Inglês, ainda está muito longe de atingir um nível desejável:

Psycholinguistics [...] has been insufficiently influential on our profession as a foundation discipline, losing out in importance to linguistics and sociolinguistics. [...] I am not arguing that linguistics and sociolinguistics are not relevant for second language learning and use, but rather that their attractions have rather diverted attention from the role of psycholinguistics (Skehan 1998:1-2).

E Margareta Rogers, ao comentar um volume sobre ensino e aprendizagem de vocabulário publicado em 1990, escreve:

*information specifically on L2 vocabulary **learning** is hard to find. So we can conclude that the advice on **teaching** is based on other information, in other words on principles motivated from elsewhere, such as sociolinguistic concepts, e.g. communicative competence; or semantic theory, e.g. semantic fields, componencial analysis, collocational analysis; or on experience, i.e. what has worked in the past. This is not to say that such advice lacks interest – simply that it is not informed by some knowledge of **how L2 vocabulary is learned, stored and retrieved** (Rogers 1996:77) ⁴.*

A escolha do **Léxico** como objecto de estudo é duplamente motivada. Em Leiria 1991, estudei a aquisição dos Pretéritos Perfeito e Imperfeito enquanto expressão da oposição aspectual *acabado/não acabado*, variante da distinção mais genérica Perfectivo/Imperfectivo. Compreendi então a importância que tem para a aquisição a representação que o falante tem do verbo e de cada um destes tempos verbais. A relação entre a base lexical e os tempos verbais no passado, na aquisição do português, e a escolha que fazemos da perspectiva aspectual foi uma das questões que sempre me interessou enquanto professora de português língua não materna. Outra questão que sempre despertou a minha curiosidade relaciona-se com os

⁴ Skehan 1998, por exemplo, tem uma posição semelhante: «I am not arguing that linguistics and sociolinguistics are not relevant to second language learning and use, but rather their attractions have rather diverted attention from the role of psycholinguistics.»

'hábitos' que cada língua adquire/escolhe para 'dizer' o mundo, o que limita a aparente total liberdade que os falantes têm de escolher as palavras e as formas de dizer (cf. Slobin 1990 e Sinclair 1991, entre outros). O desafio, para o falante, não é ser original (essa possibilidade fica reservada aos escritores de alguns, poucos, gêneros literários), mas sim conformar-se o mais possível com o que é esperável numa dada situação. É esta falta de liberdade que leva muitas vezes o professor de língua estrangeira a responder a certos alunos, na 'idade dos porquês', «porque sim!, não sei..., porque é assim...» criando-se até, por vezes, situações muito desconfortáveis entre um aluno pouco habituado a aprender línguas e um professor pouco treinado a ensiná-las. É que, nessas situações, ambos julgam que "a culpa" é do professor. E é também essa mesma falta de liberdade que pode levar o professor a devolver a um dos seus alunos um texto escrito sem marcas de correcção, mas acompanhado de outro, redigido pelo professor, e precedido da seguinte observação: «A sua composição não tem erros, mas eu diria assim: ...»

A investigação que fiz para Leiria 1991 fez-me perceber a relação entre estas duas questões, como ambas estavam dependentes de propriedades do léxico e de outras questões com ele relacionadas, e como são responsáveis, em boa parte, pela zona de desvio que podemos perceber no discurso não nativo, mesmo quando a gramática, *stricto sensu*, da interlíngua já não tem nada que se lhe diga. Ora, é exactamente nesta altura que a crise na sala de aula costuma acontecer: quando ao professor parece não haver já nada para ensinar e ao aluno parece ter tudo para aprender. A estratégia usada costuma ser fazer mais 'exercícios de gramática' e reforçar a exposição a textos variados para 'alargamento' de vocabulário... Mas, o que não se costuma fazer é observar minuciosamente em que contextos linguísticos ocorrem os erros. Em Leiria 1991, pus exactamente essa questão : com que verbos é que o erro é mais frequente? Assim, foi na sequência desse estudo e motivada por questões relacionadas com o ensino que decidi (continuar a) trabalhar sobre o léxico, estendendo a pergunta a outras categorias gramaticais: em que itens lexicais e em que contextos linguísticos o 'erro' é mais frequente, e que factores concorrem para isso?

Por outro lado, e no que respeita a linguística teórica, o seu desenvolvimento tem sido no sentido de representar cada vez mais a gramática como governada pelo léxico. O mesmo tem vindo acontecer no âmbito da Psicolinguística. A tal ponto que em qualquer investigação relacionada com o processamento e com a aquisição e desenvolvimento de L1 o léxico mental pode ser relevante. Deste modo, o léxico passou a ser, e são muitos a fazê-lo notar no âmbito da investigação de L2, suficientemente abrangente para merecer a nossa atenção.⁵

Mas, se aos linguistas teóricos (e aos lexicólogos) pode interessar sobretudo a relação entre a semântica, a sintaxe e a morfologia, os que trabalham em Psicolinguística e em Linguística Aplicada à L2 têm que se confrontar com todos os aspectos do léxico, tanto para investigação como para a aplicação ⁶. Daí a necessidade de recorrer a investigação de várias áreas e conseqüentemente de utilizar bibliografia abundante e variada. E se os estudos de aquisição de L2 são por natureza abrangentes, no estado em que se encontram, em especial no que respeita o Português, ficarão com certeza prejudicados se se limitarem a olhar numa única direcção. Cabe aqui, por isso, uma chamada de atenção. Cada um dos leitores deste trabalho, em função da sua área de especialização, encontrará na bibliografia não só as habituais gralhas tipográficas, mas também a ausência de muitos títulos considerados fundamentais ⁷. Espero, para esse

⁵ “[Q]uestions concerning the second language lexicon are important in at least two ways. First, and perhaps most obviously, knowledge about how a second language lexicon is learned informs us about the nature of second language learning. Second, how learners organize and use their lexicon informs us about the nature of language in general.” (Gass 1987:129).

⁶ A diversidade de tipos de estudos e de abordagens que o léxico pode envolver está patente numa longa lista que Altman (1997:69) inclui num estudo sobre processos de aquisição do léxico: ensino de vocabulário, linhas mestras para o ensino, técnicas relacionadas com a aprendizagem para a compreensão ou para a produção, natureza do léxico, natureza da aprendizagem do léxico, organização do léxico, uso do léxico, quantidade, processamento, avaliação, dificuldade, desenvolvimento do léxico, aquisição e leitura, etc...

⁷ Do mesmo modo que, e pelas mesmas razões, espero compreensão para o facto de alguns dos títulos citados, por vezes, não o serem directamente, mas sim através de leituras de especialistas na área.

facto, a compreensão do leitor e que tenha em mente as seguintes palavras de Singleton:

“[T]he very conception of what is to be included under the heading of lexicon has expanded to the point where almost any L2 acquisition/processing research can be relevant to the mental lexicon.” (Singleton 1999:xiii-xiv).

Como se pode ver, tenho vindo a considerar neste preâmbulo informação bibliográfica exclusivamente de origem inglesa. Assim será na maior parte deste trabalho. De facto, e ao contrário do que acontece com o Inglês, e já vem acontecendo com o Espanhol desde meados dos anos 90, ao muito que se tem dito e escrito sobre a importância do Português entre as línguas do mundo não tem correspondido um interesse suficiente para que a investigação sobre a sua aprendizagem como língua não materna tenha hoje uma expressão que nos permita sequer começar a apoiar nela a formação de professores ⁸.

Este trabalho sobre ***Léxico: aquisição e ensino do Português Europeu língua não materna*** tem como objectivo, digamos, ‘prévio’, contribuir modestamente para que este quadro comece a mudar. Mas, tratando-se de investigação aplicada, é necessário começar por reflectir sobre desenvolvimentos em outras disciplinas que possam ser relevantes para os estudos de L2. É a falta de trabalhos desta natureza entre nós que justifica a extensa Parte I deste trabalho. Com ela pretendo também chamar a atenção para que a investigação em linguística teórica ou aplicada (por exemplo, em Sociolinguística, em estudos de Análise do Discurso ou de *corpora* de falantes nativos) é de valor inestimável para o ensino de L2, mas só e se enquadrada por investigação em outras áreas da Ciência Cognitiva, nomeadamente, em aquisição da linguagem.

⁸ Contrariando este panorama de vazio total, lembro já aqui os trabalhos em curso da equipa da Universidade Eduardo Mondlane de Maputo, e muito em especial os de Perpétua Gonçalves. Falarei deles mais adiante.

0.2. O estado dos conhecimentos em aquisição do léxico em L2

Research on the acquisition of vocabulary has mushroomed enormously over the last twenty years, to the extent that it is now almost impossible to keep up with the output, even if you hardly read anything else (Meara 1995: i-ii).

Quem isto afirma, em 1995, é Paul Meara, o mesmo que, em 1980, chamou a atenção para o facto de, até então, o ensino e a investigação do léxico ser uma área muito negligenciada, numa frase que se tornou célebre e que aparece citada em nove de cada dez textos sobre o assunto ⁹:

Vocabulary acquisition is part of the psychology of second-language learning that has received short shrift from applied linguistics, and has been very largely neglected by recent developments in research (Meara 1980:221).

Desde então, nunca mais deixou de, periodicamente (cf. Meara 1980, 1995, 1996 e 1997 e Schmitt & Meara 1997), fazer balanços sobre o que está feito e o que falta fazer. Ele é, provavelmente, quem melhor dá testemunho do estado da investigação. E, ainda que, de certo modo, me pareça um pouco pessimista, Meara não é o único a dizer que à quantidade não corresponde igual qualidade. Nove anos depois do primeiro texto de Meara, escreve Söderman:

Vocabulary acquisition can no longer be referred to as a neglected aspect of language learning. Yet, we know it is an area of research still much in its infancy (Söderman 1989:114).

Num texto de 1996, Meara faz uma avaliação dos trabalhos realizados entre 1900 e 1960. E comenta, mais uma vez, que muitos dos trabalhos, quer feitos no período observado quer nos anos 80 e 90, não são nem teóricos nem sistemáticos ¹⁰. Muitos são trabalhos experimentais com um número muito reduzido de palavras a serem aprendidas, com poucos sujeitos envolvidos, com contextos artificiais e com um controle muito fraco

⁹ Mais tarde, o próprio Meara comenta este facto: «A few years ago it was fashionable to describe vocabulary acquisition as a neglected aspect of language learning. Recently, however, interest in this area has unexpectedly grown at an enormous rate.» (Meara 1996:27). Ainda alguns anos depois do reparo de Meara, Susan Gass tem uma perspectiva um pouco diferente: «it is perhaps misleading to think of the lexicon as being the 'neglected component' because many areas of research impinge on questions of the lexicon.» (S Gass 1987:130).

¹⁰ «Hardly any of the material I have read recently contains results which shock or surprise the reader.» (Meara 1995, vol. 11:2) «...current research is making the same mistakes that can be found in the earlier work on vocabulary acquisition.» (Meara 1996:36).

quanto às línguas em presença (Meara 1996:32-34) ¹¹. Num texto de 1997, insiste em que, se observarmos, no seu conjunto, a investigação que tem estado a ser feita durante os anos 90, concluiremos que ela não difere muito da que foi publicada durante os anos 80 (Meara 1997:110). Alguns dos trabalhos partem do pressuposto, muitas vezes apresentado como hipótese, que o sujeito, uma vez exposto a *input* escrito ou oral, aprenderá palavras. A investigação consiste em encontrar evidência que suporte esta 'hipótese'. Com alguma ironia, comenta que as ditas 'hipóteses de trabalho' não passam de expressões do senso comum, e que estes investigadores agem como jardineiros : põem sementes num vaso para confirmarem que elas germinarão e se transformarão em flores (Meara 1997:113). E isso, afirma, deve-se ao facto de se trabalhar, nesta área, sem modelos. Os psicólogos, diz, têm uma longa tradição de pensar as questões psicolinguísticas enquadradas em modelos formais, no entanto, o mesmo não acontece entre os que trabalham em linguística aplicada (Meara 1997:109) ¹². Schmitt & McCarthy parecem ter uma opinião um pouco diferente. Segundo estes, a falta de modelos não seria a causa do atraso dos estudos, mas a consequência:

In the case of vocabulary, however, we still have only a rudimentary idea of how words are acquired, causing theoretical modelling to lag behind. (Schmitt & McCarthy 1997:325)

Muitos trabalhos sobre vocabulário têm-se preocupado em investigar qual o número de palavras de que dispõe um falante, nativo ou não, e com a discussão das dicotomias passivo/activo e receptivo/produtivo ¹³. Uma das técnicas mais seguidas para investigar o **vocabulário receptivo** usa listas de palavras, calibradas em função de vocabulários fundamentais, salpicadas de algumas pseudo-palavras. A tarefa consiste em o participante

¹¹ Ard & Gass 1987, por exemplo, um texto muito citado, relatam um estudo feito com um grupo de 23 sujeitos: 5 árabes, 1 chinês, 1 farsi, 1 grego, 5 japoneses, 1 português e 9 espanhóis. É um dos trabalhos referidos em Meara 1996.

¹² «The L2 research literature contains lots of examples of what might be broadly described as descriptive research on vocabulary acquisition, but very few examples of explanatory, model-based research, which attempts to account for this learning.» (Meara 1997:109).

¹³ «...the crucial distinction between active and passive vocabulary might simply be that active vocabulary items are connected to their parent lexicons by more than one type of connection» (Meara 1997:119).

no estudo marcar 'sim' ou 'não', conforme conhece ou não cada uma delas. Um ou outro estudo mais sofisticado pede ao sujeito que estabeleça relações paradigmáticas e sintagmáticas com outras palavras. Há também uma linha de estudos de **vocabulário produtivo** que investiga a sua extensão em composições escritas. Esses estudos usam uma série de medidas para a análise, tais como *originalidade*, *riqueza*, *densidade*, *sofisticação* e *variação lexical* (cf. por exemplo, Linnarud 1975, Medeiros de Carvalho 1991, Laufer & Nation 1995). Read refere a inadequação destes testes como indicadores da dimensão do léxico, especialmente no que respeita as palavras mais frequentes, já que estas podem ter um grande número de sentidos e de usos (Read 1997:315). Mas Read não é o único a criticar esta metodologia.

No já distante ano de 1976, Richards chamou a atenção para o facto de **aprender uma palavra** ser muito mais do que aprender o seu significado¹⁴, e foi o primeiro a fazer uma lista dos diferentes tipos de conhecimento de que um falante de L2 tem de dispor para a usar como se fosse um falante nativo¹⁵. Antes de mais, é preciso reconhecê-la como tal e introduzi-la no nosso léxico mental. E a maior ou menor dificuldade nesse reconhecimento depende, entre outros factores, do canal. Nick Ellis faz notar que há **diversos léxicos especializados para diferentes canais de input e de output**: para a compreensão do discurso, para a leitura, para a produção oral e para a produção escrita. Ou seja, a sua correcta introdução no léxico mental exige o seu estabelecimento numa série de sistemas de representação, de que se destacam o fonológico, o semântico, o conceptual e o sintáctico (Ellis 1996:93-4 e 1997:126).

A aquisição de vocabulário surge assim como uma actividade cumulativa e não como uma questão de tudo ou nada. Palavras desconhecidas são aquelas que não fazem parte de nenhum dos sistemas de representação, que não têm qualquer conexão com o léxico do sujeito.

¹⁴ «...many people still believe that if you know a word's meaning (even if it is only one of several), you 'know' that word.» (Schmitt & McCarthy 1997:325).

¹⁵ Meara 1996 (texto disponibilizado na Internet) comenta que os oito critérios de Richards 1976 são muito mais de carácter linguístico do que de carácter psicolinguístico ou pedagógico.

Palavras conhecidas são aquelas que têm conexões, mas o número destas conexões pode variar. Certas palavras têm poucas conexões. O sujeito tem delas um conhecimento pobre; outras têm muitas conexões e, por isso, são bem conhecidas (Meara 1997:118-9) ¹⁶. Aquilo que torna difícil a aquisição do vocabulário é o facto de demorar tempo e exigir esforço estabelecer as conexões e, conseqüentemente, desenvolver um léxico bem organizado. Este problema não se põe quando o léxico a ser aprendido contém apenas uma meia dúzia de palavras (Meara 1996:34) ¹⁷.

Afigura-se conveniente reflectir aqui sobre o que se entende por **vocabulário** e por **léxico**. Parece que, até ao momento, tenho usado ambos os termos de forma aleatória. Não é bem assim. Hatch & Brown 1995 iniciam a obra *Vocabulary, Semantics, and Language Teaching* afirmando que é costume começar qualquer trabalho pela definição do campo de estudo. Lamentam, no entanto, que contrariamente ao que acontece com outras áreas de estudo, tais como a fonologia, a sintaxe ou a análise do discurso, não haja um termo que abranja a semântica, o léxico e o vocabulário, e que cada um tenha a sua definição:

*The term **semantics** refers to the study of meaning and the systematic ways those meanings are expressed in language. "The term **lexicon** refers to the overall system of words forms and, when we include morphology, the study of word formation in language. The term is also used to refer to the way forms might be systematically represented in the brain, that is, the **mental lexicon**.*

*The term **vocabulary** refers to a list or set of words for a particular language or a list or set of words that individual speakers of a language might use. (Hatch & Brown 1995:1) ¹⁸.*

¹⁶ «Let us suggest [...] that **an acquisition event consists of** the building of a connection between a newly encountered word, and a word that already exists in the learner's lexicon. This connection might be a link between the new L2 word and its L1 translation equivalent, or it might be a link between the new L2 word and an already known L2 word.» (Meara 1997:118-9).

¹⁷ Ver, por exemplo, Yang 1997: aprendizagem de uma gramática rudimentar – “keki” – e de um léxico de 67 palavras. Ou Gerganov & Rangelova 1993. Trabalhos como este são feitos porque «It is much easier to think of the lexicon as an unstructured list of words, rather than as a complex, interlocking structure» (Meara 1996:38).

¹⁸ No *Dicionário de Termos Linguísticos* (Xavier & Mateus, org.), pode ler-se: «Em gramática generativa o termo [**léxico**] designa a componente da gramática que contém a especificação abstracta morfofonológica de cada item lexical e os seus traços sintácticos, incluindo os traços categoriais e os contextuais.» (fonte: Chomsky 1981) (DTL:1992, vol II:225). **Vocabulário:** «Lista exaustiva das palavras de um corpus. O vocabulário distingue-se do léxico que é entendido como o inventário de todas as lexias de um dado estado de língua.

Assim, léxico “não é uma outra palavra para vocabulário – é um conceito muito mais rico e que não tem sido explorado” (Lewis 1993/1996:193).

Meara (1993:293) afirma que, na área da L2, há, de facto, uma tendência para ver o léxico como uma lista não estruturada a que se vão juntando mais palavras que vão sendo aprendidas. Mas, à medida que as teorias linguísticas têm ido na direcção de considerar o léxico como central na construção da gramática interna dos falantes, os estudos de L2 têm-se desenvolvido, cada vez mais, sob este conceito ‘rico’ de léxico ¹⁹. Uma linha de investigação que pretende responder a este conceito de léxico tem-se ocupado em observar quais os factores que afectam a aquisição de uma palavra; ou seja, o que é que faz que seja mais fácil ou difícil estabelecer as conexões correctas dentro de cada um dos sistemas. São investigados certos factores de natureza mais formal, como a pronunciabilidade, a ortografia, a extensão, a morfologia, a semelhança com outro ou outros itens lexicais; ou de natureza semântica: o facto de serem concretas ou abstractas, o grau de especificidade, restrições de registo e idiomaticidade (ver, entre outros, Ellis & Beaton 1993 e Laufer 1990 e 1997).

No já referido trabalho publicado em 1997, Schmitt & Meara afirmam que a investigação não tem explorado as inter-relações entre os vários tipos de conhecimento e reivindicam estar a fazê-lo pela primeira vez. Concluem que pelo menos os dois tipos de conhecimento investigados, associação de palavras e sufixos, estão não só relacionados entre si, mas também com a dimensão do léxico e com o nível geral de conhecimento da língua (Schmitt & Meara 1997:31). Afirmam ainda que o facto de ter sido demonstrada a inter-relação entre diferentes tipos de conhecimento de palavras abre um novo campo de investigação: examinar como é que essas diferentes

(fonte Guilbert 1971) (DTL:1992, vol. II:399-400). Levelt 1989: 6 define **léxico mental** como “the store of information about the words in one’s language.»

¹⁹ Termo usado, por exemplo, em Scaramucci 1997; trabalho feito na Unicamp, mas sobre falantes de PB a aprenderem Inglês.

componentes que constituem o conhecimento se relacionam entre si de modo a construir aquilo que sabemos sobre as palavras (Schmitt & Meara 1997:30) ²⁰. O aspecto mais original deste estudo está em estabelecer a relação entre nível geral de conhecimento da língua, dimensão do léxico e tipos de conhecimento disponíveis para cada palavra. Está em inter-relacionar a dimensão do vocabulário com a profundidade do conhecimento do léxico. Está em relacionar quantidade e qualidade do conhecimento lexical. E, segundo os autores, nunca nos podemos esquecer de que há um conjunto de competências que concorrem para o conhecimento total de uma palavra, que «each individual acquisition event is small [...] this small events add up to a rich vocabulary structure» (Meara 1997:119).

Por isso mesmo, advertem: a investigação futura tem de desenvolver métodos que permitam avaliar o conhecimento parcial das palavras, e, provavelmente, usando para cada estudo tarefas produtivas e receptivas; tem de observar grupos, mas ter também em consideração a variação individual (Schmitt & Meara 1997:33) ²¹. Meara resume a sua proposta numa frase:

Maybe it is time for some of us to stop being vocabulary acquisition gardeners, and become applied linguistic microbiologists instead...
(Meara 1997:121).

0.3. *Natureza e objectivos do estudo*

Paul Meara é de opinião que têm sido feitas poucas tentativas para **avaliar como é que, globalmente, se desenvolve a estrutura do léxico de L2, e como é que a dimensão do léxico afecta o modo como novas**

²⁰ «This study examines how two types of word knowledge, word associations and grammatical suffix knowledge, change over time both receptively and productively. Ninety-five secondary and postsecondary Japanese students were tested on three word associations and inflectional and derivational suffixes for each of 20 verbs, once near the beginning of their academic year and once near the end. The results showed their average vocabulary gain was 330 words. The students showed rather poor knowledge of the allowable suffixes for the verbs, especially the derivative suffixes. [...] Word association knowledge and suffix knowledge were shown to correlate with each other and with total vocabulary size....» (Resumo que introduz Schmitt & Meara 1997).

²¹ Porque «important information about individual behavior is often lost in groups analyses» (Schmitt & Meara 1997:20), Meara propõe a um grupo de investigadores que apresentem trabalhos sobre aquisição individual e reune-os no vol. 11(2) da revista *Second Language Research*.

palavras são aprendidas (Meara 1996:38). Há, afirma, muito poucos estudos que comparem falantes de diferentes L1 a aprenderem a mesma L2, bem como de falantes da mesma L1 a aprenderem diferentes L2. Faltam também, segundo ele, trabalhos sobre a aquisição do léxico por falantes nativos de línguas não indo-europeias (Meara 1996:37).

A investigação que me proponho desenvolver procura evitar alguns dos aspectos criticados por Meara e atender a algumas das suas propostas. Para isso, **estudarei um corpus transversal constituído por composições escritas, produzidas por sujeitos falantes de cinco diferentes L1, uma das quais é não indo-europeia, a aprender a mesma língua, o Português Europeu.**

Este trabalho tem como objectivos centrais

- (1) avaliar que vocabulário usa e que conhecimento manifesta sobre ele, em material escrito, um grupo de aprendentes de português língua não materna, em contexto de aprendizagem semi-formal, depois de um mínimo de 120 horas de ensino formal, durante um período de mais ou menos 3 meses;
- (2) apreciar quais os itens lexicais e quais as componentes e as características destes que mais resistência oferecem à aquisição; e
- (3) sugerir aplicações dos resultados do estudo ao ensino e à produção de materiais.

O estudo pretende assim inter-relacionar conhecimento quantitativo, ou seja, dimensão do léxico, com profundidade desse mesmo conhecimento. Para satisfazer esse objectivo, foi construída uma base de dados com o vocabulário disponível e com os diferentes tipos de desvios detectados no léxico. Deste modo, combina-se, neste projecto, uma linha de investigação mais tradicional, a que se tem chamado **aprendizagem de vocabulário**, com uma linha de investigação que tem vindo a ser desenvolvida mais recentemente e que assume que **a aquisição e a construção do léxico mental** é uma actividade complexa, já que cada palavra tem uma série de

propriedades que demoram tempo e esforço até estarem devidamente estabelecidas ²².

Estuda-se material escrito e assume-se que o conhecimento de palavras para esta modalidade envolve certas componentes específicas, diferentes das do oral. Entre elas, destaca-se a ortografia. Assume-se, no entanto, que o texto escrito pode fornecer alguma informação sobre aquisição da fonologia e da fonética.

Estudam-se composições de que, no processo de selecção, se controlaram as temáticas, de modo que todos os *sub-corpora* apresentem características semelhantes. E para que os temas abordados fossem suficientemente variados, procurou-se que cobrissem o mais possível os abordados nas entrevistas do *Português Fundamental*. Constituiu-se um grupo de controle de falantes nativos de Português Europeu, de modo a poder estabelecer comparações quanto à dimensão e natureza do léxico, não só entre os diferentes grupos mas também em relação a falantes nativos.

As L1 dos sujeitos envolvidas no estudo são o Espanhol e o Francês, o Alemão e o Sueco, e o Chinês ²³. Quase todos conhecem pelo menos mais uma língua, uma boa parte conhece mesmo duas e alguns três ou mais. Assume-se esse facto já que esse é o perfil habitual dos estudantes de Português Europeu L2. No entanto, foram excluídas da amostra todas as composições de sujeitos com um bom conhecimento de outra ou outras línguas românicas (para além da sua L1). Pretende-se, deste modo, criar um leque que vai de uma língua muito próxima do Português, o Espanhol, ao Chinês, língua não indo-europeia, com o objectivo de avaliar o papel dos

²² «...most often research carried out by applied linguists treats the lexicon as synonymous with ‘vocabulary’ learning; in this way, the lexicon is often understood to mean ‘word list’. Hence, research has focused on word lists, L2 acquirers’ knowledge of and ability to remember words, and occasionally, issues related to the traditional notion of collocation, and transfer of cognates [...] Such studies are extremely important and should be seen a complement to the perspective [Linguistic theory, language acquisition, and the lexicon] taken in this volume.» (Juffs 1996:1-2).

²³ Uso o termo ‘Chinês’ porque o *sub-corpora* a que ele se refere é constituído por documentos de falantes de Mandarim e de Cantonês. No entanto, estou consciente de que «the Chinese dialectal complex is in many ways analogous to the Romance language family in Europe» (Norman 1988:187), e de que o critério para considerar uma dada variedade linguística como língua ou dialecto é de natureza socio-política e não linguística.

conhecimentos prévios no desenvolvimento do léxico e de inter-relacionar dimensão e profundidade do conhecimento.

Não se trata, no entanto, de um estudo contrastivo. Trata-se de, pondo a tónica em factores intralexicais, avaliar quais as constantes no processo de aprendizagem do Português Europeu L2. Ou seja, quais as características de certas palavras (formais ou semânticas) que fazem com que umas sejam mais fáceis e outras mais difíceis de aprender. Trata-se, portanto, de um estudo cujos resultados, quando devidamente ponderados, poderão eventualmente ser extensíveis a aprendentes de outras L1, já que a amostra é relativamente variada a este nível.

Nos comentários de Paul Meara, aquilo que, em geral, parece merecer bastante a sua desaprovação é o facto de muito do trabalho realizado, para além da dificuldade que constitui compatibilizar os seus resultados, não reunir «a challenging combination of real-world constraints and rich theory.» (Meara 1996:39). Creio que a metodologia escolhida para o estudo que aqui apresento assegura a presença das condicionantes impostas pelo mundo real: é um estudo que

- (1) se apoia num *corpus* equilibrado, produzido por falantes de línguas diferentes, todos em situação semelhante de aprendizagem semi-formal; e que
- (2) pretende avaliar o desenvolvimento global do léxico nas suas duas vertentes: no seu conjunto e palavra a palavra; é, assim, um estudo original porque combina avaliação quantitativa com avaliação qualitativa.

Será, por isso mesmo, um estudo de tipo microbiologista. E, como é sabido, a ‘microbiologia’ (mas, sejamos justos, também a jardinagem...) é grande consumidora de tempo e de paciência. Por isso, muitas células ficaram por analisar, e muitas outras foram observadas ao microscópio, mas, por diversas razões, dessa observação não se dá conta aqui. O que pretendi, fundamentalmente, foi estabelecer um quadro prévio em que os meus trabalhos futuros se possam ir encaixando.

0.4. *Enquadramento teórico*

Tenho vindo a citar Paul Meara pelo que ele representa no contexto dos estudos sobre aquisição do léxico por aprendentes de L2. No entanto, ele não é o único a referir a não existência de modelos de aquisição, de produção e de compreensão de léxico, e até mesmo de modelos de aquisição de L2, reconhecidos por um número representativo de investigadores da área quer da L2 quer da Psicolinguística ²⁴. Skehan (1998), por exemplo, reviu recentemente três dos mais reconhecidos modelos de aquisição de L2 – ‘Universal Grammar’, ‘Multidimensional Model’ e ‘Analysis-Control’ ²⁵ – e concluiu que cada um dos modelos enfatiza só uma das três componentes que, segundo ele, devem ser tidas em consideração num modelo de aquisição de L2, respectivamente: representação, aprendizagem e processamento (Skehan 1998:87). Para se chegar à construção desse modelo, ou simplesmente para apoiar a investigação, são muitos os que advogam a necessidade de combinar informação de diferentes disciplinas. As mais referidas são a Linguística e a Psicologia:

Theories of acquisition can usefully draw on linguistic theories for detailed descriptions of linguistic phenomena, but they can't, in principle, use them as the sole source for predictions about performance (Clark 1993:259).

Linguists have the edge on grammar, just as psychologists do on learning, which is why I believe that the study of SLA should be a collaborative enterprise (Ellis 1996:120).

Para outros, qualquer modelo pressupõe também o recurso a investigação em Neurobiologia. Jacobs & Schumann são da opinião que os investigadores de aquisição da linguagem devem começar a incorporar a realidade neurobiológica na sua percepção dos processos de aquisição e que esta perspectiva os ajudará a encontrar uma estrutura para avaliação e integração das várias abordagens de aquisição (Jacobs 1988:305; Jacobs & Schumann 1992:295).

²⁴ «There is as yet no integrated model of comprehension and production in L2.» (Green 1993:270).

²⁵ Para ‘Universal Grammar’, ver White 1989; para ‘Multidimensional Model’, ver Pienemann 1984 e 1989 e Pienemann *et alii* 1988, e para ‘Analysis-Control’ ver Bialystok 1981 e 1990 e ainda Hakuta 1994.

Como é sabido, a Psicolinguística é a ciência que se ocupa da compreensão, da produção e da aquisição e desenvolvimento da linguagem (Aitchison 1994/1987). No próprio nome está implícita a colaboração entre Linguística e Psicologia. Ao longo da sua história de 50 anos, a proeminência de cada uma das áreas tem variado. Durante os anos 60 (durante o chamado período linguístico), assumia-se que o estudo da competência seria crucial para a compreensão da performance. Deste modo, porque a frase era considerada como a unidade básica, a gramática ocupava um lugar central nos estudos psicolinguísticos. Um enorme interesse, durante os anos 70, pela aquisição e desenvolvimento da linguagem infantil, fez com que este período desse lugar a empresas mais interdisciplinares. Bever (1970) e Slobin (1973), «talvez os melhores representantes deste período – o período cognitivo» (Kess 1992:21), rejeitavam a centralidade e a independência da gramática, argumentando que a linguagem é um de entre vários produtos resultantes de processos cognitivos fundamentais que governam a aquisição de estruturas linguísticas e que essas estruturas linguísticas não são aprendidas independentemente de conceitos semânticos e de funções discursivas.

Entretanto, durante os anos 80, um outro paradigma entrou em jogo: o paradigma computacional. A mente humana, tal como os computadores, manipula símbolos e toma decisões com base em inferências e em informação armazenada. Graças a estes e outros desenvolvimentos, não há, actualmente, uma ciência que prevaleça sobre a outra. Em vez disso, há uma forte actividade interdisciplinar com os investigadores muito atentos à investigação desenvolvida em campos adjacentes. De qualquer modo, alguns linguistas continuam a favorecer a hipótese da modularidade da linguagem, enquanto que, aos psicólogos, atrai mais a interacção da linguagem com princípios gerais de processamento da informação e de aprendizagem. Mas, na verdade, os paradigmas generativo, cognitivo e computacional são todos paradigmas cognitivos na medida em que todos têm o compromisso de estudar o conhecimento de que dispõem os seres humanos no uso da linguagem. Assim, actualmente, a Psicolinguística faz parte de um muito vasto programa de investigação: a natureza do conhecimento, a estrutura das representações mentais e a forma como

estas são usadas nos processos mentais e nas tomadas de decisão (Kess 1992:26-29). A esta actividade interdisciplinar tem-se chamado **Ciência Cognitiva** ²⁶. Aitchison define Ciência Cognitiva como «um campo de estudo que cobre áreas do conhecimento relacionadas com a capacidade de pensar e de falar, nomeadamente, a inteligência artificial, a linguística, a psicologia e a filosofia» (Aitchison 1992:29). Lakoff 1987 define-a e também enuncia as suas principais questões:

***Cognitive science** is a new field that brings together what is known about the mind from many academic disciplines: psychology, linguistics, anthropology, philosophy, and computer science. It seeks detailed answers to such questions as: What is reason? How do we make sense of our experience? What is a conceptual system and how is it organized? Do all people use the same conceptual system? If so, what is that system? If not, exactly what is there that is common to the way all human beings think? The questions aren't new, but some recent answers are (Lakoff 1987:xi).*

A investigação em bilinguismo e em L2 pode certamente colaborar na resposta a estas questões.

No início dos anos 70, investigadores como Corder (1967, 1974), Bailey, Madden e Krashen (Bailey & Madden & Krashen 1974) começaram a mostrar que as interlínguas são sistemas organizados, ainda que particularmente instáveis. Desde então, a investigação em L2 tem sido considerada como podendo fornecer informação útil não só para aplicação ao ensino mas também para o conhecimento da natureza da linguagem humana ²⁷. Três grandes temas de debate têm ocupado os investigadores desta área: **(1)** a existência ou não de restrições impostas pela idade à aprendizagem de uma L2; **(2)** a questão da representação das várias línguas, única ou separadas, nos falantes bilingues ou multilingues. Mas o que mais tem ocupado os investigadores, quer da L1 quer da L2, é aquilo a

²⁶ «Psycholinguistics is normally taken to cover the questions of how humans understand, produce and acquire speech. But its boundaries are somewhat fluid, and in recent years psycholinguistics courses have tended to be supplemented by information about language and the brain (technically neurolinguistics) and about speech disorders (clinical linguistics, aphasiology). They have also been influenced by the new, interdisciplinary area of cognitive science – an amalgam of linguistics, psychology, philosophy and artificial intelligence.» (Aitchison 1992:7).

²⁷ Embora Susan Gass chame repetidamente a atenção para que as outras disciplinas raramente têm em conta a investigação desenvolvida nesta área (Ver, entre muitos outros textos da autora, Gass 1993).

que Corder (1983/1992:24) chama ‘the question of the starting point’, ou seja, **(3)** a aquisição e desenvolvimento da linguagem consistem na actualização de uma gramática universal ou são princípios cognitivos gerais que determinam o modo como a linguagem é adquirida? ²⁸

No que diz respeito ao terceiro tema em debate, de facto, até mesmo os investigadores inicialmente com opções mais claras quanto a uma ou outra das posições têm vindo a revê-las e a fazer certas concessões à outra parte. Em Leiria (1991:58-61) salientei este facto, exemplificando com o LMC (*‘Language Making Capacity’*) de Slobin, por um lado, e com uma afirmação de Lydia White, que aqui repito, por outro: «as duas perspectivas não são incompatíveis: é provável que alguns aspectos da aquisição sejam explicados pela gramática universal, outros por princípios de processamento, e outros ainda por nenhum destes» (White 1989:17) ²⁹. Mais recentemente, uma investigadora da mesma família teórica, Juana Liceras, diz que

cualquiera de las lenguas naturales se realiza como una estructura cognitiva, resultado de la interacción de los principios de la GU y el medio. [...] La tradición chomskiana mantiene no sólo que el lenguaje es modular sino que es autónomo, es decir, independiente de los otros módulos cognitivos [...] lo cual no quiere decir que esté aislado de los otros módulos cognitivos, pero sí que tiene su propio sistema de funcionamiento (Liceras 1997:300-1).

Mas, se são muitos os que admitem ou insistem em que a aquisição é influenciada por factores cognitivos, são também muitos aqueles que nos recordam que uma perspectiva exclusivamente psicológica, por si só, não pode explicar as restrições linguísticas. Por isso, uma abordagem cognitiva

²⁸ «A major controversy in language acquisition research (both first and second) is whether or not acquisition can best be characterized by means of innateness. One view holds that a child comes to the learning task with a Universal Grammar (UG) which allows him or her to construct a grammar of a language on a basis of limited data. Another view maintains that language acquisition is a form of social interaction.» (S Gass 1988:198).

²⁹ «Even the strong argument for universal grammar reaches limits. The features of language that are indeed universally realised across all known languages are relatively few. White (1989) cautiously delineate what those are. The linguistic features that seem most controlled by UG are those relating to syntax, although even here it seems that only a part of syntactic structure is under UG control. Those aspects that appear to be outside UG are lexicon, and possibly discourse and pragmatics.» (Bialystok 1994:557). Em Leiria 1991 fiz uma introdução às duas abordagens: a de tradição chomskiana e a funcional tipológica (Capítulo 2: 45-60).

deve ser vista, não como competitiva, mas sim como complementar de uma abordagem linguística:

[A] cognitive psychological description of second language learning provides, none the less, a partial account, and needs to be linked to linguistic theories of second language acquisition. By, itself, for example, the cognitive perspective cannot explain such linguistic constraints as are implied in markedness theory or that may result from linguistic universals (McLaughlin 1990:126).

Por outro lado, e como vimos, a aquisição de uma palavra envolve componentes, pelo menos, de natureza fonológica, semântica e sintáctica. Como é sabido, certas áreas da investigação linguística, como é o caso da sintaxe ou da fonologia, têm-se desenvolvido preferencialmente no âmbito de teorias mais formais, enquanto outras, como é o caso da semântica, o têm feito sobretudo no quadro de abordagens mais empíricas, funcionalistas, tipológicas ou cognitivas. A recomendação “Be collaborative!”, feita por vários investigadores da área da L2 (cf. Singleton 1999:276), pode traduzir-se em interdisciplinaridade, mas também, no âmbito da linguística, na consideração de várias teorias. A abrangência do léxico justifica plenamente seguir este caminho ³⁰.

Na verdade, é hoje claro para muitos que a complexidade inerente à aquisição de uma língua, e muito em especial de uma L2, só poderá vir a ser explicada se combinarmos, com critério e sensatez, aquilo que a investigação desenvolvida pelas várias disciplinas que se ocupam da cognição humana põe à nossa disposição. A informação é muito rica e abundante e, por vezes, contraditória.

[Sensitive period for language learning, multilingual representation, and the epistemological status of linguistic knowledge] share a central role in the field of second language acquisition. Each is fundamental to some of our most basic questions of second language learning and use, and each has important practical consequences for language

³⁰ Salvas as devidas proporções, e a título de exemplo, Castilho 1995 assume a ausência de um quadro teórico único da *Gramática do Português Falado*: “Os grupos de trabalho de Fonética e Fonologia, Morfologia e Sintaxe das Relações Gramaticais elegeram a percepção mais ‘formal’ das tarefas, enquanto que os grupos de trabalho de sintaxe de classes de palavras e organização textual-interactiva se mostraram mais ‘funcionalistas’. Convencionou-se, naquele momento, que os diferentes capítulos da gramática advertiriam o leitor a respeito das diferentes opções tomadas.” (Castilho 1995:77). Segundo o autor, o facto de terem partido de “abordagens intuitivas” permitiu-lhes “operar com critérios originários de uma literatura variada.” (Castilho 1995:80).

*policy and pedagogy. Although important research and theorising have developed in recent years in all three of these areas, **the choice between competing perspectives seems no more clear than before.** In fact, as the arguments become more convincing and the evidence more documented, the competing positions become more entangled (Bialystok 1994:567-8).*

Segundo a autora, em qualquer dos casos dos temas em debate, o que se afigura como correcto não é tomar posição a favor de uma ou outra das teses mas sim «encontrar as condições sob as quais cada um dos pontos de vista fornece a interpretação correcta» (Bialystok 1994:552).

0.5. Organização

Assim, esta obra é constituída por duas partes. Na **Parte I**, *Léxico, bilinguismo, aquisição e ensino de L2*, apresenta-se, organizada em quatro capítulos, uma revisão da literatura considerada como relevante para a investigação em aquisição de léxico em L2. No **capítulo 1**, *Léxico*, faz-se uma breve referência ao papel dinamizador desempenhado pela linguística computacional e pelos estudos de *corpora* nos estudos de léxico, e revê-se, de modo muito sucinto, investigação oriunda de várias abordagens da linguística teórica (generativa, funcional e cognitiva). No **capítulo 2**, *Léxico e bilinguismo*, assume-se um conceito abrangente de bilinguismo, considera-se o léxico mental como o elo central no processamento e na produção, e caracteriza-se o léxico mental de falantes bilingues. No **capítulo 3**, *Léxico e aquisição de L2*, dá-se uma panorâmica breve da investigação realizada, e procura mostrar-se como, à semelhança de outras áreas da Ciência Cognitiva, ela tem vindo progressivamente a centrar-se mais no léxico. No **capítulo 4**, *Léxico e ensino de L2*, faz-se uma viagem rápida por teorias e metodologias de ensino para avaliar qual o lugar que o vocabulário / léxico tem tido na sala de aula, e quais as propostas que têm sido feitas quanto ao papel do ensino no processo de aprendizagem.

A **Parte II**, *Léxico, aquisição e ensino de PE L2*, que constitui o estudo propriamente dito, está organizada em três capítulos. No **capítulo 5**, *Recolha e tratamento de dados de aquisição de PE L2*, informa-se sobre as características da população e da amostra que constitui o *corpus*, e sobre a transcrição, organização, etiquetagem e tratamento informático dos documentos que conduziram à criação da base de dados. No **capítulo 6**,

Vocabulário disponível e léxico deficitário, tendo presente que a aquisição e, conseqüentemente, a construção do léxico mental é uma actividade cumulativa, distingue-se vocabulário disponível de léxico deficitário. Apresentam-se resultados que informam sobre a dimensão do vocabulário e sobre o número e natureza dos desvios de cada grupo. Analisam-se as diferentes estratégias usadas para compensar o vocabulário indisponível; e, no vocabulário disponível, identificam-se e comentam-se certos desvios em aspectos formais e em propriedades do léxico. No caso dos aspectos formais, presta-se especial atenção às pistas que a escrita pode fornecer sobre o processo de aquisição do sistema fonológico; e, no que respeita os desvios em propriedades do léxico, apresentam-se resultados que se comentam brevemente no caso dos nomes, adjectivos, verbos de alta frequência, preposições, e em outras categorias. O capítulo termina com o parágrafo *O léxico para além do léxico* em que se acentua a abrangência e a centralidade do léxico na aquisição e na produção. No **capítulo 7**, *Léxico e ensino de PE L2*, fazem-se algumas sugestões para a investigação e para o ensino do português L2 e defende-se que a *abordagem cognitiva*, apoiada em dados de *corpora* (de falantes nativos e não nativos) que guiem uma abordagem preferencialmente lexical, deve servir de respaldo ao ensino, à avaliação da aprendizagem, e à produção de materiais de L2.

Parte I

Léxico,

bilinguismo

aquisição e ensino de L2

Capítulo 1

Léxico

1.1. Introdução

*Why is it that the element of language which the native speaker feels that he knows best is the one about which linguists say the least? To the untutored person, speaking is putting **words** together, writing is a matter of correct **word**-spelling and **word**-spacing, translating is getting **words** to match **words**, meaning is a question of **word** definitions, and linguistic change is merely the addition or loss or corruption of **words**. Is the reason why the average individual embraces the **word** but the linguist shuns it, a single coin with two sides? (Bolinger 1963:113).*

A preocupação de Bolinger com as palavras ³¹, ao colocar-se na dupla posição de linguista e de falante comum, parece ser o despertar de uma atitude em relação ao léxico diferente daquela que, até meados dos anos 60, dominava os estudos. É verdade que um conjunto de actividades metalinguísticas têm feito do léxico um objecto central: é o caso da Filosofia da Linguagem, da Psicologia, da Psicologia Cognitiva, da Antropologia, da Sociologia da Expressão e da Comunicação e, claro, da Linguística Teórica, da Lexicologia e da Lexicografia (Rey 1991), mas, por todas, nomeadamente pela Linguística, o léxico foi, durante muito tempo, encarado como uma entidade vasta e imprecisa, estruturado segundo leis mal conhecidas e

³¹ «...one would look in vain for a simple definition of the **word** concept» (Singleton 1999:10). **1)** “What we call ‘words’ in one language may be units of a different kind from the ‘words’ of another language” (Lyons 1968:206); **2)** o conceito de palavra pode ser abordado do ponto de vista: **ortográfico**: “a continuous string of characters uninterrupted by a space”(Nelson 1997), **mas** a palavra não é um produto da literacia...; **fonético**: “any segment of a sentence bounded by successive points at which pausing is possible” (Lyons 1968:206), **mas** este critério de pausa potencial não pode ser considerado como uma definição teórica...; **semântico**: “the minimum meaningful unit of language” **mas** “there are single units of meaning which are conveyed by more than one word...” (Carter 1987:5); **grammatical**: “seems to be least problematic [...] The grammatical approach uses the criteria of ‘positional mobility’ and ‘internal stability’ – which essentially goes back to Bloomfield’s (1933) definition of the word as a ‘minimal free form’, **mas** então artigos e preposições não seriam verdadeiras palavras...” (Lyons 1968:206). Para esta discussão ver, entre outros, Singleton 1999:9-14.

diferentes segundo o ponto de vista. Para documentar alguns desses pontos de vista, aparecem transcritas, em trabalhos mais recentes, afirmações como estas:

The lexicon is really an appendix of the grammar, a list of basic irregularities (Bloomfield 1933).

Regular variations are not matters for the lexicon, which should contain only idiosyncratic items (Chomsky & Halle 1968:12).

The lexical items of a language can indeed be presented as a mere list (Kempson 1977).

If conceived of as the set of listemes, the lexicon is incredibly boring by its very nature. It contains objects of the single specifiable types (Words, VPs, morphemes, perhaps intonational patterns, and so on), and those objects that it does contain are there because they fail to conform to interesting laws. The lexicon is like a prison – it contains only the lawless, and the only thing that its inmates have in common is lawlessness (Sciullo & Williams 1987:3).

Mas a posição em relação ao léxico tem vindo a mudar em todas as disciplinas que por ele se interessam. A prová-lo está o recente comentário de Alan Juffs sobre a última dessas afirmações:

The lexicon may be a jail for the lawless, but even jails have rules (Juffs 1996:79).

Esta não é uma opinião isolada. Na verdade, o estatuto do léxico e o interesse pelo seu estudo têm vindo a mudar, muito em especial nos anos 90, na Linguística, mas também na Psicolinguística e em várias áreas da Linguística Aplicada, nomeadamente na Lexicologia. E, à medida que os estudos têm avançado, tem sido reconhecido ao léxico um papel cada vez mais central na construção da gramática interna do falante de qualquer língua natural. Esta centralidade deve-se sobretudo ao facto de a Linguística, a partir de Chomsky 1965, vir progressivamente a considerar o léxico como uma componente cada vez mais abrangente de propriedades e funções, até há algum tempo atribuídas em particular à sintaxe.

[R]ecent developments in principles and parameters grammar have tended to push the parameters away from the syntax towards the lexicon. Chomsky's succinct 1989 position was that 'there is only one human language apart from the lexicon, and language acquisition is in essence a matter of determining lexical idiosyncrasies' (Chomsky 1989) [...] Perhaps parameters only belong to the lexicon (Cook 1994: 498-9).

De facto, nos últimos anos, grande parte das teorias têm atribuído ao léxico uma posição central na teoria linguística. O léxico deixou de ser um apêndice da gramática para se converter numa das componentes preferenciais da descrição linguística. Um dos motores desta mudança de atitude tem sido a Linguística Computacional e a sua necessidade de responder adequadamente a questões como estas:

Should we have confidence even in the assumption that discrete units of text, such as words, can be reliably associated with units of meaning? (Sinclair 1991: 3).

[H]ow to provide, fully and adequately, the systems with the lexical knowledge they need to operate with the proper degree of efficiency [?]. The answer, to which the community is converging today, is to extract the lexicon from texts themselves (Boguraev & Pustejovsky 1996:3).

Graças à possibilidade de armazenar grande quantidade de textos em computador tratados em programas de concordâncias ³², a resposta encontrada para a segunda questão, isto é, a extração de léxico de textos, teve, segundo Sinclair (1991), uma consequência inesperada. Na construção de discurso, operamos com recurso a dois princípios: o *princípio da livre escolha* (*open-choice*) e o *princípio idiomático* (*idiom*). Operando em exclusivo com o princípio da livre escolha, não produziríamos textos normais porque, frequentemente, as palavras parecem ser escolhidas aos pares ou em grupos. São esses grupos, essas combinatórias (*collocations*) ³³, que se apresentam com bastante frequência, que ilustram o princípio idiomático. Por isso, «a model of language which divides grammar and lexis, and which uses the grammar to provide a string of lexical choice points, is a secondary model» (Sinclair 1991:110-115). O impacto desta conclusão foi significativo nas várias teorias linguísticas, e permitiu assumir que

It is thus perfectly legitimate to describe language both in terms of the system we use and our use of this system. (Kennedy 1998:270).

³² «A **concordance** is a collection of occurrences of a word-form, each in its own textual environment.» (Sinclair 1991:32).

³³ «**Combinatória** é uma sequência de palavras arbitrária e recorrente» (Benson 1990). Pode apresentar diferentes graus de coesão: *grupos cristalizados* (como expressões idiomáticas), *grupos semi-cristalizados*, e *coocorrentes privilegiados* (que podem ser de dois tipos: combinatórias lexicais e combinatórias gramaticais; nestas, um elemento lexical dominante coocorre com elementos gramaticais, como é o caso das regências) (Pereira 1994).

Este modo de olhar a relação entre o léxico e a gramática, ainda que, por vezes, traduzindo-se em posições mais moderadas, tem tido um forte impacto tanto na Lexicografia e nos estudos de análise de *corpora*³⁴ como na linguística computacional e na linguística teórica. E, como tem acontecido frequentemente, a linguística teórica, muito mais do que os estudos de psicolinguística, embora muitas vezes com considerável atraso, tem tido uma forte influência nas teorias de L2. Por isso, se essa influência ainda não se fez sentir de forma evidente no que respeita a prática da sala de aula, ela já se faz sentir no âmbito da investigação e das propostas para o ensino. O melhor exemplo disso é a *Abordagem Lexical* (Nattinger & DeCarrico 1992 e Lewis 1993) ao partir do pressuposto que

Language consists of grammaticalised lexis, not lexicalised grammar
(Lewis 1993/1996:vi).

Podemos dizer que esta progressiva mudança de posição, que trouxe consequências inevitáveis quanto à definição do âmbito do léxico e da sua relação com a sintaxe, com a semântica, com a morfologia e até com a pragmática, se deu em três linhas de investigação dentro da linguística – as **abordagens generativista, funcional e cognitivista** – e no diálogo que estas têm mantido com a linguística computacional, com os estudos de *corpora* e com outras disciplinas que se ocupam da linguagem humana enquanto actividade cognitiva. Nas páginas que se seguem, procurarei mostrar, de forma muito sintética, e tendo presente a perspectiva da sua aplicação nos estudos de L2, como a investigação linguística centrada no léxico tem vindo a evoluir, e como cada uma das abordagens desenvolveu investigação que pode ser aplicada com muito proveito nos estudos de aquisição de L2.

1.2. A abordagem generativa

É hoje reconhecido por todos que foi Chomsky quem, em 1965, colocou a aquisição da linguagem no centro do debate linguístico.

³⁴ O grupo de trabalho do *Corpus de Referência do Português Europeu*, do Centro de Linguística da Universidade de Lisboa, coordenado por M. Fernanda Bacelar do Nascimento, tem desenvolvido investigação nesta perspectiva. Ver, entre outros, Pereira 1994, Nascimento & Gonçalves 1995, Nascimento & Pereira 1995 e Neto 1995. Ver também Biderman 1995, Silva 1994 e Iriarte Sanromán 1996.

«Enquadrando-se claramente no ‘paradigma cognitivo’, as várias teorias surgidas no âmbito da Gramática Generativa têm potenciado pesquisas interdisciplinares com outras ciências, como a Psicologia Cognitiva, a Inteligência Artificial, a Psicolinguística, a Neurolinguística, sobre os mecanismos da aquisição, do funcionamento e do uso do conhecimento humano.» (Brito 1999:49). É de todos sabido que a Sintaxe sempre teve um lugar de destaque no âmbito da Gramática Generativa. Brito 1998 sintetiza assim o modo como, no quadro da Gramática Generativa, se têm perspectivado as relações entre a Sintaxe e a Semântica Lexical:

[P]odemos dizer que se encontram aí duas fases distintas; uma em que o programa não estava preparado para articular o Léxico e a Sintaxe (o modelo das Syntactic Structures e em grande medida também o modelo dos Aspects); outra fase, a partir dos anos 70, em que essa relação é explicitamente assumida, embora em moldes diversos (Brito 1998:417).

Se em *Aspects of Theory of Syntax* (Chomsky 1965) ainda se mantém a autonomia da Sintaxe ³⁵, proposta em *Syntactic Structures* (Chomsky 1957), o modelo de gramática «tem já alguns mecanismos capazes de dar conta de restrições de selecção entre palavras que compõem as construções linguísticas» (Brito 1998:380). É essa a primeira fase.

Esses mecanismos apoiam-se na **análise de traços semânticos**, desenvolvida no âmbito da *Teoria Standard*, sobretudo por Katz (Katz & Fodor 1963, Katz 1972 e 1977). Esta análise é importante na medida em que tenta estabelecer uma teoria semântica integrada nas teorias da sintaxe e da fonologia. Ela considera que o significado de cada palavra resulta de um conjunto de traços e que esse conjunto é diferente para cada palavra; e que a gramaticalidade ou não de uma frase decorre de o item lexical em causa respeitar ou não os traços de subcategorização ³⁶. É necessário, portanto, criar um conjunto de restrições de selecção que permita predizer as potencialidades de qualquer palavra em todos os contextos. Mas, esta

³⁵ **O Princípio da Autonomia da Sintaxe** impõe que as regras fonológicas e morfológicas devem fazer referência à informação proporcionada pelas regras sintácticas, mas não o contrário. Ou seja, a gramaticalidade de uma frase deve ser garantida pela componente central da teoria, a sintaxe, sem ter de atender à informação fornecida pelas componentes periféricas. (cf. Martín Arista 1999:17).

³⁶ «The correctness of this model [Chomsky 1965] is argued by Katz 1981 – even though Chomsky and other syntacticians have abandoned it, because of it perceived inadequacies, for a syntax which Katz’s semantic theory cannot use.»(Allan 1992:397).

análise, a **análise componencial**, que tem as suas origens em Trubetsky e na Escola Linguística de Praga, postula que certos traços semânticos têm consequências na gramática e outros não. Marcadores semânticos são os traços que têm consequências na gramática de certas línguas (por exemplo, <+/- contável> , <+/- humano> , <+/- animado>). Mas há marcadores semânticos que não têm consequências a nível da gramática (são os 'distinguishers', como por exemplo, <líquido-sólido>, <rígido-flexível>). A estes, a linguística tem prestado menos atenção, embora sejam igualmente importantes para os semanticistas e para os professores de língua. De facto, os investigadores em aquisição da linguagem têm constatado que a análise de traços semânticos é útil para a compreensão da aquisição de sinónimos e antónimos. Seria possível usar indiferentemente sinónimos se todos os traços fossem os mesmos, no entanto, falantes nativos seleccionam um ou outro para uma circunstância particular, o que mostra que há pelo menos um traço que distingue os dois termos. Membros positivos (alto, feliz, bom...) são geralmente mais fáceis do que os negativos correspondentes (cf. Hatch & Brown 1995:16). Um dos aspectos mais controversos desta teoria consiste em alguns componencialistas defenderem que as componentes semânticas básicas são universais (cf. Singleton 1999:34). Mas há muitos outros que têm suscitado discussão e que levaram ao seu abandono: **(1)** a dificuldade em estabelecer o que tem ou não consequências a nível da gramática, **(2)** o que é realmente necessário indicar na descrição, sem cair na redundância; **(3)** como encontrar termos suficientemente abstractos para nomear os traços; **(4)** como interpretar os sentidos metafóricos; e, em particular, **(5)** como enfrentar a questão de muitos traços serem escalas e não dicotomias (Hatch & Brown 1995:28-9).

Esta metodologia de análise foi estendida à descrição de campos semânticos. A análise de campos semânticos usa os traços semânticos para pôr em evidência a relação entre diferentes itens lexicais dentro de um mesmo campo semântico. Do ponto de vista da L2, esta análise é de grande utilidade, na medida em que o facto de se partir de uma lista permite ao aprendente comparar, intra e interlinguisticamente, analisar e seleccionar um ou outro termo dentro do campo. No entanto, não parece muito provável

que faça uma análise decomposicional ao escolher um dos itens ³⁷. Mas pode ser de grande utilidade para o professor fazer notar distinções entre palavras.

Como reacção a Chomsky 1965, Gruber 1965 e Fillmore 1968 propõem que a entrada lexical deve contemplar traços sintácticos e relações semânticas ou temáticas, tais como: agente, instrumento, objectivo e tema (Brito 1998:381). Jackendoff 1972 e Lakoff 1971, na linha destes, e no âmbito da **Semântica Generativa**, abandonam a teoria componencial, defendem que a Semântica tem um papel central e tentam integrá-la na teoria sintáctica. Jackendoff 1972 defende que a entrada lexical deve conter traços sintácticos e **relações temáticas** e propõe que a projecção em Sintaxe não pode violar a '*Hierarquia Temática*': agente; locativo/origem/objectivo; tema (Brito 1998:384-5). Lakoff 1971, por seu lado, defende que a estrutura lexical e a estrutura sintáctica têm subjacentes estruturas de **primitivos semânticos** universais (Brito 1998:382-3). Na investigação que desenvolve para a definição desses primitivos, não segue a análise componencial, mas, de certo modo como Jackendoff porque também recorre à relação entre a linguagem e outras actividades cognitivas, sugere que a categorização é essencialmente uma questão de experiência e de imaginação humanas e que mudar o nosso conceito de uma categoria não significa exclusivamente uma mudança na nossa mente, mas também na nossa visão do mundo (Lakoff 1987:8). E, em 1983, também Jackendoff afirma que o desenvolvimento lexical não pode ser descrito exclusivamente como a aprendizagem de um conjunto de regras. Há um nível de representação linguístico distinto da estrutura sintáctica: o nível conceptual, que é constituído por um inventário de primitivos semânticos (tais como: EVENTO, ESTADO, COISAS, PROPRIEDADES, LUGARES, QUANTIDADES) e por regras de combinação (regras de boa formação semântica) (cf. Brito 1998:394). O pressuposto de base é que há um nível de

³⁷ A abordagem cognitiva reconhece a análise componencial como «um instrumento heurístico indispensável numa primeira fase da descrição semântica»; o que ela rejeita é «a tese de que a significação de um item lexical pode ser definida sob a forma de um conjunto de traços semânticos, sendo cada um deles 'necessário' para definir essa significação e todos juntos 'suficientes' para a delimitar de outra, e o princípio consequente da autonomia desse nível 'semântico' relativamente a outros dados qualificados de enciclopédicos» (Silva 1999:74).

representação mental em que informação linguística, sensorial e motora são compatíveis e sujeitas à mesma estrutura conceptual. Estudos de percepção visual mostraram que aquilo que percebemos com os nossos olhos é o resultado da interacção entre *input* fornecido pelo meio envolvente e um conjunto de princípios mentais activos que estruturam esse *input*. Jackendoff 1983 propõe que o significado das palavras está organizado segundo princípios semelhantes aos que operam no domínio da percepção. Ou seja, a expressão linguística resulta de entidades mentais que são projectadas no nosso conhecimento, e não da projecção directa dos objectos do mundo real. Esta abordagem evita muitos dos problemas que a análise componencial põe, nomeadamente permite a vaguidade (*fuzziness*) e a alteração (*revisability*) do sentido ³⁸. A informação lexical pode ser encarada como estruturada pela interacção de três condições: *necessidade*, *centralidade* e *tipicidade*. E as palavras «can differ widely in which kinds of conditions are most prominent» (Jackendoff 1983:121). Apoiar assim a organização do léxico num sistema de «preference rule». A informação dentro de uma entrada lexical está organizada em função desses três princípios, que interagem entre si, com diferente importância relativa conforme a palavra ³⁹.

O principal contributo para este progressivo reconhecimento da importância da representação lexical veio, como se pode ver, de semanticistas a trabalharem no quadro generativista, e resulta do debate quanto ao modo como os aspectos semânticos do léxico determinam a forma sintáctica. E se, em *Remarks on Nominalization* (1970), Chomsky começa a dar mais peso à informação lexical, é em *Lectures on Government and Binding* (1981) ⁴⁰ que a noção de papel temático é integrada na teoria e que a articulação entre o Léxico e a Sintaxe é assegurada pelo ‘Princípio de Projecção’:

³⁸ Para ilustrar a sua proposta de vaguidade, Jackendoff 1983 usa o exemplo clássico de Labov 1978 *chávena/tigela*.

³⁹ Para uma aplicação de Jackendoff 1983 a L2 ver Spolsky 1989, *Conditions for Second Language Learning*. Apresentação muito sucinta de Spolsky 1989 pode ser encontrada em Leiria 1991:38-43.

⁴⁰ Como é sabido, Chomsky dirá mais tarde que «‘Government-Binding (GB) Theory’ [is] a misleading term that should be abandoned» (cf. Chomsky 1989:70, n. 1) e propõe a designação: Gramática de Princípios e de Parâmetros.

As representações de cada nível sintáctico (i.e. FL, Estrutura-P e Estrutura-S) são projectadas a partir do Léxico, respeitando as propriedades de subcategorização dos itens lexicais (Chomsky 1981:29)

Ou seja, o falante sabe não só como se diz uma palavra e o que ela significa mas também o seu comportamento sintáctico, tendo a sintaxe de respeitar as características de cada item lexical. O 'Princípio de Projecção', juntamente com os outros princípios que o modelo de gramática prevê, são a estrutura básica da faculdade da linguagem. A Gramática Universal (GU) consiste num sistema de princípios comuns a todas as gramáticas e de parâmetros com valores diferentes consoante as línguas. Chomsky dirá mais tarde,

If this approach is correct [...] a language is not, then, a system of rules, but a set of specifications for parameters in an invariant system of principles of universal grammar (UG); and traditional grammatical constructions are perhaps best regarded as taxonomic epiphenomena, collections of structures with properties resulting from the interaction of fixed principles with parameters set on one or another way. There remains a derivative sense in which a language is a 'rule system' of a kind; namely, the rules of L are the principles of UG parametrized for L (Chomsky 1989:43).

Adquirir uma língua particular consiste em fixar valores para todos esses parâmetros e em adquirir entradas lexicais e as suas propriedades, a partir dos dados disponíveis; consiste em ir construindo um léxico mental. É assim que, a partir dos anos 80, se estabelece um diálogo ainda mais estreito entre a teoria da gramática e a aquisição de L1 e de L2, no âmbito da Gramática Generativa, mas também com a Psicolinguística e a Neurolinguística. Desde então, têm sido desenvolvidas, neste quadro, diversas teorias que fazem diferentes propostas quanto às fronteiras e à articulação entre o Léxico e a Sintaxe. Todas elas (Teoria de Regência e Ligação, Programa Minimalista, Gramática Lexical Funcional, GPSP) têm em comum o facto de abandonarem a ideia da autonomia absoluta da Sintaxe, e de considerarem que, pelo menos em parte, a estrutura da frase depende da semântica dos predicadores (Brito 1999:49). De modo que

Uma das propostas actualmente mais fortes é a de que as entradas lexicais deverão conter dois tipos de informações: a 'representação lexical sintáctica', 'estrutura de predicado-argumentos'

ou ‘*estrutura argumental*’ e a ‘*estrutura lexical conceptual ou semântica*’ (Brito 1999:33).

Esta posição, que resulta, muito em especial, do desenvolvimento das teorias da Semântica Lexical, vê o significado dos verbos (porque, de facto, é sobretudo do verbo que estas teorias se ocupam) como “uma estrutura multidimensional” (Brito 1998:417). Algumas delas (Grimshaw 1990, Pustejovsky 1991, Tenny 1994, Erteschik-Shir & Rappoport 1995) propõem que a dimensão aspectual dos itens lexicais, e a sua articulação com a estrutura argumental, deve ser tida em consideração. Para Pustejovsky 1991, por exemplo, é possível propor um número limitado de mecanismos generativos que construam as expressões semânticas. Segundo este autor, uma das preocupações centrais de uma teoria de Semântica Lexical é o papel do aspecto e dos tipos de ‘eventos’ (neles, inclui os estados). A descrição semântica deve incluir a interpretação aspectual de cada item lexical, do sintagma e da frase ⁴¹. A ‘estrutura de evento’ deve articular as propriedades aspectuais de cada item, o papel dos advérbios, o papel da estrutura argumental e a articulação do Léxico com a Sintaxe (Brito 1998:400) ⁴².

À chegada a posições como esta, que vêem o aspecto como um fenómeno de natureza composicional e que têm em consideração outros constituintes da frase, também não é, com certeza, alheia a experiência de Pustejovsky e de outros investigadores no campo da Linguística Computacional, enquanto ciência aplicada. A colaboração que tem existido entre linguistas teóricos e os investigadores em Inteligência Artificial tem tido a grande vantagem de permitir aos primeiros testarem as suas teorias

⁴¹ Em Leiria 1991:100, apoiei-me em Moens & Steedman 1988 e em Timbarlake 1982. Este tipologista vê a **estrutura semântica de um evento** como “o resultado de um certo número de níveis encaixados uns nos outros” (‘nested levels’). Distingue dois macroníveis: o lexical e o proposicional. Do lexical, fazem parte: a base (a semântica inerente à base lexical); o verbo (a base + certas operações semânticas; por exemplo, a marcação do início da acção ou a estativização); o predicado (o verbo + os seus argumentos sintácticos, incluindo o sujeito e os complementos). Do proposicional, a proposição (o predicado + a sua posição no espaço temporal e modal) e a narrativa (a proposição + a sua relação com outros eventos).

⁴² Até 1990, trabalhos sobre o **aspecto**, no âmbito da GG, não eram muitos e não estavam suficientemente divulgados. Em Brito 1998, podemos encontrar em “Relação léxico-sintaxe” um bom guia para este tema; Grimshaw 1990 afirmava: “a estrutura argumental e a dimensão aspectual dos itens lexicais devem estar representados no léxico”; em Pustejovsky 1991 lê-se: “uma das preocupações centrais de uma teoria de semântica lexical é o papel do aspecto e dos tipos de evento; uma descrição semântica deve incluir a interpretação aspectual de cada item lexical, do sintagma e da frase».

de modo a tornarem-nas operacionais. Nesse sentido, é ela também largamente responsável pelos recentes desenvolvimentos que têm aproximado a Sintaxe do Léxico no quadro generativo ⁴³. O *Programa Minimalista* (Chomsky 1995) minimiza os mecanismos gramaticais; vê a gramática das línguas como um sistema computacional que articula o Léxico e níveis de interface (a interface articulatória-perceptual e a conceptual-intencional); e propõe que esses dois níveis de interface são concebidos como “instruções” para os sistemas de “performance”, os sistemas de uso da língua (Brito 1999:41). Uma entrada lexical contém informação fonológica e semântica, e os traços de selecção categorial são derivados das propriedades semânticas, nomeadamente, dos papéis temáticos (Brito 1998:393-4).

Assim, nesta segunda fase, embora não haja «consenso quanto ao modo como o significado lexical determina a forma sintáctica, é comum a todos os modelos referidos a ideia de que o significado lexical determina (parcialmente) a Sintaxe» (Brito 1998:418) ⁴⁴. Ou seja,

In all mainstream generative theories (the Principles and Parameters approach, Lexical Functional Grammar, and Generalized Phrase Structure Grammar) the lexicon is of central importance, since they all assume that ‘clause structure is largely predictable from the semantics of predicates (Juffs 1996: 5)

Mas,

... a ‘dissidência’ representada pela Semântica Generativa anunciava de facto mudanças profundas. Enquanto Chomsky partiria, nos anos 70, para a edificação da Teoria Standard Alargada e , nos anos 80, para a Teoria da Regência e da Ligação, os subscritores da Semântica Generativa ajudariam a construir o que é hoje conhecido pela Gramática ou Linguística Cognitiva (Brito 1998:384).

Mas não só. As propostas de Chomsky, influentes e dinamizadoras, para além de proporcionarem o debate interno, e criarem cisões, também influenciaram abordagens (e foram influenciadas por elas) que se regem por

⁴³ «Many, if not most, theories used in computational linguistics today assume some degree of *functionality* in language, with corresponding structures at the syntactic and semantic levels. This assumption is so pervasive that it can be difficult to separate the syntactic and semantic descriptions» (Waterman 1996:147).

⁴⁴ Mas nenhuma teoria da GG aceita, como acontece com as teorias cognitivistas, de que falaremos em seguida, a **dependência da Sintaxe** em relação à Semântica. Brito 1999:47 apresenta alguns argumentos contra esta posição de dependência da Sintaxe em relação à Semântica.

princípios teóricos e metodológicos distintos. É o caso da abordagem funcional.

1.3. *A abordagem funcional*

Durante o século XX, o termo ‘funcional’ foi usado para qualificar práticas metalinguísticas diversas. Usaram-no os membros do Círculo Linguístico de Praga, Hjelmslev e os seus seguidores, a *Société Internationale de Linguistique Fonctionnelle*, e, mais recentemente, linguistas, como Simon Dik, influenciados por teorias, pelo menos à partida, generativistas (Martinet, 1995:290).

Consequentemente, e como seria de esperar, há muitas diferenças entre as várias abordagens que se dizem ‘funcionais’ ou ‘funcionalistas’. No entanto, todas assumem que a linguagem é uma actividade cognitiva e social. A investigação funcional tem como objectivo clarificar a relação entre forma e função, e determinar a natureza das funções que influenciam a estrutura gramatical. O facto de assumirem que a gramática de uma língua é regulada por considerações comunicativas e cognitivas tem como consequência prática imediata não terem como objectivo a construção de modelos formais (cf. Thompson 1992, vol. 2:37) ⁴⁵.

Simon Dik 1968, na linha de Fillmore, critica aquilo que considera serem alguns pontos fracos da primeira fase de Chomsky, nomeadamente a autonomia da Sintaxe, pois considera que alguns fenómenos gramaticais requerem explicações semânticas. Dez anos depois, Dik 1978 propõe um novo modelo linguístico, a que chamou *Functional Grammar*. Nesse modelo, estavam já definidas as três grandes linhas que iriam guiar as posteriores investigações nesse mesmo quadro: a exclusão da componente transformacional, a definição de um único nível de representação e a negação da autonomia da sintaxe. A partir de então, a Gramática Funcional tem-se desenvolvido no sentido de validar os critérios pragmáticos, psicológicos e tipológicos definidos em Dik 1989. Martín Arista define-a assim:

⁴⁵ «...‘functionalists’ have not been involved in constructing formal models of grammar. Thus, some approaches to grammar with the term ‘functional’ in their labels are actually formal models, and not share the assumptions just outlined» (Thompson 1992, vol 2: 37).

... la Gramática Funcional (GF) de Dik (1989, 1997b) es una teoría lingüística de base léxica en la que las relaciones de tipo semántico, sintáctico y pragmático sustentan un conjunto de condiciones de gramaticalidad que garantizan la validez comunicativa de las expresiones lingüísticas de las lenguas naturales. Dicho de otro modo, los marcos predicativos de la teoría son el punto de partida de la expansión que da lugar a los enunciados (Martín Arista 1999:172).

Na sequência dos trabalhos de Dik, têm surgido várias teorias funcionais. Martín Arista identifica como teoria funcional qualquer uma que cumpra os seguintes critérios: **(1)** tem em conta que o fim último da linguagem humana é a interação social por meio da comunicação; **(2)** presta especial atenção às funções sintáticas, semânticas e pragmáticas e, acima de tudo, às relações entre elas; e **(3)** assume a separação entre o léxico e a gramática (Martín Arista 1999:14-15) ⁴⁶.

Assim, e apesar de reconhecer grandes diferenças entre elas, Martín Arista considera como gramáticas funcionais, na tradição do funcionalismo norte-americano, a *Role and Reference Grammar* de Foley e Van Valin 1980 e os trabalhos de Givón no âmbito da *gramática funcional tipológica* (entre outros, 1979 e 1985); e, na tradição do funcionalismo europeu, a *Gramática Sistémica* de Halliday (1985) e a *Gramática Funcional* de Simon Dik (1978 e 1989). Todas elas assentam no pressuposto de que, a partir da informação contida no léxico, é possível predizer o comportamento sintático dos predicados.

No modelo da *Gramática Funcional*, o ponto de partida para a análise de uma expressão linguística é a construção da predicação nuclear. Para a construção das predicações nucleares de uma língua, é preciso dispor de um conjunto de termos e de predicados dessa língua. A essa informação dá-se o nome de *Fundo* e compõe-se de **(1)** um *léxico*, que contém os predicados (categorias lexicais) e termos (categorias sintáticas) básicos, ou

⁴⁶ Pela aplicação do segundo critério, e apesar do nome, o autor não considera a *Gramática Funcional Lexical* de Bresnan como uma teoria funcional, já que é um modelo sintacticista, no qual «el componente semántico-pragmático es más bien periférico»; mas, atendendo a este mesmo critério, considera como tal a *Role and Reference Grammar* [Foley e Van Valin 1980, 1984]. Pela aplicação do terceiro critério, e «aunque las propuestas de la Lingüística Cognitiva son, en muchos aspectos, compatibles con las del funcionalismo, no se he tenido en cuenta los modelos cognitivos por su tendencia a identificar la gramática con el léxico». Pelo contrário, a *Gramática Funcional Tipológica* é considerada como funcional porque «aunque aboga por una progresiva disolución de la sintaxis en la semántica y en la pragmática cumple el segundo criterio, es decir, la consideración del texto desde la triple perspectiva sintáctica, semántica y pragmática y de la relación entre las funciones que se reconocen en estos ámbitos.» (Martín Arista 1999:14-5).

seja, os que não são derivados por meio de regras sincronicamente produtivas, **(2)** um conjunto de *regras de formação de predicados* que geram os predicados derivados; e **(3)** um conjunto de *regras de termos* que produzem os termos derivados (Mairal Usón 1999:45). Sendo assim, na descrição das entradas lexicais não se especifica informação morfosintáctica redundante (como, segundo Dik, acontecia nos primeiros modelos de Chomsky). Essa informação pode predizer-se a partir da informação armazenada no léxico.

Do sistema de notação semântica que dá conta das propriedades combinatórias e do significado de cada uma das peças lexicais que constituem **o léxico**, fazem parte: **(1)** a categoria sintáctica (N, V, A, ADV, etc.); **(2)** a valência quantitativa, que especifica o número de argumentos subcategorizados por um determinado predicado; **(3)** a valência qualitativa, que especifica a função semântica dos argumentos do estado de coisas estabelecido pela predicação; e **(4)** as restrições de selecção, que especificam a natureza dos argumentos que podem aparecer na posição de sujeito e de complemento. Alguns modelos incluem também na entrada lexical (5) uma representação do significado de cada um dos predicados (Mairal Usón 1999:46-65).

As *regras de formação de predicados*, que são outra das componentes do *Fundo*, ocupam-se do **léxico derivado**. Essas regras podem actuar sobre qualquer um dos parâmetros inscritos no léxico. Dos efeitos da aplicação dessas regras, pode resultar, por exemplo, uma redução ou uma extensão da valência quantitativa, uma mudança da função semântica dos argumentos do predicado, ou uma mudança nas restrições de selecção do predicado. Assim sendo, não há no modelo um mecanismo que trate de forma autónoma os fenómenos morfológicos. Dois tipos de regras, de expressão e de formação de predicados, dão conta, respectivamente, dos processos flexionais e dos derivacionais. Deste modo, as formações derivacionais são tratadas do mesmo modo que fenómenos como a causativização perifrástica ou a incoativização (Mairal Usón 1999:66)⁴⁷.

⁴⁷ No que respeita os modelos generativistas, tanto para a *Lexical Functional Grammar* como para a Teoria da Regência e Ligação, o Programa Minimalista e a Teoria da Optimidade, é comum a consciência da importância da morfologia flexional. Mas não há convergência quanto ao nível a que certas propriedades morfológicas devem ser tratadas. Uns tendem a vê-las operando no Léxico, outros na Sintaxe, e outros na Forma Lógica. (Brito 1999:48).

Muito resumidamente, são estes, segundo Mairal Usón (1999), alguns dos aspectos comuns mais relevantes dos modelos funcionais. Para o mesmo autor, a componente léxico deve, no entanto, ser enriquecida, já que os actuais parâmetros não são tão exaustivos que dêem conta do conhecimento lexical de um falante (Mairal Usón 1999:67). Por isso, faz várias sugestões para o seu enriquecimento. Destaco duas, de natureza diferente: o recurso à linguística de *corpora* e uma descrição onomasiológica do léxico de modo a agrupar os predicados em classes ou domínios semânticos.

Para exemplificar como pode ser enriquecido o léxico com descrição onomasiológica ⁴⁸, apresenta o *Modelo Lexemático Funcional* (MLF) desenvolvido por Martín Mingorance em variadíssimos trabalhos (por exemplo, 1984, 1987, 1995). Um dos primeiros objectivos deste modelo foi propor uma organização onomasiológica em que os lexemas aparecessem classificados em domínios e subdomínios, o que permite, por um lado, relacionar semântica e sintaxe, e por outro, criar uma interface entre a estrutura linguística e a conceptual (Mairal Usón 1999:70). Uma entrada lexical é encarada como um discurso em que confluem factores linguísticos, textuais, culturais e antropológicos. O léxico é representado com todas as suas propriedades sintácticas, morfológicas, semânticas e pragmáticas. Obtém-se assim uma base de dados que dá conta das relações sintagmáticas e paradigmáticas dos lexemas e que, ao contrário de um dicionário, reflecte o modo como organizamos o nosso léxico mental (Cf. Aitchison 1987) ⁴⁹. Essa base de dados, ou seja, o estabelecimento da arquitectura semântica de uma língua, implica três tipos de análise: **(1)** a estruturação de cada domínio lexical em subdomínios; **(2)** o estabelecimento de hierarquias de significado e análise dos parâmetros

⁴⁸ ONOMASIOLOGIA: estudo do significado que consiste em partir de uma noção (ou de um conceito) e procurar o (ou os) signo(s) linguístico(s) que dão conta dessa mesma noção. SEMASIOLOGIA: metodologia de carácter semântico que parte do signo para analisar aquilo que ele designa.

⁴⁹ O modelo adopta uma versão enriquecida da Lexemática de Coseriu e integra-a no léxico da GF de Dik (Cf. Dik 1978 e Coseriu 1981) (ver Mairal Usón 1999:73).

semânticos de cada subdomínio; e **(3)** o estabelecimento dos parâmetros sintáctico-semânticos (Mairal Usón 1999:78) ⁵⁰.

POSSESION (tener)	<ol style="list-style-type: none"> 1. Tener algo (<i>poseer</i>) <ol style="list-style-type: none"> 1.1. Tener algo en las manos [<i>coger1</i>] 1.2. Llegar a tener algo [<i>obtener</i>] <ol style="list-style-type: none"> 1.2.1. Obtener dinero [<i>ganar, extorsionar</i>] 1.2.2. Obtener algo mediante dinero [<i>comprar, adquirir</i>] 1.3. Llegar a tener un número de personas o cosas juntas [<i>reunir</i>] 1.4. Llegar a tener algo en las manos [<i>coger2</i>] <ol style="list-style-type: none"> 1.4.1. Coger algo con violencia [<i>capturar</i>] <ol style="list-style-type: none"> 1.4.1.1. Capturar a alguien [<i>apresar</i>] 1.4.1.2. Capturar un animal [<i>cazar</i>] 1.5. Volver a tener algo [<i>recuperar</i>] 1.6. Seguir teniendo algo [<i>mantener1</i>] 1.7. Tener algo dentro [<i>contener</i>] 1.8. No tener algo [<i>carecer</i>] 1.9. Dejar de tener algo [<i>perder</i>] 2. Hacer que alguien o algo tenga [<i>dar</i>] 3. Hacer que alguien tenga una cosa..... ..
----------------------	---

Mairal Usón exemplifica as suas ideas com a arquitectura do domínio lexical de POSSE, que reproduzo parcialmente. Apresentei aqui este desenvolvimento da teoria pelo interesse que pode vir a ter para os investigadores de L2, já que não só parece reflectir certos aspectos da organização cognitivo-conceptual do léxico como permite a comparação entre duas ou mais línguas. E, nesse sentido, ela apresenta uma forte inspiração tanto da abordagem cognitiva como do funcionalismo tipológico de Givón.

⁵⁰ O MLF distingue os seguintes domínios semânticos: percepção física, percepção mental, actos de fala, existência, posse, mudança, sentimento, acção e movimento (cf. Mairal Usón 1999:78).

1.3.1. A abordagem funcional tipológica

A abordagem funcional tipológica ⁵¹, iniciada por Humbolt em 1836 e retomada por Greenberg em 1966, tem sido uma das perspectivas mais dinâmicas dentro da abordagem funcional. Por ter como objectivo não só estabelecer tipologias linguísticas mas também encontrar universais, tem sido designada por ‘universais e tipologia’. E, pelo seu carácter comparativista, é a que mais tem influenciado os estudos de L2. Destacam-se, por isso, da sua fase mais recente, os trabalhos de Bybee 1985, Comrie 1976 e 1981, Dahl 1985, Dixon 1977, Givón 1979 e 1984, Halliday 1985, e Talmy 1985⁵².

Talmy Givón 1979 contribuiu de forma decisiva para a reflexão sobre os fenómenos de **lexicalização** e de **gramaticalização** ⁵³, ao assumir que sincronia e diacronia não são realidades estanques e ao relacioná-las com a aquisição e uso da linguagem:

[L]anguage – within the minds of speakers, rather than as some abstract system of langue – is always in the middle of change in lexicon/meaning, syntax, morphology and phonology. Language as a cognitive map is thus not only a system of coding knowledge, but perhaps primarily a system of re-coding, modifying and re-structuring existing knowledge and integrating into it newly-acquired knowledge (Givón 1982:112).

Givón 1985 propõe que todas as construções sintácticas complexas surgem ontogenética, diacrónica e, com certeza, também filogeneticamente

⁵¹ Segundo Croft 1990, o termo tipologia é usado com diferentes valores, o que conduz muitas vezes a alguma confusão quanto ao seu conteúdo. (1) Taxinomia ou classificação; este conduz a um sentido (2) classificação tipológica das línguas em função da sua morfologia; este introduz (3) a tipologia propriamente dita: o estudo de características que são encontradas por comparação linguística (cf. Greenberg 1966 e *universais implicacionais unidireccionais*).

⁵² Em Leiria 1991 (capítulo 2) reflecti sobre o apoio teórico que esta abordagem, combinada com os universais de processamento (Slobin 1973 a 1990), tem fornecido aos estudos de L2. Discuti também a relação entre *universais linguísticos* e *universais cognitivos*; ou seja, entre a teoria de Givón e o *método interlinguístico* desenvolvido por Slobin (1973 e 1985), Peters (1985) e outros.

⁵³ «**Grammaticalization** is the process by which full lexical items become grammatical morphemes» (Croft 1990: 230). Ver também, em 1.4 *Abordagem Cognitiva*, a definição de Silva 1996:14. «The most important functional grammaticalization process is called **desemanticization** by Heine and Reh. This is the process by which lexical meanings shift to the meanings of grammatical items. [...] The term ‘desemanticization’ implies that the process involves a loss of meaning.» (Croft 1990: 236).

como resultado da sintactização de construções paratáticas do modo pragmático. Este, apesar de, do ponto de vista do seu processamento, ser mais lento, tem, sobre o modo sintático, a vantagem de ser mais transparente; o modo sintático, em contrapartida, é mais rápido, mas exibe uma menor e mais complexa correlação entre código e mensagem. Todas as línguas conhecidas têm, afirma Givón, os dois extremos, para além de uma série de modos intermédios. As crianças adquirem primeiro o modo pragmático e vão-no gradualmente sintactizando; os adultos dispõem de um leque de modos, ancorados no modo pragmático, que usam apropriadamente segundo as condições (Givón 1985:1018). Abre assim espaço para o estudo e análise sistemática dos *pidgins*, dos crioulos e de gramáticas intermédias, porque, se a nível sintático as línguas divergem enormemente, a nível do seu modo pragmático elas evidenciam espantosas semelhanças (Givón 1979:107).

A componente psicolinguística desta teoria, como veremos no Capítulo 3, defende que a criança, para desenvolver o seu conhecimento linguístico, usa princípios de atenção e de armazenamento, no sentido de extrair e segmentar pequenas parcelas – *chunks* – do *input* disponível (cf. Slobin 1973, 1985; Peters 1985). Esses processos de extracção e segmentação obedecem a princípios de natureza linguística e psicolinguística, tais como *frequência*, *saliência*, *relevância* e *marcação*.

Seguindo Givón 1979, Bybee 1985 defende, na sua obra *Morphology*, que a relação entre forma e função não é completamente arbitrária e não estabelece uma linha de fronteira entre léxico, morfologia e sintaxe. Em vez disso, propõe um *continuum* linguístico que vai da expressão lexical à sintáctica e fá-lo depender de *saliência* cognitiva e cultural:

A meaning element is relevant to another meaning if the semantic content of the first directly affects or modifies the semantic content of the second. If two meaning elements are, by their content, highly relevant to one another, then it is predicted that they may have lexical or inflectional expression, but if they are irrelevant to one another, then their combination will be restricted to syntactic expression. [...]
Relevance depends on cognitive and cultural salience: no matter to what extent an entity, event or quality is decomposable into semantic features, if it is perceived as discrete from surrounding entities, events or qualities, it can have a lexical item applied to it. So two semantic elements are highly relevant to one another if the result of their combination names something that has high cultural or cognitive salience. **If cognition shapes both language and culture,**

and cultural and linguistic variation is patterned and constrained by cognition, then we would expect lexicalization patterns to differ across languages, but in a principled and patterned way (Bybee 1985:13).

Faz assim depender as línguas, as suas gramáticas e os seus padrões de lexicalização de aspectos pragmáticos, culturais e cognitivos.⁵⁴ Ou seja, a variação intralinguística e a interlinguística têm uma motivação funcional e estão relacionadas com processos individuais e colectivos.

Também Talmy, que centra a sua investigação no verbo, observa que as línguas são comparáveis quanto à informação que veiculam, mas que elas diferem bastante quanto ao modo como o fazem (Talmy 1985:57). Em conformidade, assume que é possível identificar separadamente elementos no domínio do significado (tais como 'Motion', 'Path', 'Figure', 'Ground', 'Manner' e 'Cause') e elementos da estrutura de superfície (tais como 'verb', 'adposition', 'subordinate clause' e 'satellite'). E também ele, como Bybee, assume que o modo como as línguas o fazem é regido pelo conceito de saliência (Talmy 1985:122). Ou seja, enquanto que, numa dada língua, uma componente do significado, devido ao tipo de representação linguística, é posta em destaque, noutra língua pode ser-lhe prestada menos atenção. Assim, **os padrões de lexicalização característicos⁵⁵ variam consoante as línguas de modo consistente e regular em cada uma delas ou dentro de um grupo tipológico. Esses padrões dependem dos elementos de significado que se fundem ('conflate')⁵⁶ numa mesma raiz verbal ou que são confiados a outros elementos da estrutura de superfície.**

Dentro da classificação de Talmy, destacam-se os **verbos de movimento**⁵⁷. Segundo o autor, o Chinês⁵⁸ e todas as línguas indo-

⁵⁴ «[T]he **relevance** of the meaning of the affix to the stem determines the degree of fusion between affix and stem. Further, the relevance of the morphological notion to the stem determines how cognitively distinguishable or discrete the resulting concept is.» (Bybee 1985:208).

⁵⁵ «Here, 'characteristic' means that: (i) It is colloquial in style, rather than literary, stilted, etc. (ii) It is frequent in occurrence in speech, rather than only occasional. (iii) It is pervasive rather than limited, that is, a wide range of semantic notions are expressed in this type» (Talmy 1985:62).

⁵⁶ «One terminological note: we will refer to the meaning-in-form relation with three terms. They are 'lexicalization' from McCawley (e.g. 1968); 'incorporation' as used by Gruber (1965); and 'conflation', a term that was coined for the purpose by the author (Talmy 1972) and that has now gained general currency. These terms have different emphases and connotations [...] but all refer to the representation of meanings in surface forms» (Talmy 1985:60).

⁵⁷ Batoréo 2000 apresenta detalhadamente a teoria de Talmy.

européias, excepto as línguas românicas, seguem o mesmo padrão. Assim, enquanto o Inglês, que Talmy considera como o exemplo perfeito do primeiro grupo, pode codificar, «with its particular verb-conflation pattern and its multiple satellite capability», numa mesma frase quatro componentes ('Motion', 'Path', 'Manner' e 'Cause'), o Espanhol, exemplo perfeito do segundo grupo, «with its different verb-conflation pattern and almost no productive satellites, can background only one of the four English components, using its main verb for the purpose; any other expressed component is forced into the foreground in a gerundive or prepositional phrase.» Talmy ilustra com os seguintes exemplos:

- (1) The man ran back down into the cellar
- (2a) El hombre corrió al sótano
- (2b) El hombre volvió al sótano corriendo
- (2c) El hombre bajó al sótano corriendo
- (2d) El hombre entró al sótano corriendo

De facto, o Inglês consegue codificar, numa frase coloquial e frequente, informação que o Espanhol, tal como o Português, para respeitar os mesmos critérios de aceitabilidade, tem que omitir, confiar à inferência ou a outros momentos do discurso (Talmy 1985:122-123).

Segundo Ake Viberg (1993:373) que, no âmbito da tipologia linguística, tem desenvolvido um projecto lexicológico contrastivo (*Contrastive Lexicology Project*)⁵⁹, este diferente padrão de lexicalização dos verbos protagonizado pelas línguas românicas representa, provavelmente, a mais abrangente divisão tipológica entre as línguas da Europa e tem, certamente, reflexos notórios ao nível da aquisição. O objectivo do seu projecto consiste em estudar a organização e aquisição do **léxico básico** numa perspectiva interlinguística (mas também cognitiva). As referências teóricas do seu trabalho sobre a estrutura lexical são a análise dos campos semânticos, a marcação tipológica e as hierarquias de lexicalização. Assume, seguindo Miller e Johnson-Laird 1976, que um campo semântico se organiza em torno de um conceito nuclear (*core concept*).

⁵⁸ Para opinião ligeiramente diferente quanto a 'conflation' em Chinês, ver Juffs 1996.

⁵⁹ Ver também Viberg 1984, 1986, 1993 b, e 1993c.

A partir de estudos de frequência em onze línguas da Europa ⁶⁰, concluiu que uma característica tipológica dessas línguas, ao contrário de outros grupos que têm uma classe de adjetivos e de advérbios muito reduzida, consiste em cada uma das classes abertas (N, V, ADJ, ADV) ter um conjunto de características próprias que a individualiza e de ser composta por um grande número de membros. Concluiu também que um pequeno número de campos semânticos cobre a maior parte das palavras mais frequentes de cada uma das classes gramaticais. E concluiu ainda que, além disso, há muitas semelhanças entre essas línguas quanto ao significado básico dos membros mais frequentes de cada um desses campos. Muitos desses significados são tipologicamente não marcados e ocupam lugares cimeiros nas hierarquias universais de lexicalização (Viberg 1993a:341-2).

Viberg concentra-se também nos **verbos**. Dentro de cada campo semântico, é possível, afirma, isolar um ou alguns sentidos que lexicalizam como verbos nucleares num certo número de línguas. **Define verbos nucleares como os tipologicamente menos marcados dentro do respectivo campo semântico** (Viberg 1993a:347); por exemplo, ‘go’ para o campo ‘movimento’, ‘give’ e ‘take’ para ‘posse’, ‘see’ para ‘percepção’ e ‘say’ para ‘comunicação verbal’ ⁶¹. Os verbos nucleares combinam algumas características não marcadas, tais como: alta frequência, prototipicidade sintáctica, e domínio semântico em hierarquias de polissemia ⁶². Termos tipologicamente não marcados, como é o caso dos verbos nucleares, constituem um subconjunto das palavras básicas de qualquer língua. Se a estes verbos nucleares acrescentarmos algum ou alguns outros específicos duma língua (ou do grupo de línguas a que ela pertence) teremos os **verbos básicos** dessa língua. Assim, Viberg propõe um conjunto de 13 verbos básicos nas 11 línguas analisadas. Desses, 9 verbos reflectem tendências universais, são os verbos nucleares (*go, give, make, say, see, know, want,*

⁶⁰ As 11 línguas são: Inglês, Alemão e Sueco; Francês, Espanhol, Italiano e Romeno; Russo, Polaco; Finlandês e Húngaro.

⁶¹ Usa verbos ingleses porque «many of the most basic characteristics of English lexical structure are typical of European language in general»(Viberg 1993:341).

⁶² «Verbs in general, and especially the most frequent verbs, tend to belong to the most polysemous elements in a language.» (Viberg 1993 a: 349)

come, take), a que acrescenta outros 4 verbos básicos: 3 específicos da área europeia (*be, can, must*) e um específico de sub-áreas (*have*)⁶³.

Os verbos nucleares de Viberg podem ser equiparados à entrada lexical de um domínio de Martín Migorance de que falámos antes. Assim, enquanto que a perspectiva deste privilegia, à partida, a representação **paradigmática** do léxico, a de Viberg privilegia a representação **sintagmática**.

Como veremos no Capítulo 3, muitos investigadores têm reconhecido à tipologia linguística e ao conceito de marcação potencialidades para a formulação de hipóteses sobre a aquisição de L2. Se cruzarmos estas duas perspectivas, sem esquecermos a abordagem de Talmy para os padrões de lexicalização dos verbos, obteremos um quadro interessante para abordar estudos de léxico por falantes de L2 em dois possíveis estádios de aquisição: as abordagens de Viberg e de Talmy num primeiro estádio, até ser atingido um nível básico de competência e de correcção morfossintáctica; e um segundo estádio, com a proposta de Martín Mingorance, em que haveria uma expansão do léxico e uma aplicação das estruturas prototípicas. E este seria um dos caminhos possíveis para que a reconhecida centralidade do léxico contribuísse para um avanço significativo dos conhecimentos sobre sequências de aquisição em L2⁶⁴.

1.4. *A abordagem cognitiva*

Qualquer uma das duas abordagens consideradas até ao momento tem uma perspectiva cognitiva, na medida em que ambas se ocupam da faculdade da linguagem enquanto atributo humano. No entanto, há um grupo de investigadores que, nos últimos vinte anos, tem desenvolvido uma abordagem para a qual tem reivindicado a designação de **Linguística**

⁶³ «The most frequent verb in practically all European languages is the copula **BE**, which clearly is such an areal-specific characteristic. [...] The second most frequent verb in a large group of languages in the centre of Europe is **HAVE**. But actually, this verb is completely missing from the Celtic languages at the western fringe of Europe and from languages to the east such as Finno-Ugrian languages. As is well known, relatively few languages outside Europe have developed an equivalent which makes the presence or absence of HAVE in a large group of European languages an important sub-areal characteristic.» (Viberg 1993:348-9). Odlin 1989:76 exemplifica com o Irlandês. A tradução mais aproximada de *I have money* é *Tá airgead agam / Is money at me*.

⁶⁴ Ver, em Meara (1993:293-4), relação entre dimensão e estrutura interna do léxico; entre 'plateaux' e reestruturação do léxico. Ver também, a seguir, Goldberg 1995.

Cognitiva ⁶⁵. Augusto Soares da Silva (1999:11) diz que Gibbs responde assim à questão ‘What’ s cognitive about cognitive linguistics?’:

Cognitive linguistics is especially deserving of the term cognitive not solely because of its commitment to incorporating a wide range of data from other cognitive disciplines, but because it (a) actively seeks correspondences between conceptual thought, bodily experience, and linguistic structure, and (b) because it seeks to discover the actual contents of human cognition (Gibbs 1996:49).

A Linguística Cognitiva representa, em alguns aspectos, uma linha de continuidade e, em outros aspectos, uma ruptura em relação a certas propostas de outras abordagens e modelos linguísticos. Ela integra muita da investigação realizada pela tipologia funcional, mas insurge-se contra o modelo chomskiano, no sentido em que o conhecimento linguístico não pode ser considerado isoladamente, sem ter em conta as outras capacidades cognitivas. E o facto de a linguagem ser um domínio que faz parte da cognição humana, e estar relacionada com outros domínios cognitivos, aconselha investigação interdisciplinar. Por isso, ela insere-se no quadro da **Ciência Cognitiva**, na medida em que abarca conhecimentos de várias áreas, de que se destacam a Filosofia, a Psicologia Cognitiva, a Neurociência, a Inteligência Artificial, a Antropologia, e, claro, a Linguística. Ana Maria Brito (1998:384) afirmava antes que «os subscritores da Semântica Gerativa ajudariam a construir o que é hoje conhecido pela Gramática ou Linguística Cognitiva». Isto significa que, enquanto no âmbito da Teoria Standard, Katz 1963 e 1975 desenvolve uma teoria semântica baseada na ideia da composicionalidade do significado, como vimos, a Semântica Gerativa, pela mão, entre outros, de Jackendoff e de Lakoff, passa a atribuir à semântica, e não à sintaxe, o papel central. Para a Semântica Gerativa, a sintaxe não é arbitrária, mas motivada pela semântica. É nesta linha de pensamento que se desenvolve a **Semântica Cognitiva**.

Vimos que Lakoff 1971 sugere que a categorização é essencialmente uma questão de experiência e de imaginação humanas e que mudar o nosso

⁶⁵ «Mas há uma diferença essencial: A Gramática Gerativa interessa-se pelo conhecimento *da* linguagem (tomando-a, portanto, não como meio, mas como objecto da relação epistemológica) e procura saber como é esse conhecimento é adquirido, ao passo que a Linguística Cognitiva interessa-se pelo conhecimento *através* da linguagem e procura saber como é que a linguagem contribui para o conhecimento do mundo.» (Silva 1999:12, nota 4).

conceito de uma categoria não significa exclusivamente uma mudança na nossa mente, mas também na nossa visão do mundo (Lakoff 1987:8). Esta filosofia tem por detrás a **teoria do protótipo** de Eleanor Rosch (1973, 1975, 1978). Esta estabelece que em geral qualquer categoria tem um membro mais ‘típico’, um exemplar ‘ideal’ (Aitchison 1994:55), com um conjunto ideal de características, em relação ao qual todos os candidatos a serem incluídos na mesma categoria são comparados. Por exemplo, *oliveira* é mais prototípico da categoria *árvore* do que *palmeira*. É assim para um português, mas, para um africano, talvez seja o contrário. Isto significa que a categorização, mesmo de entidades concretas, pode depender da saliência cultural ⁶⁶. Podemos então considerar que alguns modelos mentais podem ser (relativamente) universais enquanto que outros são culturalmente específicos ⁶⁷. Sendo assim, para os cognitivistas não há lugar para a distinção, como propõem os estruturalistas e os generativistas, entre conhecimento semântico, enquanto conhecimento linguístico associado às formas linguísticas, e conhecimento enciclopédico resultante da experiência individual, social e cultural do falante.

Mas a maior parte das categorias não são de objectos, mas de entidades abstractas (Lakoff 1987:6). Entidades abstractas envolvem conceitos mais complexos. Ou seja, as categorias organizam-se radialmente à volta de membros centrais de uma categoria, ou membros prototípicos, que cumprem um máximo de propriedades associadas a essa categoria ⁶⁸. As categorias mais complexas são originadas a partir das mais simples: padrões metafóricos e metonímicos operam sobre esquemas pré-linguísticos baseados na experiência sensorial do falante, o que permite elaborar conceitos complexos a partir de conceitos simples. A **metáfora** e

⁶⁶ Ver Bybee 1985:208, citada em *Abordagem Tipológica Funcional* : «Relevance depends on cognitive and cultural salience.»

⁶⁷ Ou podem até depender da **idade**: «When asked to select the ‘best examples’ of a category, English 11-14-year-olds differed from adults in a number of ways. The youngest children did not make any consistent selection, and older ones tended to give a top ranking to items which were important to them. English children tended to give high priority to potatoes on the vegetable list [...] and to parrots on the bird list..» (Aitchison 1996:18).

⁶⁸ Silva 1999:41-44 diz que, no que respeita a representação da estrutura semasiológica de itens lexicais, para além do *modelo radial*, popularizado por Lakoff 1987, outros dois têm sido propostos: o *modelo de grupos em sobreposição* de Geeraerts (por exemplo, Geeraerts 1989) e o *‘network’* de Langacker 1988.

metonímia surgem assim como processos universais na linguagem, interpretáveis pelos mesmos processos que o sentido literal. «Metáfora e metonímia constituem dois importantes processos cognitivos, duas importantes estratégias de conceptualização. A nível lexical, elas **estão na base da extensão semântica dos itens lexicais (isto é, do alargamento de um mesmo item a outros significados ou aplicações semânticas)** e do processo de mudança semântica» e, por isso mesmo, ocupam «um lugar central na abordagem cognitiva da significação lexical» (Silva 1999:44-5). Porque se apoiam na experiência humana, na experiência física, elas são estruturadas com base em um número limitado de esquemas imagéticos (*image schemas*) que podem ser vistos como estruturas cognitivas pré-linguísticas universais ⁶⁹. Assim, é um determinado **esquema imagético**, dominante numa comunidade, que determina, em função da sua saliência para essa comunidade, o grau de prototipicidade de um dado conceito e que, por conseguinte, explica os diferentes padrões e hierarquias de lexicalização, e conseqüentemente a **polissemia** ⁷⁰, específicos dessa língua. E é deste modo que muito do importantíssimo trabalho sobre a relação entre conceitos e sua lexicalização, desenvolvido no âmbito do funcionalismo tipológico (por Anderson, Bybee, Comrie, Givón, e Talmy, entre outros), de que me limitei a dar apenas uma pálida imagem no ponto

⁶⁹ «**Image schemas** provide the structures used in those [complex cognitive] models. [...] These schemas [CONTAINER, SOURCE-PATH-GOAL, LINK, PART-WHOLE, CENTER-PERIPHERY, UP-DOWN, FRONT-BACK] structure our experience of space. What I will be claiming is that the same schemas structure concepts themselves.» (Lakoff 1987: 282-3). «Thus, what has been called abstract reason has a bodily basis in our everyday physical functioning.» (Lakoff 1987: 278). Batoréo 2000:365 adverte: «É, sem dúvida, em Talmy [por exemplo, 1983] que se devem procurar as raízes do *Esquema Imagético* (= 'image schema'), largamente operacionalizado e discutido por Lakoff, Johnson e outros autores, sem, em princípio, reconhecer a 'paternidade' talmiana.»

⁷⁰ «It is common for a single word to have more than one meaning. In some cases the meaning are unrelated, like the two meanings of *bank* – the place where you put your money and the land along the edge of a river. In such cases, there is not one word but two. They are called instances of **homonymy**, where two words with two totally different meanings happen to be pronounced the same way. In other cases, the senses are related, often in such a close and systematic way that we don't notice at first that more than one sense exists at all. Take the word *window*, for example. It can refer either to an opening in a wall or to the glass-filled frame in that opening. Or take the word *open*. We open doors and we open presents, and though the actions described by the words are very different, we would normally have to think twice to notice the difference. Or the word *run*. It is very different for Harry to run into the woods and for the road to run into the woods.[...] Such cases are called instances of **polissemia**. They are cases where there is one lexical item with a family of related senses.» (Lakoff 1987:416).

anterior, é relido e integrado na Semântica Cognitiva. Lakoff resume deste modo o efeito da prototipicidade:

Linguistic categories, like conceptual categories, show prototype effects. Such effects occur at every level of language, from phonology to syntax to the lexicon. I take the existence of such effects as prima facie evidence that linguistic categories have the same character as other conceptual categories (Lakoff 1987:67).

A teoria do protótipo estende-se, portanto, não só à metáfora mas também a entidades não individuais: os **eventos** (Fillmore 1976, Schank & Abelson 1977) ⁷¹. Imaginemos um. Por exemplo, ‘comer fora’. Uma ida ao restaurante envolve um *script*, isto é, uma sequência estereotipada de eventos, constituído por caminho para o restaurante, encomendar, comer, beber, conversar e pagar; envolve um *frame*, um contexto espacial que é o restaurante: mesas, cadeiras, guardanapos e flores; e certos *schemata*: a bebida, por exemplo, neste contexto, é alcoólica – vinho ou cerveja – mas no contexto de uma festa de crianças não é. Este exemplo é de Singleton (1999:35-7). Ao *script* deste autor, eu acrescentei ‘conversar’, porque a minha representação de ida ao restaurante inclui sempre outra ou outras pessoas; de *frame* retirei ‘velas’, porque, geralmente, nos restaurantes portugueses, a que costumo ir, elas não existem; incluirei ‘velas’ se pensar num jantar na Suécia; do mesmo modo que, se eu fosse muçulmana, ‘bebida’, para mim, seria chá ou qualquer refresco.

Por outro lado, são estes modelos mentais, estes *scripts* ou guiões, que guiam a atribuição a uma dada expressão linguística, metafórica ou metonímica, de uma interpretação mais literal ou mais idiomática. Para essa interpretação, a expressão é confrontada, a nível sintagmático e paradigmático, com todas as outras com que está relacionada. A idiomaticidade surge assim não como uma questão de tudo ou nada, mas como um fenómeno básico e gradual na linguagem. E até as chamadas expressões idiomáticas, tradicionalmente vistas como casos especiais

⁷¹ «[W]e must add to the description of grammar and lexicon a description of the cognitive and interactional ‘frames’ in terms of which the language user interprets his environment...» (Fillmore 1976:23, citado por Anderman 1996:51). Estruturas de conhecimento (**modelos cognitivos ou modelos culturais**) têm sido designados de modo diferente. Além dos de Schank & Abelson 1977, ver, entre outros: Johnson-Laird 1983 que fala de modelos mentais; Fillmore 1985 de ‘scenes-and-frames’; Langacker 1987 de ‘domínios cognitivos’; Lakoff 1987 de ‘modelos cognitivos idealizados’; e D’Andrade 1989 de ‘modelos culturais’.

dentro das línguas, podem ser integradas nesse *continuum* de idiomaticidade. Esta gradação pode ir de expressões variáveis, a nível lexical e morfossintáctico, e que, à semelhança das expressões literais, partilham propriedades linguísticas das construções envolvidas, até às mais fixas, muito próximas de fórmulas. Estudos de processamento revelam a existência de pontos de idiomaticidade, lugares onde cada expressão idiomática se torna inequivocamente identificável como tal (Faria 1999a, 1999b, em prep.).

Assim sendo, e para resumir, a abordagem cognitiva, apoiada na teoria do protótipo, tem, sobre a teoria tradicional dos campos lexicais, algumas vantagens. Soares da Silva destaca as seguintes: « (1) dar conta da vaguidade e da polissemia, (2) mostrar a complexidade de sobreposições de diferentes categorias e descrever valores prototípicos e suas extensões, (3) descrever a parte conotativa dos lexemas e integrar a base sócio-cultural relevante.» (Silva 1999:63).

Exactamente porque têm uma base perceptiva, muitas metáforas são universais. Além disso, elas não são só cognitiva mas também socialmente determinadas (cf. Hatch & Brown 1995:105). Consequentemente, as mais prototípicas reflectem mais o senso comum, reflectem modos semelhantes de olhar a vida e o mundo. E línguas e culturas mais próximas terão, com certeza, mais metáforas comuns. Tendo em consideração que *script*, *frame* e *schemata* representam um nível conceptual, que activa determinados itens lexicais, e apresentam diferenças individuais, sociais e culturais, é fácil imaginar a importância que esta extensão do teoria do protótipo a eventos pode ter no contexto de aquisição e uso de L2. Eles constituem modelos mentais prontos a serem ‘transferidos’ para a conceptualização do discurso em L2, quer a nível da compreensão quer da produção ⁷². Vimos a propósito de Viberg ⁷³ que os verbos nucleares tendem a ser lexicalmente não marcados, que são aprendidos mais cedo por falantes de L1 e L2, que funcionam como protótipos sintácticos, que são muitas vezes usados com

⁷² Em Leiria 1991, comentei um exemplo interessante que envolve diferentes *schemata*. No reconto de uma curta história, em que se falava de, a meio de uma viagem, parar para tomar um café, muitos chineses ‘ouviram’ chá; e só os suecos repararam em dois detalhes: os viajantes levavam consigo uma manta e a estrada tinha muito pó.

⁷³ Ver *Abordagem Tipológica Funcional*.

um sentido não literal, e que fazem parte do nível lexical básico de uma dada língua. No entanto, muitos dos sentidos metafóricos podem ser comuns à L1 e à L2. Isto pode significar que as metáforas mais prototípicas, porque cognitivamente mais salientes, serão mais fáceis de aprender para o falante de L2. No entanto, o falante de L2, até ter tido evidência suficiente sobre este facto pode duvidar da coincidência e usar em relação a elas estratégias de evitação (*avoidance*). Questão a investigar é se, a nível da aquisição, o percurso consiste em, através de conceitos concretos (nomeadamente de ordem física e espacial), aprender conceitos abstractos (Silva 1999:51) ⁷⁴. Mas é óbvio que, neste caso, a questão se porá de modo diferente para aquisição de L1 e de L2.

Slobin (1977:185) ensinou-nos que processos individuais e colectivos de variação e de mudança linguísticas estão inter-relacionados. Assim, para os desvios na produção de L2, é válido o que Soares da Silva nos diz sobre as propostas de Geeraerts (1989, 1997) quanto aos efeitos da prototipicidade na mudança semântica. «[A] prototipicidade envolve quatro características prototípicas, decorrentes da correlação de duas dimensões: por um lado, a *não-igualdade* entre os elementos (referentes, significados e suas propriedades) de uma categoria (item) lexical, isto é, os seus diferentes graus de saliência e a estrutura interna da categoria sob a forma de um centro e uma periferia, e ainda a *não-discrção*, ou seja, a flexibilidade desses elementos e dessa categoria e as dificuldades de demarcação daí resultantes; por outro lado, a distinção entre o nível *extensional* ou referencial (campo referencial de aplicação de uma categoria lexical ou de um significado individual dessa categoria) e o nível *intensional* (dos significados e da sua definição) de um item lexical.» (Silva 1999:29-31).

Do cruzamento destas características da prototipicidade resultam diferentes formas de **mudança** que tanto podem ocorrer no centro prototípico como na periferia de uma dada categoria. No centro, elas dão-se

⁷⁴ Para designar «a hipótese, segundo a qual as expressões espaciais são mais básicas, tanto do ponto de vista gramatical como semântico, do que várias outras expressões de carácter não espacial» usa-se o termo **localismo**, proposto por Anderson 1971 (Batoréo2000:45). Batoréo 2000 desenvolve largamente esta questão por ser «esta uma das linhas mestras» da sua dissertação. Sobre esta teoria Silva 1996:117 comenta: «...a tão propalada tese do 'localismo' tem várias limitações (embora o espaço seja o domínio-origem de muitas funções gramaticais, ele não é, ao contrário do que muitos defendem, a fonte 'primária' da gramaticalização).»

sob a forma de modulações da estrutura do significado. Quanto aos membros periféricos, quanto mais afastados estiverem do centro menos propriedades terão em comum com ele e, por conseguinte, diacronicamente, menos estabilidade apresentarão. O desaparecimento de significados periféricos pode envolver a reestruturação de todo o conjunto ou de parte dele (Silva 1999:66-7) ⁷⁵.

À luz de Slobin 1977, é lícito pensar que um padrão semelhante de reestruturação se dê no processo de aquisição de uma L2. Ou seja, e de forma muito simplificada: itens lexicais mais prototípicos, mais salientes, mais estáveis tenderão a ser adquiridos com mais facilidade, embora possam apresentar, no discurso de L2, alguns efeitos de desvio em certas modulações. Aspectos semânticos periféricos (efêmeros, contextuais) serão adquiridos tarde (ou nunca...) e, caso o sejam, serão muito instáveis, muito sujeitos a variação, no que respeita a sua disponibilidade num mesmo falante. É ainda lícito pensar que este padrão de reestruturação resulte do efeito combinado de condições de prototipicidade intra e interlinguísticas (ver Viberg).

Ao considerarem a metáfora e a metonímia como processos básicos da extensão semântica, os cognitivistas enfrentam a relação léxico-gramática. A **gramaticalização**, tal como a idiomatidade, porque tem por detrás a teoria do protótipo, também é vista como um processo gradual e contínuo. «Por gramaticalização entende-se, à luz destes estudos recentes, o processo gradual e contínuo (diacrónico, como uma espécie de mudança linguística, e sincrónico, como variação de uso num mesmo estágio de língua), pelo qual um item lexical, uma construção, um elemento discursivo-pragmático, ou mesmo um item gramatical se tornam, em determinados contextos linguísticos, uma categoria gramatical (ou mais gramatical, no último caso) passando a codificar uma função gramatical que, ou ainda não tinha sido codificada, ou fora codificada de modo diferente.» (Silva 1996:114).

É inevitável lembrar aqui o que disse sobre os trabalhos de tipologistas como Bybee ou Givón, e sobre o conceito de *continuum* linguístico aplicado a vários níveis. Há, diz a primeira, três modos de

⁷⁵ Para uma visão mais completa da questão, ver Silva 1999.

codificar elementos semânticos: lexical, flexional e sintático; mas adverte: eles não constituem categorias discretas, mas formam um *continuum* que vai da expressão lexical à sintática e que depende de *relevância* cognitiva e cultural (lexical - derivacional - flexional - gramatical livre - sintática) (Bybee 1985:12). Ao nível do discurso, Givón, como vimos, propõe um *continuum* de modos que vai do paratático, mais transparente e menos gramaticalizado, ao sintagmático, menos transparente, mais rápido e mais gramaticalizado. Também para Lakoff 1987, as **construções**, sendo o resultado de um processo de gramaticalização, favorecem um processamento rápido e económico das expressões linguísticas. É a hierarquia que relaciona as variantes mais prototípicas (ou não marcadas) de uma dada construção com as variantes menos prototípicas (ou marcadas) dessa mesma construção favorece o armazenamento da informação. Partindo de uma característica lexical comum, dá conta das diferenças e semelhanças entre todas as expressões em que aparece um dado segmento por meio de relações de natureza metafórica (Martín Arista 1999:206-7).

Têm surgido recentemente algumas propostas de Gramáticas Cognitivas ⁷⁶ que têm como objectivo articular a investigação realizada no âmbito da Linguística Cognitiva. A designação Gramática Cognitiva, em sentido estrito, identifica o modelo de Langacker (1987, 1991). Afirma-se contra as teses da autonomia e da modularidade da linguagem. É uma gramática compatível com o modelo conexionista ⁷⁷, simbólica, de valores conceptuais e semânticos e de imagens (Silva 1995a, 1995b, 1995c). «Gramática e léxico formam um continuum de unidades simbólicas, só arbitrariamente divisíveis em componentes separadas, e a semântica (que naturalmente absorve a pragmática) constitui a dimensão básica da gramática. A estrutura semântica é a expressão convencional da estrutura conceptual; a sintaxe e a morfologia são a simbolização convencional da estrutura semântica.» (Silva 1995a:402). Langacker reduz assim a gramática a relações simbólicas entre estruturas semânticas e fonológicas.

⁷⁶ Além desta, de Langacker, há pelo menos mais dois modelos: Kay & O'Connor 1988 e Goldberg 1995. Mas, enquanto que os primeiros seguem a linha proposta por Fillmore 1982, a segunda, de que falaremos a seguir, segue Lakoff 1987.

⁷⁷ Ver 'conexionismo', em Capítulo 2.

Goldberg 1995 – *A Construction Grammar Approach to Argument Structure* – é também uma Gramática Cognitiva. Afirma-se como uma gramática generativa, mas não transformacional e não inatista («knowledge of language is knowledge», Goldberg 1995:5). Assume que não existe uma divisão estrita entre léxico e sintaxe, e entre semântica e pragmática («Information about focused constituents, topicality, and register is represented in constructions alongside semantic information», Goldberg 1995:7). E assume que as construções de uma língua constituem um conjunto altamente estruturado de entidades com informação inter-relacionada (Goldberg 1995:5) ⁷⁸. Afirma apoiar as suas propostas em investigação em aquisição da linguagem (Clark 1978, Slobin 1985 e Bowerman 1989) e, como Lakoff 1987, também ela estabelece relações entre construções que partilham um elemento qualquer e afirma que este facto reflecte a experiência humana («Simple clause constructions are associated directly with semantic structures which reflect scenes basic to human experience.» Goldberg 1995:5) ⁷⁹.

Enquanto que numa abordagem lexical pura se considera que estruturas sintáticas são a projecção da semântica do léxico, neste caso «the meaning of an expression is the result of integrating the meaning of the lexical items into the meaning of constructions.»

*This work has argued, counter to the current trend, that **an entirely lexically based approach to grammar is inadequate**, and that lexically unfilled constructions must be recognized to exist independently of the particular lexical items which instantiate them. By recognizing the existence of **meaningful constructions**, we can avoid the claim that the syntax and semantics of the clause is*

⁷⁸ «In Construction Grammar, no strict division is assumed between the lexicon and syntax. **Lexical constructions and syntax constructions differ in internal complexity**, and also in the extent to which phonological form is specified, but both lexical and syntactic constructions are essentially the same type of declaratively represented data structure: both pair form with meaning. It is not the case, however, that in rejecting a strict division, Construction Grammar denies the existence of any distinctly morphological or syntactic constraints (or constructions). Rather, it is claimed that there are basic commonalities between the two types of constructions, and moreover, that there are cases, such as verb-particle combinations, that blur the boundary.» (Goldberg 1995:7).

⁷⁹ Refere outras **abordagens construcionais** e, de entre elas, revê Montague 1973, Wierzbicka 1988 e Jackendoff 1990. A propósito deste, diz: «On the present account, constructions play a more central theoretical role than on Jackendoff's account. For Jackendoff, correspondence rules are only required in exceptional cases, when the correspondence rule itself contributes an argument. For the majority of cases, he supposes that argument structure is determined on the basis of verbal semantics in isolation.» (Goldberg 1995:221).

projected exclusively from the specifications of the main verb (Goldberg 1995:224).

Ou seja, a semântica do verbo deve juntar-se à semântica da construção para explicar o significado de uma expressão. Além disso, e para atingir a interpretação completa da expressão no contexto, também é necessário conhecimento extralinguístico ⁸⁰ (Goldberg 1995:16 e n. 12). Isto quer dizer, em última análise, que as representações que temos das várias construções em que um dado item lexical costuma aparecer são activadas no processamento e que, por conseguinte, não existe uma divisão estrita entre léxico e construções. Esta posição retira, assim, parte da ênfase que tem sido posta na semântica lexical e, em especial, na semântica do verbo, e, ao cruzar a semântica com a sintaxe, combina o plano paradigmático e o sintagmático de um modo que parece reflectir a representação mental que temos do léxico ⁸¹. Por outro lado, a Gramática das Construções não só integra informação de outros paradigmas da Ciência Cognitiva, como é o caso da aquisição da linguagem, mas também parece querer fazer uma simbiose entre as três abordagens – a **generativa**, a **funcional** e a **cognitiva**.

1.5. Sumário

Neste muito breve percurso sobre o tratamento dado ao Léxico em várias abordagens linguísticas, pudemos constatar uma progressiva afirmação da semântica em relação à sintaxe. Partimos da autonomia da sintaxe, nos primeiros modelos generativistas, passámos por posições de um relativo equilíbrio, até chegarmos à dependência da sintaxe em Gramáticas Cognitivas. O léxico é visto actualmente como o repositório das propriedades sintácticas, capaz de desenvolver mecanismos de articulação

⁸⁰ [Wierzbicka 1988] «proposes that the entire semantics of any lexical item can be captured by paraphrases involving these atomic primitives combined in determinate ways. We have taken rather the opposite approach to semantics, arguing that lexical items are associated with rich frame-semantic or encyclopedic knowledge, and that decomposition into atomic elements is impossible.» (Goldberg 1995: 224)

⁸¹ «Una primera evaluación de la comparación entre las dos teorías [*Gramática Funcional* y *Gramática de las Construcciones*] parece indicar que la GC combina más acertadamente la sintagmática y la paradigmática que la GF: en la GC las construcciones sustentan la dimensión paradigmática y los verbos la sintagmática, de tal forma que las construcciones relacionan el plano semántico con el plano sintáctico.» (Martín Arista 1999:200).

entre a informação nele contida e a realização sintáctica. Essa progressiva afirmação traduz-se em uma **centralidade** e **abrangência** patentes, quer dentro de cada uma das abordagens (como é o caso da abordagem generativa), quer nos estudos linguísticos em geral. A abordagem cognitiva, que é relativamente recente, e que resultou de uma cisão dentro da generativa e de uma releitura de muitas propostas funcionais, fez da semântica o seu objecto de estudo privilegiado. No entanto, a Gramática das Construções parece anunciar uma via de conciliação da sintaxe com a semântica e uma integração dos planos sintagmático e paradigmático.

Devido a esta preponderância do léxico nos modelos de descrição linguística, os contornos da sua representação constituem, actualmente, um dos grandes desafios dos linguistas, já que a informação aí armazenada é central para o desenho do resto da teoria da gramática. No entanto, muito dificilmente a descrição oferecida por qualquer uma das teorias será tão rica que possa reflectir a competência lexical de qualquer falante. Apesar disso, se não encararmos as várias abordagens como paralelas ou contraditórias, mas antes como complementares, e se, por vezes, recuperarmos até certos caminhos abandonados ao longo do percurso, veremos que elas nos podem oferecer fundamentação adequada para muitos dos fenómenos em estudo.

Tal como seria de prever, certas abordagens oferecem-nos mais apoio para o estudo de certos aspectos da aquisição do léxico e outras mais para outros. Isso depende, em parte, da atenção que cada uma delas tem dedicado à descrição dos fenómenos linguísticos. Assim, a mais formal, a **generativa**, parece dar melhor resposta a questões relacionadas com os aspectos mais formais, mais sistémicos, mais ligados à competência, mais ligados a conhecimento procedimental, como é o caso da fonologia e dos aspectos sintácticos do léxico. Aspectos da teoria, abandonados, como é o caso dos traços semânticos, ou mantidos até aos modelos mais recentes, como é o caso dos papéis temáticos, são relevantes. As **funcionais e cognitivas**, porque preocupadas com o uso, oferecem modelos que dão conta de modo mais satisfatório da componente semântica e pragmática do léxico. Estas duas últimas, porque não estabelecem fronteiras entre léxico e gramática, parecem, portanto, oferecer enquadramentos também para o estudo de certas relações sintagmáticas e paradigmáticas do léxico, nomeadamente para o problema que mais do que uma língua em presença

coloca no que respeita a definição de palavra. A abordagem **tipológica** é aquela que mais apoio pode fornecer para os aspectos propriamente semânticos do léxico, nomeadamente para os objectivos centrais deste trabalho, já que é ela que mais se tem ocupado do tratamento da lexicalização de conceitos e, em particular, da fusão de significados. Além disso, o facto de ser comparativista torna-a um instrumento indispensável em estudos como este, estudos que envolvem mais do que uma língua. A abordagem **cognitiva**, que, na realidade, apoia muitas das suas propostas na tipologia linguística, oferece-nos fundamentação para a formulação de hipóteses no que respeita a ordem de aquisição dos diferentes aspectos do léxico e, em última instância, para a ordem de aquisição, em geral, de uma língua particular, por um falante particular.

Lluís Payrató, num livrinho que dedica «a los lingüistas aplicados y a las lingüistas aplicadas (y sobre todo viceversa)», afirma:

Si es cierto que no hay nada más práctico que una buena teoría, y si también lo es que las teorizaciones angelicales no resuelven ningún conflicto de la práctica comunicativa, una concepción pertinente de la lingüística aplicada debería ser capaz de conjugar ambas afirmaciones, una tarea más bien complicada pero en absoluto desdeñable. Entre soñar quimeras y aplicar recetas de forma rutinaria, debe de quedar algún hueco para la imaginación sensata (Payrató 1998:125).

O percurso feito neste capítulo teve exactamente como objectivo evitar a escolha de um quadro teórico ‘angelical’ ou a aplicação de uma ‘receita’.

Capítulo 2

Léxico e bilinguismo

2.1. Introdução

O percurso no âmbito da Psicolinguística é paralelo àquele que encontrámos na Linguística. Nos últimos trinta e cinco anos ⁸², a investigação centrada no léxico mental não tem parado de crescer. No entanto, especialmente até ao início dos anos 90, essa investigação sobre o processamento do léxico usava como estímulos sobretudo palavras isoladas ⁸³. Mas, tendo em conta que o que se pretende é saber quais as operações que um ouvinte/falante tem de realizar para interpretar/produzir discurso em contexto natural, estudos sobre o processamento *on-line* começaram a fazer-se sentir como indispensáveis ⁸⁴.

No **processamento *on-line***, a compreensão do discurso exige que o ouvinte integre uma grande variedade de tipos de conhecimento linguístico e não-linguístico: propriedades acústico-fonéticas, lexicais, sintácticas e semânticas mais o conhecimento geral não linguístico do mundo ⁸⁵. Alguns

⁸² «It is now more than twenty years since Oldfield (1966) developed the concept of the **mental lexicon**, and only slightly less since the pioneering lexical-decision experiments of Rubenstein, Garfield, and Millikan (1970).» (Forster 1992:75).

⁸³ «[R]esearch on bilingual memory does appear to be evolving toward the use of more contextualized linguistic units, which are more likely to avoid a number of limitations associated with studies relying on isolated words or, to a lesser extent, isolated sentences tasks. [...] such studies [that employ isolated words] constitute a valuable tool in the investigation of bilingual cognitive processes. However, there is also an urgent need for studies that employ grammatically and semantically complex language units and thus allow full access to the cognitive processes involved in the normal uses of language.» (Hummel 1993:281). Ver, em 2.2.1, *frase e enunciado*.

⁸⁴ «But even in this current linguistic climate of emphasis on the importance of the lexicon, the on-line processing functions of lexical representations remain neglected – especially in the auditory domain.» (Marslen-Wilson *et alii* 1988:3).

⁸⁵ Marslen-Wilson e colegas dão testemunho da sua investigação sobre **processamento ‘on-line’** numa série de trabalhos (1980, 1988, 1991, 1995) . O conjunto destes textos do *cohort model* assenta sobre três ideias chave: (1) a representação lexical envolve aspectos

psicolinguistas têm feito questão em afirmar que «**o léxico mental é o elo central no processamento da linguagem**», e que «uma abordagem combinada, psicológica e linguística se afigura como inevitável se queremos compreender a linguagem como fenómeno psicolinguístico». É assim evidente também para a Psicolinguística, como para a Linguística Teórica e para a Linguística Computacional, que a representação lexical desempenha um papel central no processamento (Marslen-Wilson *et alii* 1988:1-2; Marslen-Wilson 1988:292).

*The **mental lexicon** is a very important part of any model of language processing. It plays **a central role** because it provides a bridge between form and meaning. In the mental lexicon information from all different linguistic levels is combined. Phonology, orthography, syntax, argument structure, morphology and lexical semantics all appear in the entries of the mental lexicon (see, e.g. Aitchison 1987). Understanding just how this information is represented and used is crucial to any model of language perception and production (Schreuder & Weltens 1994:4-5).*

Este papel central que a Psicolinguística tem reconhecido ao léxico mental traduz-se no destaque que lhe tem sido dado nos modelos de percepção e produção de discurso. No entanto, esses modelos têm tido em consideração exclusivamente sujeitos monolíngues e não bilingues.

Mas considerar que uma pessoa é ou não bilingue depende **do que se entende por bilinguismo**. Um critério que costuma ser usado para definir bilinguismo é o **contexto de aprendizagem**. Para alguns, apenas a aquisição de duas línguas em contexto informal é considerada como conducente ao bilinguismo, enquanto que, para outros, também a aprendizagem em contexto formal pode conduzir ao bilinguismo. No entanto, esta dicotomia informal vs. formal é bastante problemática. Ou seja, na maior parte dos casos, e como quase sempre acontece, não se trata de uma dicotomia mas de um *continuum* (Schreuder & Weltens 1993:2-3). Aquilo a que se chama aprendizagem de uma língua estrangeira envolve muitas vezes uma combinação de instrução formal e aquisição em ambiente natural. É o que acontece, actualmente, com as crianças e com os jovens

sintácticos, semânticos e fonológicos; (2) a representação semântica é activada muito cedo (Marslen-Wilson *et alii* 1988); (3) a representação fonológica é altamente abstracta e muito pouco específica (é a variação, o facto de uma palavra poder realizar-se de muitas maneiras diferentes, que chama a atenção para este aspecto) (Marslen-Wilson *et alii* 1995).

portugueses que aprendem Inglês na escola e são expostos diariamente a *input* abundante que lhes é disponibilizado pelos *media* ⁸⁶.

Se usarmos como critério o nível de **proficiência**, e definirmos como bilinguismo apenas o uso habitual, fluente e sem sotaque de duas línguas, muito poucos indivíduos poderão ser considerados como bilingues. Se, pelo contrário, aceitarmos que pessoas consideradas como bilingues não formam um grupo homogêneo, mas que se situam em pontos diferentes de um *continuum* multidimensional, que reconhece diferenças, quanto ao tipo e ao grau, no que respeita as estruturas e as modalidades linguísticas, então o bilinguismo será um fenómeno bastante generalizado. E incluirá também situações de bidialectalismo e de multilinguismo, isto é, qualquer situação em que mais do que uma língua ou uma variedade é usada regularmente (Paradis 1986:xi). Actualmente, no âmbito dos estudos sobre bilinguismo, a tendência é, de facto, para considerar esta perspectiva bastante alargada de bilinguismo, por isso, uma grande parte dos estudos pressupõe diferentes níveis de proficiência e diferentes oportunidades de aprendizagem e de uso. Kees de Bot diz mesmo que qualquer modelo de produção de discurso devia ter em consideração o bilinguismo, com uma versão opcional para o monolinguismo (De Bot 1992:2).

Mas, não havendo, até ao momento, um **modelo de compreensão e produção para falantes bilingues** ⁸⁷, muitos investigadores de L2 têm procurado adoptar e adaptar um modelo proposto por Levelt em 1989, que pretende dar conta de todos os processos envolvidos na produção de discurso desde a sua conceptualização até à articulação, e que dá particular destaque ao papel do léxico no processamento. Este modelo tem oferecido aos investigadores orientação para poderem reflectir sobre os problemas que se vão pondo ao falante bilingue ao longo do processo de produção de discurso, de tal modo que o trabalho mais consistente sobre os aspectos psicolinguísticos do bilinguismo tem-se desenvolvido no seu âmbito.

⁸⁶ Exactamente por isso, D'Allwright, para distinguir entre **língua estrangeira** e **língua segunda**, abandona esta dicotomia e propõe que olhemos para as oportunidades de aprendizagem como 'encounter opportunities' e 'practicing opportunities', isto é, "opportunities to *meet* whatever is to be learned' e 'opportunities to *do something* with target material' (Allwright 1991:166-168).

⁸⁷ «There is as yet no integrated **model** of comprehension and production in L2. The Levelt's model of speech production offers one route to this goal.» (Green 93:270).

Assim, como tantos outros, **assumirei um conceito abrangente de bilinguismo que considerarei adequado ao *continuum* de proficiência linguística que os aprendentes de L2 visados neste trabalho representam.** Por outro lado, o facto de também usar este modelo para organizar a reflexão que se segue não significa, no entanto, que assumo que ele, ou qualquer outro existente até ao momento, por si só, possa dar conta da complexidade de que se reveste a produção de discurso, e, em especial, a produção de discurso por falantes bilingues. Como disse antes, tendo em conta o estado dos conhecimentos, parece-me avisado reconhecer que, para já, só a colaboração entre diferentes perspectivas pode fazer alguma luz sobre a complexidade destas questões. O bilinguismo é um fenómeno (pelo menos, duplamente) complexo e variado. Tudo o que se disser sobre ele pecará por defeito e por imprecisão.

2.2. Produção de discurso por bilingues: o léxico como mediador

The assumption that the lexicon is an essential mediator between conceptualization and grammatical and phonological encoding will be called the lexical hypothesis (Levelt 1989:181).

O modelo de Levelt 1989 propõe quatro fases para a produção de discurso: a **conceptualização da mensagem**, a **codificação gramatical**, a **codificação fonológica** e a **articulação**. Estas quatro fases são desempenhadas por três componentes de processamento autónomos: o *conceptualizador*, o *formulador* e o *articulador*. Faz também parte do modelo um sistema de compreensão que permite ao falante compreender o seu interlocutor e avaliar as suas mensagens permitindo-lhe também corrigir eventuais erros. Por se tratar de um modelo de produção, as operações realizadas por este sistema não são descritas de forma detalhada.

O modelo de Levelt 1989 é serial e modular. Serial, no entanto, não significa que haja autonomia e sequencialidade absoluta entre as várias componentes; significa sim que diferentes partes da mensagem podem estar a ser processadas ao mesmo tempo, podem estar em diferentes estádios de produção e não interferem umas com as outras (cf. Towell & Hawkins 1994:168). Quanto a ser modular, Singleton 1999 faz notar que Chomsky e Fodor nunca defenderam que o módulo da linguagem não tivesse qualquer

forma de conexão com outras operações cognitivas, ou que todos os aspectos da cognição estejam organizados por módulos. É óbvio que o uso normal da linguagem requer interfaces entre a linguagem e outros aspectos do conhecimento. Fodor e Chomsky sustentam que esta interligação é levada a cabo por uma espécie de 'central', não-modular, que liga os módulos entre si e enriquece o seu *output* com a experiência acumulada pelos vários módulos em operações anteriores (Chomsky 1981 e Fodor 1983, citados por Singleton 1999:112) .

O **conceptualizador** gera a mensagem. Gerar a mensagem envolve macroplaneamento, ou seja, conceber o conteúdo da mensagem, e microplaneamento, ou seja, conceber a forma da mensagem. Para isso, usa conhecimento declarativo. Esse conhecimento declarativo envolve conhecimento enciclopédico ou conhecimento do mundo, conhecimento sobre o modelo discursivo mais adequado ao contexto comunicativo e conhecimento sobre a situação espacial envolvente. O **formulador** recebe como *input* esta mensagem pré-verbal. Então, o falante procura no **léxico mental**, «the store of information about the words in one's language» (Levelt 1989:6), os itens lexicais que expressem correctamente a mensagem planeada. Os itens são acedidos com base no seu significado, mas eles também contêm informação sintáctica, morfológica e fonológica. Além da informação semântica, o falante usa informação sintáctica, e por vezes também morfológica, e executa a codificação gramatical (daí resultando o *lema*). O codificador fonológico acrescenta-lhe a informação fonológica (daí resultando a forma, ou *lexema*). Uma vez acedida esta forma, o falante pode construir o plano articulatório ou fonético que o **articulador** transformará em fala.

A **hipótese lexical** preconiza que as operações realizadas pelo formulador, ou seja, a codificação gramatical e fonológica, são mediadas pelas entradas lexicais. As **entradas lexicais** constituem o léxico mental. O léxico mental é um repositório de conhecimento declarativo sobre as palavras da língua. Esse conhecimento é constituído, pelo menos, por quatro tipos de propriedades relacionadas com o significado, a sintaxe, a morfologia e a fonologia. Os itens estão ligados entre si de diversas maneiras. **As relações entre os itens são de dois tipos: as que se estabelecem dentro de cada entrada e as que se estabelecem entre as**

diferentes entradas. Nem todos os itens são entradas. As formas flexionadas de um mesmo verbo, por exemplo, são itens que pertencem à mesma entrada. Ou seja, estão interligados *dentro de* uma mesma entrada. As relações entre as entradas lexicais são de dois tipos: intrínsecas e associativas. As relações intrínsecas derivam dos quatro tipos de propriedades referidos (semânticas, sintáticas, morfológicas e fonológicas). Assim, pode haver conexões entre duas entradas por pertencerem ambas a um mesmo campo semântico, por razões de natureza derivacional, por haver conexões fonológicas entre elas ou por pertencerem à mesma categoria sintática. Relações associativas entre entradas não são necessariamente baseadas em propriedades semânticas; dependem sobretudo de conexões conceptuais e da frequência de coocorrência entre elas (cf. ‘collocations’).

Uma crítica feita frequentemente ao modelo diz respeito à associação exclusiva, ao léxico, de conhecimento declarativo. Sendo o léxico não uma entidade estática mas dinâmica, muitos são os linguistas, nomeadamente os conexionistas ⁸⁸, que não aceitam a distinção entre **conhecimento declarativo** e **conhecimento procedimental**, e que têm chamado a atenção para a criatividade lexical. Esta permite a formação e a compreensão de novas palavras, o que constitui, com certeza, conhecimento procedimental. Sergio Scalise tenta uma saída linguística apoiada em argumentação psicolinguística. Propõe que, para a **morfologia**, tem de existir uma diferença crucial entre ‘*memorizzato*’ e ‘*formato per regole*’: «La morfologia deve potere distinguere tra unità formate tramite regole e unità memorizzate. Questa distinzione – in termini linguistici – può essere resa con la distinzione tra il lessico e il componente delle regole» (Scalise 1994:73). Pode pôr-se como hipótese, afirma, que as palavras simples sejam todas memorizadas e que, ao contrário, as palavras complexas que tenham

⁸⁸ Ver Capítulo 3: **conexionismo**. «The notion of independence of different processing operations in **parallel models** such as the connectionist one is not, however, in any real sense comparable to the Fodorian idea of informational encapsulation. Independence of processing in parallel models refers to micro-operations, and is not to be identified with a barrier between, for example, ‘higher-level’ semantic processes and ‘lower-level’ formal computations. Parallel-processing models are usually interpreted, on the contrary, as making claims about a high degree of top-down / bottom-up interactivity.» (Singleton 1999:123). Dell 1986, por exemplo, é um modelo que se apoia no conceito de ‘*spreading activation theory of retrieval*’, que é um elemento fundamental nas propostas conexionistas.

estruturas regulares sejam construídas com o recurso a regras (Scalise 1994:72). Estabelece, deste modo, dentro do léxico, a distinção entre conhecimento declarativo, palavras memorizadas, e conhecimento procedimental, capacidade para formar novas palavras com recurso a regras. No entanto, a questão não é tão simples assim. Ficam por resolver muitas questões relacionadas com as diferentes representações, e consequentemente as diferentes relações entre léxico e morfologia, no que respeita diferentes tipos de línguas. O que é, por exemplo, uma palavra memorizada numa língua aglutinante como o Húngaro?

Outro aspecto que tem causado estranheza é a distinção entre conhecimento enciclopédico e conhecimento semântico-lexical (cf. Singleton 1999:108-9). De facto, muitos têm sido, como vimos, os investigadores que têm chamado a atenção para a impossibilidade de estabelecer uma fronteira entre **conhecimento lexical** e **conhecimento enciclopédico** (Aitchison 1987, Lakoff 1987, entre muitos outros) e que, por isso, defendem que o léxico mental não pode ser visto como um módulo separado de outras formas de conhecimento armazenadas na memória a longo prazo (Hulstijn 1997:211).

O modelo também prevê que **certas entradas podem ser constituídas por mais do que uma palavra** ⁸⁹. É o caso das expressões idiomáticas, de locuções, combinatórias frequentes e fórmulas: «idiomatic collocations are entries in the mental lexicon» (Levelt 1989:187). O seu significado pode ser mais ou menos opaco e pode não ser derivável do significado das partes ⁹⁰. Este será assim um dos casos em que o **princípio de convergência** não ficará garantido: «Convergence would be guaranteed when concepts and words would entertain a simple one-to-one relation» (Levelt 1989:200). Um dos problemas consiste, exactamente, em saber onde se situa o «ponto de idiomaticidade» (cf. Faria, citada antes). Sabendo nós que o princípio da convergência muitas vezes não se verifica, outro

⁸⁹ «Si possono distinguere varie accezioni di 'parola', a secondo del punto di vista a partire del quale si considera questo 'oggetto'. Così, la nozione di parola fonologica non coincide con la nozione di parola morfologica o di parola sintattica.» (Scalise 1994:61). Ver também Capítulo I, nota 31 ('word').

⁹⁰ «Per la **morfologia** esiste una differenza cruciale tra 'memorizzato' e 'formato' per regole. Questa differenza non esiste per la sintassi. O meglio, esiste solo come caso limite (frasi idiomatiche, frasi che hanno acquisto significati non prevedibili, non regolari).» (Scalise 1994:72).

problema igualmente importante para a problemática do léxico bilingue consiste em saber como é que, nas diferentes línguas, a informação semântico-conceitual se distribui pelos diferentes itens lexicais. Para a reflexão sobre este assunto, como vimos, deram uma contribuição valiosíssima os tipologistas, e muito em especial Leonard Talmy.

Cada entrada, como vimos, contém sempre informação semântica e sintáctica. Quanto à informação morfológica, ela difere muito conforme a língua. De facto, as línguas exploram de modos muito diferentes as possibilidades derivacionais, e isso, como disse a propósito de Scalise e de Bybee, tem importância na representação. Assim, é provável que numa língua aglutinante, como o Turco ou o Húngaro, a representação da morfologia derivacional seja diferente da de línguas como o Inglês ou o Chinês que fazem pouco uso deste recurso ⁹¹. Isto não significa, no entanto, que estas línguas não disponham de outros recursos para cumprir os mesmos objectivos comunicativos. As línguas diferem exactamente no modo como lexicalizam os conceitos, mas, na verdade, nada é intraduzível de uma língua para outra.

Languages differ not so much as to what can be said in them, but rather as to what is relatively easy to say (Hockett 1954:122)

Sendo assim, lexicalização de conceitos, diferentes padrões de lexicalização, em geral, e a convergência e a fusão de significados, em particular, são questões primordiais para a problemática do léxico bilingue. Na mesma medida em que o são também as hierarquias de lexicalização enquanto espelhos da prototipicidade e, portanto, da relevância cultural.

⁹¹ Literatura muito abundante sobre este tema reflecte a intensidade deste debate. Veja-se, entre outros: Lukatela & Carello & Turvey 1987; Kess 1992; Laudanna & Badecker & Caramazza 1992 ("there are multiple representational levels at which morphological structure is represented"); Salmons 1993; Sandra 1994. Gergely & Pléh 1994 discutem os modelos de processamento lexical (Logogen e Cohort Model) e a sua adequação a línguas como o Húngaro (tão diferente do Inglês no qual se baseia a maior parte da investigação): «From the point of view of lexical processing the major empirical question that arises in relation to **agglutinative (and inflectional) languages** is whether the morphologically complex words are represented in a holistic or a morphological decomposed form in the mental lexicon and whether lexical access requires some form of morphological parsing of the word.» (Gergely & Pléh 1994:180)

2.2.1. *A mensagem pré-verbal*

Como vimos, a mensagem pré-verbal é gerada no conceptualizador. Antes de mais, o falante tem de decidir sobre o que é que quer falar e como o quer fazer ⁹². Esta operação, a que Levelt chama macroplaneamento, envolve a elaboração de um determinado objectivo comunicativo numa série de sub-objectivos, e a busca da informação a ser expressa de modo a atingir cada um destes sub-objectivos. No microplaneamento, o falante fixa a forma (o tipo de acto de fala) e a perspectiva (o tópico e o foco) que considera mais adequadas a cada um dos *chunks* ⁹³ de informação, de modo a atrair a atenção do seu interlocutor. Para realizar estas operações usa conhecimento procedimental e conhecimento declarativo. **O conhecimento procedimental** permite ao falante construir a asserção que deseja exprimir; esta é depositada na memória de trabalho que contém toda a informação procedimental disponível para gerar e monitorizar a mensagem. Gerar e monitorizar a mensagem são actividades controladas que requerem atenção constante por parte do falante. Na memória a longo prazo está disponível **conhecimento declarativo** (e ou enciclopédico) adquirido ao longo da vida sobre a situação e sobre o modelo discursivo presentes. O *output*, a mensagem pré-verbal assim gerada no conceptualizador, será o *input* para o formulador (Levelt 1989:9-11).

Dispor-se a planear uma mensagem pressupõe **motivação** para comunicar. É essa motivação que conduz as crianças a adquirirem a sua L1. O sistema límbico é responsável, entre outras coisas, pela emoção, pelos desejos e pela motivação. A propósito, Michel Paradis observa que uma parte do insucesso que se verifica na aquisição de L2 em contexto formal pode radicar, exactamente, na falta de verdadeira motivação que caracteriza certos métodos e práticas de sala de aula. É verdade que há mais motivação em alguns dos actuais métodos do que, por exemplo, no método gramática-

⁹² Como é que os falantes coordenam planear e falar é relativamente desconhecido (cf. Meyer 1996).

⁹³ Foi George Miller (1956) quem primeiro usou o termo 'chunking'. «A **chunk** is a unit of memory organisation, formed by bringing together a set of already formed chunks in memory and welding them together into a large unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organisation of memory. Chunking appears to be a ubiquitous feature of human memory. Conceivably, it could form the basis for an equally ubiquitous law of practice.» (Newell, 1990:7)

tradução (*La plume de ma tante est sur le bureau de mon oncle...*)⁹⁴. No entanto, mesmo nos métodos mais recentes, fala-se para cumprir uma tarefa, faltando assim a motivação inicial. Em jogos ou em actividades em que o estudante tem mesmo vontade de falar, aí sim, há real motivação (Paradis 1994:406-7).

O modelo de Levelt, ao incluir no conceptualizador aspectos relacionados com o uso da linguagem, assume implicitamente que, para além de usar competência linguística, o falante tem de recorrer a outros conhecimentos, condicionados por outros aspectos cognitivos. Avaliar as características da situação, escolher o acto de fala adequado, identificar sentidos não literais de um enunciado, ou traduzir os gestos e as entoações do interlocutor, são tarefas que podem colocar sérios problemas a um falante de L2, especialmente se ele é oriundo de uma cultura muito diferente. Isto significa que, tanto para a compreensão como para a produção, certas operações realizadas pelo monitor e pelo conceptualizador envolvem **conhecimentos extra-linguísticos, para-linguísticos, sociolinguísticos e pragmáticos** que, no processamento *on-line*, exigem do falante um esforço adicional, esforço esse que ocupará ainda mais a sua atenção e a sua memória de trabalho.

Paradis 1998 revê estudos sobre pacientes com lesões no hemisfério direito e conclui que, no seu conjunto, estes pacientes parecem ter dificuldade em usar informação contextual para interpretar discurso. **Competência pragmática**, uma gramática do discurso que inclua regras de pressuposição e de inferência, que nestes pacientes não está disponível, é tão necessária em situações de interacção verbal como a competência linguística. Por isso, afirma, há razões teóricas e neuro-funcionais, para distinguir **frase** de **enunciado**. A primeira é uma entidade linguística independente do contexto e a sua interpretação semântica pode ser derivada do significado das palavras e da estrutura gramatical; a interpretação da segunda requer também operações de inferência a partir dos vários contextos (situacional, discursivo e conhecimento geral) (Paradis 1998:3). Sendo assim, algumas diferenças relatadas em trabalhos de neurolinguística, quanto ao maior ou menor envolvimento do hemisfério

⁹⁴ O exemplo é de Paradis 1994.

direito nos processos conducentes à compreensão e produção de discurso por falantes bilíngues, podem estar relacionadas exactamente com aspectos extra-linguísticos, para-linguísticos, sociolinguísticos e pragmáticos que relevam da distância entre as duas línguas. E também com o tipo de tarefas usadas na experimentação; ou seja, a qualidade e a quantidade de conhecimento declarativo que o falante tem de disponibilizar varia em função das diferentes modalidades da escrita e do oral; e é aí que residem certas diferenças advertidas *intra* e *inter* falantes de L2⁹⁵. Dito de outro modo, a nível da *performance*, um mesmo falante pode evidenciar grandes diferenças conforme a tarefa em que está envolvido, dependendo isso não só de competência linguística mas também do conhecimento enciclopédico e declarativo exigidos pelas **características do modelo discursivo**. É neste sentido que o nível cultural, os hábitos de aprendizagem formal, em geral, e de aprendizagem de línguas, em particular, podem desempenhar um papel de destaque nas diferenças evidenciadas por falantes de uma mesma L1 no processamento e na aprendizagem de uma mesma L2. Além de mais ou menos conhecimento declarativo, estão disponíveis e podem ser usadas na L2 mais ou menos estratégias (ou seja, conhecimento procedimental). Diferenças entre falantes de diferentes L1, a aprenderem a mesma L2, estão relacionadas não só com a competência linguística, mas também com o conhecimento do mundo, e com a quantidade de conhecimentos para-linguísticos, sociolinguísticos e pragmáticos aplicáveis à L2.

2.2.2. *Entre a conceptualização e a formulação da mensagem*

Vimos que, no modelo de Levelt, ao léxico mental, que constitui a segunda fonte de conhecimento declarativo, é atribuído o papel de mediador

⁹⁵ «One of the most controversial and most discussed issues over the past 20 years has been the notion that the language of bilinguals, or of some well-defined subset of bilinguals, are less asymmetrically represented in the **cerebral hemispheres** than the language of unilingual speakers. [...] This state of affairs may be exacerbated by a lack of specification of what is meant by *language*, that is, what is that these investigators suspect of being less lateralized. If they mean implicit linguistic competence, they are clearly wrong. [...] If they mean communicative competence, including pragmatics, they are probably right...[...] second language speakers, like children, are liable to compensate for the lacunae in their implicit linguistic competence in the weaker language by increasing their reliance on pragmatic inferences ... [...] Increased right hemisphere involvement in these cases, however, does not reflect the representation or processing of the language system (implicit linguistic competence), but, on the contrary, whatever nonlinguistic competence is substituted for it.» (Paradis 1997:338).

entre a conceptualização e a formulação da mensagem. O facto de o modelo fazer esta distinção – entre o *output* da conceptualização, ou seja, a mensagem pré-verbal, e a formulação – tem certas implicações para um modelo bilingue: se as línguas diferem no modo como os conceitos são lexicalizados, como é que, no formulador, se projectam as **estruturas conceptuais** da mensagem pré-verbal nas **representações semânticas** contidas no léxico? (Green 1993:251) ⁹⁶. Ou seja, como é que o sistema resolve a questão dos diferentes **padrões de lexicalização** das diferentes línguas? E como é que o falante bilingue lida com o facto de o seu poder ser um **léxico deficitário**? (De Bot & Schreuder 1993:212). São várias as propostas que pretendem, no quadro deste modelo, encontrar respostas para estas questões.

É um facto, como vimos, que as línguas diferem consideravelmente no que respeita a **lexicalização** de uma dada estrutura conceptual. A relação entre primitivos conceptuais e os elementos da estrutura de superfície de uma língua não estabelecem entre si uma relação de convergência de um para um. O que numa língua pode ser expresso pela morfologia, por exemplo, noutra pode sê-lo por itens lexicais soltos (ver Talmy 1985, Jackendoff 1991, Levin & Pinker 1991). Bierwisch & Schreuder 1992 (seguidos por De Bot & Schreuder 1993) argumentam que o conceptualizador e a mensagem pré-verbal não contêm especificação linguística. Ou seja, o conceptualizador não fornece ao formulador uma lista de conceitos lexicalizados mas sim uma cadeia de informação formada por primitivos conceptuais que tem de ser acomodada em *chunks* lexicalizáveis, antes de o léxico ser acedido. Antes de mais, e por causa dos diferentes padrões de lexicalização, a mensagem pré-verbal deve ser informada quanto à língua que vai ser escolhida. Para resolver o problema da não convergência entre conceitos e padrões de lexicalização, acrescentam ao modelo unilingue de Levelt um dispositivo – o *verbalizador* (VBL) – cuja tarefa consiste em projectar as estruturas conceptuais da mensagem pré-verbal em representações semânticas contidas no léxico, isto

⁹⁶ «There must be a mapping between the conceptual representation and the specification of word meaning in the formulator and such mapping might differ between languages because language differ in terms of how concepts are lexicalized.» (Green 1993:251). Em De Bot 1992 não há o dispositivo VBL. De Bot sugere que a mensagem pré-verbal contém informação linguística para ser lexicalizada pelo formulador.

é, dividir a mensagem pré-verbal em *chunks* que possam ser lexicalizados. Quando o VBL termina esta tarefa, e tendo em consideração a disponibilidade ou não de itens lexicais específicos, é seleccionado o lema (De Bot & Schreuder 1993:205).

Poulisse 1993 e Poulisse & Bongaerts 1994 resolvem a questão de outro modo. Defendem que a escolha da língua é feita no conceptualizador, porque essa escolha é influenciada pelo conhecimento da situação e esse tipo de informação só está disponível no conceptualizador. A especificação da língua assume a forma de uma componente dessa mesma língua que desempenha um papel na activação dos itens lexicais. Em consequência, defendem, seguindo De Bot 1992 ⁹⁷, mas ao contrário de De Bot & Schreuder 1993, que a mensagem pré-verbal gerada no conceptualizador já contém os *chunks* com informação sobre a língua seleccionada (Poulisse 1993:176). Ou seja, a informação conceptual contida na mensagem pré-verbal, porque já contém informação linguística específica, é transferida directamente para o formulador não exigindo qualquer outra forma de processamento para activar os lemas. Isto evita a adição do componente VBL (Poulisse 1997).

Ao distinguirem estruturas conceptuais de representações semânticas, estes autores adoptam a **three-store hypothesis** de Paradis (1978, 1980, 1985) ⁹⁸. Esta hipótese considera que o significado e a forma das palavras (propriedades semânticas e forma fonológica) constituem a **representação lexical**, juntamente com um terceiro nível de **representação conceptual**, independente da linguagem, e, portanto, das línguas. Ou seja, o falante bilingue dispõe de um léxico para cada uma das línguas e de um arquivo conceptual autónomo. Um falante de duas línguas dispõe assim de três sistemas isoláveis e inter-relacionáveis. Os conceitos, uma vez

⁹⁷ De Bot sugere que a mensagem pré-verbal contém informação linguística para ser lexicalizada pelo formulador. «De Bot [1992] suggests that whereas macroplanning is language-independent, microplanning is language-specific.» (Green 1993:266).

⁹⁸ «In the 1960s and 1970s, the question was whether the bilingual subject possessed two lexical stores, a separate one for each language, or one common store where both languages would be indiscriminately stored. Much contemporary modeling of **bilingual memory** distinguishes between two levels of representation for each language, one lexical, storing word forms, and the other conceptual, storing meanings.» (Paradis 1997:335). «Evidence from paroxysmal aphasia establishes the dissociation between semantic and conceptual representations.» (Paradis: Seminário, Lisboa, Abril de 1998).

adquiridos, são armazenados e podem ser acedidos independentemente da linguagem. «A speaker of two languages, therefore, has three interconnected isolable systems» (Paradis 1997:335-7).

2.2.2.1. *O léxico bilingue*

Seguindo Weinreich (1953/1974), os investigadores costumam distinguir entre três diferentes tipos de **relação entre os dois léxicos e o sistema conceptual**: composta, subordinada e coordenada. A composta e a subordinada assumem um único sistema conceptual partilhado pelos dois léxicos; no caso da composta, as palavras da L2 acedem à sua representação conceptual directamente, enquanto que na organização subordinada o acesso à memória conceptual faz-se através da palavra correspondente na L1. A coordenada assume a existência de dois sistemas conceptuais, cada um deles associado a cada um dos léxicos ⁹⁹ (cf. A de Groot 1993:110). Esta proposta tem sido largamente discutida e muito trabalho experimental tem sido realizado à volta dela. Alguns investigadores encontraram evidência para defender que diferentes **contextos de aquisição** conduzem a diferentes sistemas (por exemplo, Lambert *et alii* 1958:240, citado por A de Groot 1993). Mas Hummel 1993, entre outros, chama a atenção para que o tipo de tarefa e de material usado na experimentação terá, com certeza, consequências nos resultados ¹⁰⁰. Ou seja, será que diferentes tipos de palavras produzem diferentes padrões de resposta? De Groot 1993 afirma que há muito poucos estudos que tenham em consideração o tipo de palavras usadas na experimentação. Há razões, afirma, para acreditar que as traduções de pares de **palavras concretas** apresentem mais semelhanças do que as de **palavras abstractas** ¹⁰¹; e que este facto terá certamente reflexos no modo como estas palavras estão representadas, pelo menos, no léxico de alguns bilingues. Assim, para as

⁹⁹ Ervin e Osgord (1954) propuseram que a **organização da L2** dependeria dos contextos de aquisição e de uso. Sugerem que os bilingues se distribuem ao longo de um *continuum* que vai de um sistema composto puro (*pure compound system*) até um sistema coordenado puro (*pure coordinate system*). (em De Groot 1993:34)

¹⁰⁰ No que respeita a **organização do léxico** em bilingues, Hummel 1993 conclui: «lexical tasks are likely to lead to quite different results».

¹⁰¹ Ellis & Beaton 1993 consideram que mais importante do que a distinção entre concreto e abstracto é a existência ou não de uma representação imagética associada ao conceito.

palavras concretas poderá ser encontrada uma representação conceptual comum para as duas línguas, uma representação composta; enquanto que as palavras abstractas aparecerão como dependentes das línguas, com uma representação coordenada (de Groot 1993:41). Além disso, segundo a *dual coding theory* de Paivio, as palavras abstractas não estariam representadas no sistema imagético. No mesmo texto, a autora refere ainda que o facto de se tratar de **cognatos** ou **não cognatos** também conduz a diferenças no armazenamento (de Groot 1993:45). Não cognatos estão representados como unidades específicas na memória conceptual, enquanto que cognatos partilham uma representação conceptual (de Groot & Nas 1991). É preciso, portanto, mais investigação, com um maior controle do material usado como estímulo (de Groot 1993:45).

Kees de Bot e colegas (1995) estão de acordo com esta necessidade. Afirmam que a relação entre os dois léxicos é um dos temas mais tratados na literatura sobre os aspectos psicolinguísticos do bilinguismo e que, na maior parte das publicações, basicamente, o que os autores tentam defender ou atacar é uma de duas hipóteses: a da interdependência ou a da independência dos léxicos. No entanto, segundo eles, a maior parte da literatura sobre o assunto é bastante imprecisa. Quando falam de armazenamento independente ou partilhado, a que é que se estão a referir? ao processamento? aos sistemas conceptuais? aos léxicos? ou a uma combinação de tudo isto? E, para aclarar a situação, também eles se propõem discutir esta questão dentro do quadro de Levelt. Ou seja, distinguir entre **conceitos, lemas e lexemas**. Haverá ou não conceitos, lemas e lexemas específicos de cada uma das (duas) línguas?

Quanto aos conceitos, já vimos que, neste quadro, eles são independentes das línguas. No que respeita o processamento lexical, a activação de um lema (que contém especificação semântica e sintáctica) pode ser ou não distinta da activação do lexema (que também envolve a forma fonológica) ¹⁰². Ou seja, se tivermos em consideração as palavras cognatas, não é claro em que medida é que os lexemas são ou não

¹⁰² Marslen-Wilson & Nix & Gaskell 1995 propõem que a **representação lexical** é altamente abstracta e que, no processamento *on-line*, interage com processos de inferência fonológica para formarem a representação interna do conteúdo lexical do discurso. O que os conduz a esta hipótese é o facto de uma mesma palavra poder realizar-se foneticamente de um grande número de modos.

específicos de cada língua. A não especificidade, que implica a coincidência absoluta entre as duas formas, pode depender do modo de apresentação, oral ou visual. De facto, formas coincidentes são muito mais frequentes no **escrito** do que no **oral**. Isto sugere, antes de mais, que são necessárias diferentes teorias e modelos para explicar o **processamento visual e o auditivo**. De Bot e colegas afirmam que isto não é uma surpresa. A Neurolinguística apresenta evidência no sentido de mostrar que *input* visual e auditivo estimulam partes diferentes do cérebro e descreve casos de afasia em que funcionamento diferencial entre as duas modalidades foi observado (De Bot *et alii* 1995:18).

É claro que o número de formas coincidentes depende, como é evidente, das duas línguas em causa e da **relação tipológica** entre elas. Por isso mesmo, Grosjean 1989 chamou a atenção para que um bilingue não deve ser considerado como dois monolingues num só corpo, mas sim como um falante/ouvinte específico. O sistema linguístico de uma pessoa bilingue em Inglês e Turco deve ser bastante diferente do de um bilingue em Francês e Espanhol ¹⁰³. E por isso mesmo, o modo de lidar com as representações de cada entrada lexical, na produção e na compreensão, são necessariamente muito diferentes (Grosjean 1989). Além das diferentes línguas representadas na mente do falante (e das relações tipológicas entre elas), outros aspectos que Grosjean 1989 diz ser necessário ter em consideração são o **nível de proficiência** da ou das línguas desfavorecidas, já que o sistema é muito mais instável que o da L1 de um sujeito adulto, e a **idade** em que foi / foram aprendida(s). Por tudo isto, se os bilingues estabelecem diferentes tipos de sistemas ou se, em certos aspectos, podem funcionar como monolingues, é para ele uma questão empírica. Green (1993:250)

¹⁰³ «There seem to be four **patterns of cognacy**. The simplest, and probably the most widespread case, is one where the L1 and the L2 share very few cognates: English and Arabic, Turkish and Dutch, Finnish and Swedish, for example. [...] The second simple case is found where the L1 and the L2 are closely related, and the bulk of the vocabularies of the two languages are mapped by some simple phonological rules. Italian and Spanish, Russian and Polish, Danish and Swedish, are obvious examples of this pattern (though perhaps less clear cut in practice than one might imagine). [...] The first mixed case is typified by English and the Germanic languages. In this case, a very large proportion of the basic vocabulary of the L2 is cognate with the L1, but the rarer, more difficult items in the L2 are noncognate with the L1. [...] The second mixed case is found where the cognates words do exist, but where their use is restricted to certain domains or registers. This happens, for example, in the case of Romance speakers learning English: Romance words in English tend to be low frequency items used in formal situations.» (Meara 1993:285).

acrescenta que, de facto, há estudos que apontam no sentido de que indivíduos que aprendem uma L2 quando crianças estabelecem representações fonéticas-fonológicas distintas para cada uma das línguas; e, neste aspecto, funcionam como dois monolingues num mesmo corpo. Enquanto que os que aprendem uma L2 em adultos estabelecem um sistema diferente de qualquer um dos grupos monolingues. Ou seja, quanto mais equilibrado é o grau de bilinguismo maior parece ser a independência no que respeita o armazenamento dos dois léxicos (de Bot 1995:16).

Admite-se, em geral, que o hemisfério esquerdo é dominante para a linguagem na maior parte dos monolingues. A investigação em Neurolinguística do bilinguismo tem procurado eventuais diferenças entre sujeitos monolingues e bilingues no que respeita a intervenção de cada um dos hemisférios na linguagem. Quanto à questão da **localização**, isto é, saber se cada uma das línguas está localizada ou não num determinado ponto do hemisfério esquerdo, Paradis diz-nos: «One would have to assume that languages change place over time». E quanto à questão da sua **organização no hemisfério esquerdo**, uma em relação à outra, (em um, em dois, ou em três sistemas, ou então em dois subsistemas), segundo ele, é preciso mais investigação (Paradis 1997:340). A hipótese da representação num único sistema (*the extended system hypothesis*) sustenta que as línguas estão armazenadas na mesma região do córtex; isso implica que, à medida que a L2 vai sendo adquirida (ao mesmo tempo ou depois da L1) novos elementos vão sendo adicionados ao conjunto disponível para a L1. Estes novos elementos são tratados como variantes da L1: alofones, alomorfes, e variantes sintáticas são processados como se de variantes estilísticas se tratasse. A hipótese que defende a existência de dois sistemas (*the dual system hypothesis*), pelo contrário, assume que os elementos das várias línguas estão armazenados separadamente em sistemas independentes, o que não significa que não estejam localizadas na mesma área do cérebro. Para a hipótese dos três sistemas (*the tripartite hypothesis*), os elementos comuns às duas línguas estariam representados num sistema e haveria dois outros sistemas com os elementos diferentes de cada uma das línguas. Por último, a hipótese do subconjunto (*the subset hypothesis*) estabelece que as duas línguas são servidas por dois subsistemas do grande sistema que é a competência linguística implícita. Segundo Paradis

(1997:342), esta última hipótese parece ser compatível com todos os padrões de recuperação de afásicos bilingues.

Tendo em conta o estado actual dos conhecimentos, duas coisas podem concluir-se. **Os bilingues não constituem um grupo homogêneo e vários são os factores que podem ter impacto no processamento e na representação das línguas: o contexto e a idade de aquisição, a relação tipológica entre elas, o nível de proficiência, o contexto de uso e até o grau de envolvimento afectivo.** Quanto à organização do léxico bilingue (coordenada, composta ou subordinada), isso pode depender do contexto de aquisição. A frequência de uso de um dado item tem com certeza influência na sua activação e, por conseguinte, na sua disponibilidade. O nível de proficiência determinará o grau de necessidade de recurso a estratégias compensatórias. E o grau de motivação determinará o maior ou menor envolvimento do sistema límbico (Paradis 1997:348). Como corolário desta conclusão, a uma outra podemos chegar, secundando Bialystok:

One of the crucial keys to understanding how language is learned will be to understand how language is represented, and it seems that linguistic representation will turn out to be more complex than we had imagined (Bialystok 1994:568).

Para além das questões a que a Neurolinguística tem procurado responder, e que dizem respeito à relação entre os dois léxicos e sua representação no cérebro, outras questões, mais especificamente relacionadas com cada um dos subconjuntos do léxico mental de um falante bilingue, e relevantes para este trabalho, se podem pôr. Importante entre elas é, com certeza, saber quais as características que os distinguem. Ou seja, **que relação existe entre dimensão, organização, conteúdo de cada entrada, tipos de ligações entre as várias entradas, e desenvolvimento de cada um dos léxico?** Para lhe responder é preciso considerar e comparar o léxico da L1 com o da L2.

2.2.2.2. O léxico da L1

Vimos que os itens estabelecem, entre si, relações de natureza diversa **dentro de cada entrada e entre as diferentes entradas.** Mas, nem todos os itens são entradas; as formas flexionadas de um mesmo verbo, por exemplo, são itens que pertencem à mesma entrada.

O trabalho dos linguistas, especialmente ao discutirem a relação entre a semântica, a sintaxe, e a morfologia, tem-se centrado, como vimos no Capítulo I, naquilo que está, e como está, representado numa entrada lexical. E hoje parece não restarem muitas dúvidas de que cada uma delas contém basicamente informação semântica e sintáctica, a que se junta informação fonológica, morfológica, ortográfica (no caso de sujeitos alfabetizados), independentemente da forma que, em cada um dos casos, e segundo cada uma das teorias, essa informação pode ter. O léxico mental de um sujeito adulto não só é rico dentro de cada entrada mas também em conexões entre as diferentes entradas. São relações paradigmáticas e sintagmáticas que se estabelecem e que criam essa rede de conexões. As primeiras são de tipo semântico e são aquelas que se estabelecem entre membros intersubstituíveis da mesma categoria gramatical; as outras são de tipo sintáctico e são aquelas que se estabelecem entre diferentes expressões de diferentes categorias gramaticais. Relações paradigmáticas estabelecem-se entre sinónimos, antónimos, hipónimos, e hiperónimos ou super-ordenadores. Por exemplo, *rosa* e *leão* são hipónimos, respectivamente, dos hiperónimos ou super-ordenadores *flor* e *animal* (Lyons 1995:124). Mas a activação de um item lexical também pode resultar de conexões de tipo imagético, enciclopédico ou afectivo ou de outro qualquer que esteja associado à entrada em causa. Tudo isto faz parte do conhecimento que um falante nativo adulto tem de uma palavra. Assim, para um falante nativo adulto, saber uma palavra consiste em saber os seus possíveis significados, as suas extensões metafóricas, as suas restrições pragmáticas, a categoria a que pertence e as restrições sintácticas a que está sujeita numa frase ou num enunciado, as suas características fonético-fonológicas, a sua estrutura morfológica, formas derivadas e flexão, e as suas relações paradigmáticas e sintagmáticas com outras. E, no caso de sujeitos escolarizados, a tudo isto se junta saber escrevê-la. No entanto, não é obrigatório que para todas o conhecimento seja igualmente rico. Essa riqueza varia de palavra para palavra e de indivíduo para indivíduo. Há, portanto, para além do factor frequência de uso, uma série de factores intralexicaís que podem afectar a sua ‘aprendibilidade’: os seus traços semânticos, a sua categoria sintáctica, ser

mais fácil ou mais difícil de pronunciar, e apresentar maior ou menor complexidade flexional (ver, entre outros Laufer 1997).

Vimos antes que nem todos os itens são entradas, mas vimos também que não é fácil determinar quais os itens que são entradas. Esta dificuldade arrasta consigo uma outra: estabelecer a dimensão do léxico mental de um falante de uma língua particular. Exactamente por isso, são muito díspares os números fornecidos por cada um dos investigadores interessados na questão e é muito difícil responder a questões do tipo: Quantas palavras tem uma língua? Quantas palavras sabe um adulto falante nativo dessa língua? De quantas dispõe uma criança nos primeiros anos de vida? Eve Clark (1993:13-14) diz-nos que, entre os 2 e os 6 anos de idade, uma criança adquire 10 palavras novas da sua L1 por dia, até chegar mais ou menos a 14.000; entre os 10 e os 11 anos, encontra aproximadamente 10.000 palavras novas; entre os 9 e os 15 é exposta a uma média de 85.000 raízes/bases diferentes e mais ou menos 100.000 significados distintos nos textos escolares; até aos 17 anos, adquire mais ou menos 3.000 palavras novas por ano; assim, um adulto falante nativo de Inglês dispõe de uma produção vocabular entre 20.000 e 50.000 formas, e, para a compreensão, este vocabulário é consideravelmente mais vasto. Aitchison (1994:5), por seu lado, diz que «children pick up words like a magnet pick up pins - possibly over ten a day», de modo que, com dois anos, pode usar à volta de 500 palavras, com três anos 1.000, e com cinco para cima de 3.000; e estes números são muito mais modestos do que os que se referem aquelas que é capaz de compreender; com seis anos, por exemplo, pode compreender pelo menos 14.000 palavras; um adulto educado pode conhecer 150.000 palavras e ser capaz de usar 90 % destas. Outros investigadores chegam a números menos ambiciosos: Goulden e Nation e Read (1990) e Zechmeister e colegas (1993) calculam entre 14.000 e 17.000 o número de palavras de que dispõe um estudante universitário. Se assim é, concluem, um jovem com 18 anos aprendeu uma média de 3 palavras por dia e não 60. Estas grandes disparidades resultam do que se considera uma palavra ou uma entrada e da metodologia usada na estimativa. O método mais usado tem consistido em seleccionar uma amostra de um dicionário e perguntar aos sujeitos envolvidos na experiência quais daquelas palavras conhecem. Mas é claro que os

resultados dependem muito das características da amostra seleccionada (ver Nation 1993).

Perante uma tal dimensão, para que o **acesso** ¹⁰⁴ a um determinado item lexical seja rápido, tanto para a compreensão ¹⁰⁵ como para a produção, o léxico mental tem de ter uma qualquer forma de **organização**. A organização e a acessibilidade parecem depender de uma série de características das próprias palavras e da rede de conexões que se estabelece entre elas. Ou seja, de factores intralexicaes (como a classe gramatical, características fonético-fonológicas, e do facto de serem concretas ou abstractas) e, claro, da sua frequência de uso e, portanto, do seu nível de activação. Resulta claro, portanto, que dificilmente se pode comparar o léxico mental com um dicionário, tanto pela sua organização como pela sua riqueza.

2.2.2.3. O léxico da L2

Da dimensão do léxico da L2 depende com certeza a sua organização e acessibilidade. Testes de associação de palavras têm procurado encontrar as diferenças entre a organização do léxico da L1 e da L2. Inicialmente, os investigadores defendiam que a organização do léxico de um falante de L2 se assemelhava à das crianças falantes de L1, já que em ambos os casos eram frequentes associações sintagmáticas e '*clang associations*', isto é, respostas que resultavam de associações com certos traços fonológicos da

¹⁰⁴ Uma boa parte do trabalho sobre **acesso e reconhecimento lexical** foi feito, inicialmente, usando como estímulo palavras escritas e isoladas. É o caso do *logogen model* de Morton 1979 que propunha que cada palavra armazenada no léxico mental continha informação fonológica e semântica. Investigação usando como estímulo a palavra oral tem, ultimamente, merecido a atenção de muitos. É o caso do *cohort model* de Marslen-Wilson 1989. Alguns investigadores, entretanto, também têm tentado desenvolver modelos que integrem a componente visual e oral do léxico. É o caso do modelo conexionista de Seidenberg e McClelland 1989. Forster 1992 comenta que foi realizado muito trabalho durante os últimos trinta anos, por isso, distinguir e classificar os vários modelos de acesso lexical é uma tarefa quase impossível, já que cada um deles ficou mais complexo e desenvolveu zonas de sobreposição com outros. Forster inclui numa mesma grande família o "logogen model" e os modelos conexionistas e comenta: «this theories are tied together by their shared commitment to a totally parallel detector-based system, and to the notion of spreading activation». Noutra grande família - "lexical search theory" - inclui, entre outros, o *cohort model* e *the two-stage bin model* de Forster: «what ties these theories together is a collective commitment to the analogy between lexical access and a computer-implemented information-retrieval system.» (Forster 1992:75-76).

¹⁰⁵ Gerrig 1986 defende dois modos de **acesso lexical**: limitado e exaustivo (cada entrada lexical contém diferentes tipos de informação, mas uma parte dessa informação não precisa de ser acedida em todas as situações de compreensão).

palavra estímulo sem que houvesse qualquer relação semântica entre o estímulo e a resposta (ver Meara 1980, 1983) ¹⁰⁶. Mas, actualmente, consideram que as semelhanças e diferenças estão muito mais relacionadas com a dimensão e com a profundidade do conhecimento do léxico do que com o factor idade. No entanto, as respostas em L2 tendem a ser menos homogêneas do que as respostas de um grupo comparável de falantes nativos ¹⁰⁷. Isto acontece, com certeza, porque o léxico da L2 é menos estável, embora, como sempre, essa instabilidade dependa também do nível de conhecimentos que o falante tem em relação à L2.

Só se aceitarmos como mais realistas os números mais modestos propostos para a L1, há esperança de um falante que comece a aprender uma L2 quando adulto poder vir a ser um bilingue linguisticamente equilibrado. Por outro lado, e como se pode concluir de tudo o que ficou dito até aqui, para dispor de um léxico organizado, é preciso muito mais do que saber um de vários possíveis significados de uma palavra. Por isso mesmo, os métodos usados para avaliar a dimensão do vocabulário em L1, quando transpostos para a **avaliação da dimensão do vocabulário em L2**, podem dar resultados muito enganadores do conhecimento lexical. Viberg (1993:340), referindo-se à L2, afirma que «the fundamental problem with the lexicon is its size.» Ou seja, **a relação entre dimensão do vocabulário e profundidade do conhecimento lexical tem de ser avaliada tendo em consideração o nível de proficiência linguística do falante, o canal e a relação tipológica entre as duas línguas.**

Numa fase inicial da aprendizagem da L2, a uma dimensão relativamente restrita do vocabulário, quer para a compreensão quer para a produção, corresponderá, com certeza, um léxico pouco estruturado. À

¹⁰⁶ «It is customary to claim that **word association** responses generally fall into two main classes called syntagmatic associations and paradigmatic associations. [...] Personally, I have found that this distinction is very difficult to work in practice, especially when you cannot refer back to the testee for elucidation, but this difficulty is not generally commented on in the literature.» (Meara 1983:1-2).

¹⁰⁷ Fiz uma pequena experiência de **associação de palavras** com falantes de português L1 e L2, que, devido à sua pouca extensão e ao pequeno número de sujeitos envolvidos, não pode ser considerada como representativa e muito menos conclusiva. No entanto, várias respostas, em particular dos falantes menos proficientes, foram *clang associations*. Quanto a heterogeneidade, talvez por estarem em contexto de aprendizagem semi-formal, ou pelos estímulos escolhidos, não se verificou ser muito maior que a das respostas dos falantes de L1.

medida que os diferentes sistemas (fonológico, sintáctico, semântico, etc...) vão ficando estabelecidos e que a profundidade do conhecimento vai aumentando, o esforço de aprender uma palavra nova pode ser cada vez mais o que corresponde a aprender o seu significado e as suas eventuais extensões metafóricas e a introduzi-la na rede de conexões já estabelecida. Mas é claro que, se tivermos em consideração um bilingue com duas línguas muito próximas, como é o caso do Português e do Espanhol, o **léxico disponível para a compreensão**, porque usa como fonte o léxico da L1, será muito mais extenso do que **para a produção**. Francine Melka 1997 cita uma série de estimativas, a maior parte delas datadas dos anos 70, que concluem que o vocabulário receptivo é, pelo menos, o dobro do vocabulário produtivo em bilingues de Inglês-Francês, mas que esta distância vai diminuindo à medida que o tempo de exposição vai aumentando. No entanto, há outros que propõem uma diferença muito menor. Melka (1997:93-96) apresenta algumas razões para estas grandes discrepâncias: os itens incluídos na amostragem, os critérios para considerar uma resposta como certa ou errada, a escolha do contexto em que o item aparece inserido, o controle ou não do uso de estratégias de evitação, e o papel dos cognatos. Por isso, e por tudo aquilo que envolve saber uma palavra, também ela acha que a distinção entre vocabulário receptivo e produtivo deve ser evitada ¹⁰⁸. Assim, **qualquer estimativa do conhecimento de vocabulário / léxico só terá resultados fidedignos se avaliar cada uma das componentes que constituem saber uma palavra** (em especial, a fonético-fonológica, a semântica, e a sintáctica), **separadamente, e tanto para o conhecimento receptivo como para o produtivo, tendo em conta o canal (oral ou escrito)**.

O que quero dizer ficará muito mais claro se pensarmos num par de línguas tão afastadas como são o **Português e o Chinês**. Para o processamento de discurso oral, aquilo que o ouvinte/falante recebe é uma cadeia de sinais sonoros a que tem que associar, antes de mais, um conteúdo semântico. Ou seja, a **componente fonético-fonológica** do léxico

¹⁰⁸ «Paribakhat and Wesche (1993) developed a **Vocabulary Knowledge Scale** (VKS) which attempts to capture stages of knowledge. The VKS is a five-level scale in which the subjects' responses range from total unfamiliarity to the ability to use words (semantically and syntactically) accurately in sentences.» (Melka 1997:98).

funciona como a porta de entrada para a componente semântica e, portanto, para a eventual descodificação da mensagem. Essa componente lexical constitui conhecimento procedimental ou implícito ¹⁰⁹ que, por isso mesmo, no caso de línguas com um sistema fonológico bastante diferente ¹¹⁰ e com um grau de cognaticidade nulo, demora muito tempo e esforço até ser adquirido e automatizado ¹¹¹.

Em todas as variedades linguísticas conhecidas pela designação genérica de Chinês, a distinção entre sílabas e morfemas parece ter um peso diferente na modalidade oral e visual. O facto de ser uma língua tonal faz com que um mesmo segmento tenha diferentes realizações fonéticas e, portanto, significados diferentes, e representações escritas também diferentes. Zhou e Marslen-Wilson (1994:398) são da opinião que, quando uma palavra composta (geralmente dissilábica) é apresentada visualmente, a relação de um para um entre caracteres e morfemas é, com certeza, mais relevante para o acesso ao léxico do que a unidade sílaba. Mas que, quando é apresentada oralmente, tanto a sílaba como o morfema devem ser factores importantes no acesso: a sílaba porque a ambiguidade das formas fonológicas no *input* pode levar ao acesso de múltiplos morfemas e palavras, e o morfema porque, em qualquer ponto do processo, tal como na modalidade escrita, o morfema alvo tem de ser identificado pelo ouvinte. Este facto tem certamente consequências na representação lexical destes falantes, e também no modo como abordam a fala noutras línguas.

¹⁰⁹ Ellis (1997:122) defende que «the acquisition of a words form, its I/O lexical specifications, its collocations, and its grammatical class information all result from predominantly unconscious (or **implicit**) **processes** of analysis of sequence information; the acquisition of a word's semantic and conceptual properties, and the mapping of word form labels onto meaning representations, results from conscious (or **explicit**) **learning processes**.» Ver também Nick Ellis (1994, 1995, 1996). Tratarei desta questão no Capítulo 3, *Léxico e aquisição de L2*.

¹¹⁰ «[I]t is surprising to find so little work done so far on the possible **effects of the tonal properties** of Chinese on the learning of English speech contours. Intuitively, it seems plausible that speakers of Chinese, a tone language, would have trouble with English, a language in which tone is non-phonemic. [...] a Cantonese learner of English may be 'predisposed' to hear as several words the same English word uttered with different intonation patterns.» (Wong 1988:3)

¹¹¹ É claro que outras diferenças tipológicas também são responsáveis pelas dificuldades no processamento. Wong (1988:6) destaca, entre outras, o facto de a frase canónica em Chinês ser **tópico-comentário**, ao contrário do Inglês, ou do Português, que são línguas de sujeito-predicado.

Como é sabido, e para agravar a situação, o PE caracteriza-se por um débito muito rápido e por um enfraquecimento sistemático do vocalismo átono, o que torna a segmentação de unidades na cadeia sonora particularmente complexa. Se a componente fonético-fonológica, que constitui a porta de entrada, estiver bloqueada, ficaremos com a ilusão de que o conhecimento lexical de todas as outras componentes é inexistente, o que pode não coincidir com a realidade.

Pelo contrário, no acesso a **discurso escrito**, a componente fonético-fonológica pode não interferir ¹¹². É possível associar à imagem gráfica o significado, que depende sobretudo da memória e que constitui conhecimento declarativo ou explícito ¹¹³. Além disso, e como se sabe, os falantes do nosso exemplo, os chineses, devido ao sistema de escrita da L1, desenvolvem bastante mais a memória visual do que os falantes de escritas alfabéticas, o que lhes permite memorizar um grande número de pares imagem gráfica-conteúdo semântico em relativamente pouco tempo ¹¹⁴. E isso, juntamente com o que disse antes sobre sílaba vs. morfema, explica o muito mais elevado grau tanto da compreensão como da produção de texto

¹¹² «There is now abundant evidence that **phonemic mediation** is not necessary to obtain access to meaning even with an alphabetic orthography [...] One implication [...] from this evidence is that phonemic mediation is just one of the strategies for obtaining access to meaning, rather than an obligatory stage.» (Tzeng & Hung & Wang 1977). Ver também Castro 2000 e Castro-Caldas 2000.

¹¹³ «Chinese speakers confronted with English may simply apply what they do in **learning to read in Chinese**, i.e., rote memorization of the meaning(s) associated with a word of a certain configuration, to the new system, which is inefficient but 'safe' » (Wong 1988:13)

¹¹⁴ Actualmente vigoram **dois tipos de escrita**: uma logográfica - *báihuà* - e uma outra alfabética-fonética, *pinyin*, adoptada em 1957 e usada com fins pedagógicos, ambas baseadas no *putonghuà*. Inicialmente, há mais ou menos 2000 anos, o sistema de escrita logográfica *wényán*, embora com um estilo muito elaborado, estava relativamente próximo da língua falada na época, mas, à medida que a língua foi mudando, a distância foi-se acentuando. O *báihuà* surge mais tarde como um novo sistema de escrita para funções culturais consideradas mais baixas e o *wényán* fica reservado para a literatura. A reforma escolar, levada a cabo entre 1910-1920, para que aprender e usar a língua escrita fosse mais fácil para as massas, adoptou o *báihuà* como sistema de escrita para todas as funções. O maior objectivo desta reforma consistia em eliminar, ou pelo menos reduzir, a distância entre a língua falada e a escrita. De qualquer modo, só um número pouco significativo de falantes de variedades do sul sabe *putonghuà*, particularmente em Macau e Hong-Kong. É certo que, nesta área, as normas para a escrita apresentam algumas diferenças inevitáveis não só no léxico, mas também na gramática (Chen, 1993). Embora não seja correcto assumir que os caracteres chineses são todos pictográficos, no entanto, a relação fonética-escrita pode ser considerada, no mínimo, como muito opaca (Hasuike, 1986). Por outro lado, o sistema alfabético *pinyin* adequa-se ao dialecto dos falantes de Pequim, mas não aos outros, muito em particular aos do sul.

escrito na L2 manifestada pela, atrevo-me a dizer, totalidade dos falantes de Chinês. Diz Meara:

My guess is that lexical access is in fact a key element in most language skills, playing a particular important role in listening comprehension and in oral fluency. Learners whose lexical access skills are poor would be expected to perform badly on tasks which involve real-time processing. (Meara 1993:293)

Este exemplo, com línguas tipologicamente muito afastadas, parece ser bastante elucidativo da necessidade de ter em consideração a qual dos canais e a qual das componentes nos estamos a referir, quando falamos de conhecimento receptivo e produtivo. Por outro lado, ele faz luz sobre todos os casos de pares de línguas que se situam ao longo do *continuum* de distância linguística entre si. Quero com isto dizer que aquilo que consideramos como vocabulário receptivo, na modalidade oral, é determinado, pelo menos em fases iniciais de aprendizagem da L2, pela capacidade que o falante tem para descodificar a fala e que este é resultado de habilidade pessoal, mas também, e em grande medida, do apoio que a L1 pode facultar.

Na verdade, não há dúvida de que uma parte do contacto que muitos falantes bilingues têm com o léxico da L2, nomeadamente os que nos interessam aqui, porque em situação de aprendizagem semi-formal, se faz por exposição a material escrito. Sendo assim, a porta de entrada para as outras componentes não é fonético-fonológica, mas **visual**. Vários estudos sugerem que **o sistema de escrita** usado para a transcrição da L1 desempenha um papel importante no processamento lexical da L2 e que diferentes sistemas de escrita, e até mesmo diferentes línguas com o mesmo sistema, requerem diferentes estratégias. Um desses estudos foi realizado por Green & Meara 1987 e envolvia três sistemas diferentes: alfabético, silábico e logográfico. Os sujeitos (espanhóis, árabes e chineses) utilizaram estratégias de processamento diferentes ao realizarem uma tarefa de busca visual na L1 ¹¹⁵. Resultados de um outro estudo (Suarez & Meara 1989) mostraram que essas mesmas estratégias parecem ser usadas para o processamento de material do Inglês como L2. Os espanhóis tentaram usar uma abordagem fonológica para ler palavras inglesas. Ann Ryan (1997:183)

¹¹⁵ Ryan & Meara 1991 falam, no caso dos árabes, de “vowel blindness”.

faz notar que, ao contrário do Espanhol, que tem uma ortografia bastante regular, o Inglês apresenta um duplo código: tem propriedades fonologicamente regulares, como o Espanhol, e logográficas, como o Chinês. Ryan comenta que esta forma de conhecimento procedimental, específico para a L1, deve ser encarado como um caso especial de *transfer*.

Koda afirma que os sistemas de escrita da L1 têm efeitos profundos e duradouros no modo como os materiais da L2 são processados e acrescenta que o uso de mecanismos de processamento da L1 que são tipologicamente semelhantes terão como resultado uma performance melhor e mais eficiente (Koda 1997:38) ¹¹⁶. Segundo Ping Chen (1993:525) os leitores chineses prestam mais atenção ao significado transmitido pela forma gráfica dos caracteres do que ao seu valor fonético, por isso, não é raro que reconheçam o conteúdo semântico de caracteres sem serem capazes de os pronunciar. No entanto, Hasuike, Tzeng e Hung (1986:284) lembram que vários investigadores têm chamado a atenção para o facto de, actualmente, 90 % dos caracteres chineses serem fonogramas e não pictogramas ¹¹⁷, o que constitui um argumento contra a hipótese de a leitura de caracteres poder dispensar a conversão grafema - fonema. Mas acrescentam que, de facto, a relação entre o discurso e a sua transcrição escrita não é de modo nenhum tão clara como no caso dos sistemas alfabéticos. Para Koda 1997, o facto de a L1 de um indivíduo não usar o **sistema alfabético** constitui um sério '*handicap*' tanto para a leitura como para o processamento lexical em

¹¹⁶ Estudos sobre afásicos japoneses, chineses e coreanos têm discutido os **efeitos dos sistemas de escrita no processamento**. Sasanuma & Fugimura 1972 com base em evidência de 50 afásicos e 30 não afásicos, japoneses, concluem: «these findings provide strong support for the hypothesis [...] that kana [sistema fonético e silábico] and kanji [pictórico, como o chinês] transcriptions can be processed in different modes, the former requiring some phonological device and the latter bypassing it.»; Park & Arbuckle 1977 em estudo sobre «effects of script on memory in 'biscrptural' korean subjects», concluem: «the findings are consistent with the japanese claim that there are fundamental differences in how the two systems are processed»; no entanto, Hasuike & Tzeng & Hung 1986 afirmam que devemos ser cautelosos quanto a haver efeitos da escrita na lateralização. Galloway 1983:103 ("résultats de l'analyse d'environ 400 cas cliniques d'aphasie chez les polyglottes") considera como uma das "hypothèses de travail valables": «Les orthographes phonétiques paraissent mieux traitées par l'HG [hémisphère gauche]. Les écritures idéographiques sont probablement mieux traitées par l'HD [hémisphère droit].»

¹¹⁷ «Traditionally, six **classes of characters** called 'the six writing' are distinguished [...] Of all Chinese characters, 90 per cent or more belong to the fifth class, being composed of a phonetic and a classifier. This is one reason why Chinese characters are very inadequately described as ideographs, if this notion means that the written symbol refers to an idea rather than to a linguistic unit.» (Coulmas 1999:82).

geral. O processamento ortográfico deficiente conduz a dificuldades no acesso bem como na compreensão. Sendo assim, a experiência ortográfica prévia tem um forte impacto no desenvolvimento de estratégias para o processamento da L2 e, conseqüentemente, na construção do léxico mental.

2.2.3. A formulação da mensagem

Vimos que o formulador recebe como *input* a mensagem pré-verbal e que a sua função consiste em traduzir essa estrutura conceptual numa **estrutura linguística**. Esta tradução realiza-se em duas fases: na primeira, com a informação semântica, sintáctica e morfológica, atinge-se a situação de **lema**, e na segunda, a este se acrescenta depois a informação fonológica, daí resultando o **lexema**.

Sendo o léxico o mediador, para formular a mensagem é necessário procurar no léxico mental itens lexicais que expressem correctamente a mensagem planeada e quanto mais incipiente, e portanto menos automatizado, é o conhecimento da L2 mais tempo é preciso gastar nessa tarefa de acesso ao léxico. O nível de activação de um determinado item influencia também, certamente, a quantidade de tempo necessário ao seu acesso. O alto nível de **activação da L1** ou de outra língua tem, assim, como consequência que o falante pode ter muita dificuldade em desactivar essas línguas quando pretende falar a L2 em causa, visto que, em circunstâncias normais, ou seja, em falantes bilingues que não sofrem de afasia, uma língua nunca pode ser completamente inibida (Paradis 1997:349).

Mas, um outro e mais sério problema se levanta quando o falante descobre que não tem disponível no seu léxico um item que dê conta de todas as especificações de um determinado *chunk*. No entanto, não é claro em que ponto do processo de produção o falante toma consciência deste problema. É possível que, havendo uma ligação entre o conceptualizador e o léxico mental, o falante tenha esse facto em consideração no planeamento da mensagem. Mas, a maior parte das vezes, ele só parece dar-se conta desse *deficit* aquando da codificação gramatical ao nível do formulador. Confrontado com esse problema, o falante usa uma de três estratégias: abandona a mensagem planeada, desistindo do que ia dizer, pede ajuda, ou

reformula a mensagem procurando um caminho alternativo ¹¹⁸. À última destas opções, e aquela que nos interessa, tem-se chamado estratégias comunicativas compensatórias (Poulisse 1993).

2.2.3.1. *Léxico deficitário e estratégias compensatórias*

Estratégias comunicativas tem sido um tópico bastante tratado em aquisição de L2 ¹¹⁹. Nanda Poulisse 1993 revê as tipologias de Faerch & Kasper 1983 e de Bialystok 1990 ¹²⁰ e propõe uma tipologia baseada em Levelt 1989 ¹²¹. Convém referir aqui que Poulisse segue De Bot 1992, ou seja, não introduz o módulo verbalizador e, assim, assume que a escolha da língua é feita no conceptualizador, e que os *chunks* da mensagem pré-verbal contêm informação lexical. O modelo desenhado em Poulisse 1993 apresenta duas grandes vantagens sobre os outros: tem por detrás um ponto de vista psicolinguístico e, porque reduz bastante o número de entradas, é muito mais elegante. A autora propõe três tipos de estratégias compensatórias: **substituição**, **substituição mais estratégia**, e **reconceptualização**. Estratégias de **substituição** são casos em que um ou mais traços do *chunk* da mensagem pré-verbal são substituídos ou omitidos. Poulisse considera que esta decisão é tomada, provavelmente, no **conceptualizador**. São aqueles casos em que o falante substitui o item

¹¹⁸ «The choices are between giving up, looking for assistance and devising alternative means of expression to communicate his message. It is true that speaker need not to be **conscious** of having a problem when he is confronted with deficiencies in his mental lexicon. If an alternative way of expression presents itself immediately, this is, in fact, fairly unlikely. In practice however, it will be impossible to identify cases of CS use unless the speaker is able and willing to say that he knows there is a more preferred means of expression which he would have used if he had been able to do so.» (Poulisse 1993:160)

¹¹⁹ Ver, entre outros, e além dos referidos acima, Cohen 1989 e 1990, Corder 1983/1994, Haastrup 1990, Kellerman 1990, Palmberg 1987, Ringbom 1985 e 1990, Yule & Tarone 1990, e Zobl 1980.

¹²⁰ «Part of the controversy concerning the **definition of strategies** seems to arise from the fact that Faerch and Kasper's notion of CS covers only a subset of what Bialystok considers to be the CS. Faerch and Kasper explicitly restrict the use of the term CS to cases in which the speaker (either native or non-native) attempts to overcome linguistic difficulty, while Bialystok's definition covers all attempts to reach a communicative goal, whether this goal presents the speaker with difficulties or not.» (Poulisse 1993:160).

¹²¹ Esta taxonomia, como salienta Nanda Poulisse, constitui uma **nova proposta** em relação a trabalhos anteriores da autora no âmbito do projecto de Nijmegen. «The distinction between *Conceptual* and *Linguistic Strategies* which was made in the original Nijmegen taxonomy could not be related to differences in the processes underlying these strategies. As a consequence, it had to be discarded.» (Poulisse 1993:183).

desejado por outro. São aqueles a que a literatura costuma referir como empréstimos ou *code-switching* e aproximações ou generalizações. Por exemplo ¹²², e para o primeiro caso, significa substituir *criado* por *camarero* (E18), *segurança* por *seguridad* (E28) ou *mudança* por *cambio* (E08), e para o segundo, *mão-de-obra* por *trabalho* (S 44) ou *semelhante* por *vizinho* (S37). Mas o falante pode estar consciente do déficit e também de que o seu interlocutor conhece a palavra em outra língua e decidir usá-la: «*Eu penso que os portugueses são o povo mais 'easy-going' (não sei como se diz em portuguese) do mundo...*» (C31).

O segundo tipo – **substituição** mais **estratégia** – consiste numa estratégia de substituição combinada com a aplicação de morfologia ou fonética (ou ortografia) da L2 para que o item fique mais de acordo com aquilo que o falante considera ser típico da L2. São exemplos deste tipo: *nervosidade* por *nervosismo* (F33), *regretará* por *lamentará* (F42), *sufisante* por *suficiente* (F48), ou *volontade* por *vontade* (F52). Esta estratégia parece ser, do ponto de vista do processamento, mais complexa e mais exigente do que a anterior, já que ela envolve codificação não automática. Ela implica, por vezes, criatividade morfológica, ou seja, conhecimento procedimental, visto que não parece provável que o falante tenha, pelo menos em muitos casos, o item armazenado, pronto e disponível no seu léxico mental. Assim, é provável que seja o **monitor** a tomar a decisão de aceitar ou rejeitar o item (Poullisse 1993:181).

Estratégias de **reconceptualização** implicam uma mudança na **mensagem pré-verbal** e envolvem mais do que um *chunk*. Elas podem revestir-se de várias formas. Pode ser uma definição aproximada, como em «*Todas <usa> usaram ropas <muchos> muito especial, alguém <ficau> ficou como a bebê, alguém ficou como o 'Cowboy'*» (C27) para resolver a falta da palavra *carneval*; ou «*ontem foi o dia quando os Relogios receberam o novo tempo*» (A09) para dizer que *ontem mudou a hora*; combinar duas palavras para fazer uma em falta, como em «*Depois todos os 'hippies' envolvidos foram 20 anos para <...> 'a casa dos ladrões', inocentes!*» (S26) para suprir a falta de *prisão*; e, por vezes, o falante pode também usar só gestos ou

¹²² Os exemplos são da base de dados (dos *sub-corpora*: A- Alemão, C - Chinês, E - Espanhol, F- Francês e S - Sueco).

combiná-los com discurso. Convém, no entanto, salientar que em certos casos é evidente que o falante tem consciência de que lhe falta a palavra adequada, como mostram as aspas em ‘*casa de ladrões*’, ou pode estar convencido de que a que tem disponível não é aquela que a língua requer e enganar-se nessa sua suposição, como em «...*ele [Schindler] queria salvar as muitas pessoas que era possível dos ‘campos de concentração’ (Não sei a palavra própria)*» (S25). Poulisse (1993:181-2) faz notar que as estratégias incluídas neste grupo, ainda que, à primeira vista, possam parecer muito diferentes, são, de facto, semelhantes, já que todas elas implicam mudanças na mensagem pré-verbal, e que, muitas vezes, estas mudanças são muito mais substanciais do que as usadas em estratégias de substituição. As estratégias de reconceptualização implicam a substituição, adição ou apagamento de *chunks* inteiros, o que, muitas vezes, afecta consideravelmente a estrutura do enunciado. Além disso, e exceptuando o uso de gestos, elas são muito mais exigentes do ponto de vista do processamento do que as de substituição.

A autora salienta que, como é óbvio, o número de estratégias compensatórias usadas depende do grau de proficiência. O falante, quando confrontado com um problema lexical, adere a dois **princípios comunicativos** para o resolver: o *princípio do mínimo esforço* e o *princípio cooperativo* (cf. Grice 1975). O primeiro determina que o falante use a estratégia que lhe exige menos esforço de processamento e o segundo que a mensagem seja compreensível. No entanto, por vezes, estes princípios podem entrar em conflito. A decisão de dar prioridade a um ou outro depende de vários factores impostos pelo contexto comunicativo: exigências impostas pela tarefa, o tempo disponível para a realizar, a complexidade cognitiva desta, o apoio disponibilizado pelo contexto, e o *feedback* fornecido pelo interlocutor (Poulisse 1990:148-150). A necessidade de resolver todos os problemas favorece, certamente, o uso de estratégias de reconceptualização, que, na verdade, exigem um grande esforço, mas a complexidade, as restrições de tempo, o contexto e as características do interlocutor podem favorecer o uso de estratégias de substituição e de substituição mais estratégica. Assim, no caso dos materiais em observação neste trabalho, escritos, em situação de exame, porque a pressão exercida pelo tempo é menor do que seria em produção oral, e porque o tipo de tarefa

favorece bastante a necessidade de ser bem compreendido, é provável que se verifique uma preferência por estratégias de reconceptualização, em particular sempre que o falante está consciente de que o item seleccionado em primeiro lugar pode apresentar uma deficiência de qualquer natureza.

É oportuno fazer aqui uma breve referência a um termo que faz parte da tradição na literatura em L2. Refiro-me a **transfer**. Como se pode verificar, relendo estes últimos parágrafos, ele esteve ausente da discussão. Mas, se o nosso quadro teórico fosse outro (ou não fosse nenhum...) ele teria sido usado, com certeza, com relativa frequência. Pit Corder, do mesmo modo que foi um pioneiro ao chamar a atenção para o peso excessivo que, na década de sessenta, se atribuía à L1, também o foi quando, em meados dos anos oitenta, publica um texto (Corder 1983; reimpresso em 1992) em que discute o papel da língua materna e a utilização, segundo ele indevida, que se vem fazendo de termos como *transfer* ou *interferência*, já que, como refere, estes termos pertencem a uma teoria – a análise contrastiva – e, por isso, só devem ser usados depois de cuidadosamente redefinidos.

A aprendizagem de uma língua não é, diz Corder, um processo cumulativo, como juntar objectos num armazém. É como um botão que vai desabrochando até se transformar em flor. **O papel desempenhado pela L1 é bastante mais penetrante e subtil do que tradicionalmente se tem acreditado. Se há alguma coisa a que se possa, com propriedade, chamar transfer é à transferência de conhecimento implícito da estrutura mental da L1 para a interlíngua, a qual constitui um conhecimento separado e com desenvolvimento independente da L1. Ou seja, transfer é um mecanismo de facilitação que usa por empréstimo itens e traços da L1 como uma estratégia comunicativa e que, quando é bem sucedido, conduz à sua incorporação na interlíngua.**

De facto, aquilo que Nanda Poulisse fez no texto que tenho vindo a citar foi, sem nunca os usar, responder ao apelo de redefinir os termos *transfer* e *interferência*. Dito de outro modo, **ao situarmos a maior parte das estratégias comunicativas ao nível do conceptualizador e do monitor, damos um peso decisivo à L1 enquanto fonte de conhecimento declarativo, enciclopédico e procedimental, adquirido e automatizado ao longo da vida.** Esse conjunto de conhecimentos funciona

como uma ferramenta na descoberta das propriedades formais da L2, facilitando em especial a aprendizagem do que é comum ou semelhante nas duas (Corder 1983/1992:29). Assim, sempre que o termo *transfer* for usado neste texto, é com esta perspectiva por detrás que ele deve ser entendido.

2.2.4. A exteriorização da mensagem

Performance in L2 reflects the nature of the system that has been constructed (Green 1993:270).

As características evidenciadas pelo discurso produzido, num determinado momento, por um falante particular, seja através do canal oral ou escrito, reflectem, de facto, a natureza do sistema que foi construído. Esse sistema desenvolveu-se em função dos conhecimentos linguísticos e não linguísticos prévios e das características pessoais do sujeito, bem como do contexto de aprendizagem. Dito de outro modo, os desvios que se podem detectar no discurso, em relação àquilo que seria nas mesmas circunstâncias a produção nativa, resultam do modo como o sujeito lida com todo este conjunto de processos complexos que temos vindo a seguir (e, com certeza, também de outros que não referimos, ou de que nem suspeitamos...).

Quando pensamos em exteriorização da mensagem, pensamos antes de mais no discurso **oral**. Neste, saltam ao ouvido algumas características elocutórias que nos permitem, na maior parte dos casos, identificar um falante não nativo. Entre outras, as que resultam do estabelecimento ou não de categorias fonético-fonológicas distintas para as duas línguas e, conseqüentemente, uma produção com características idênticas ou não às de um falante nativo.

Bongaerts & Planken & Schils 1995, para responderem à questão de que evidência empírica dispomos para a existência de um período crítico para a aquisição de um **sotaque** nativo numa L2, analisaram alguns estudos disponíveis e concluíram que, em geral, as crianças obtêm, a longo prazo, melhores resultados do que os adultos. No entanto, começar a aprender uma língua antes da puberdade não garante que se atinja um sotaque nativo. O que parece garanti-lo é que o início da aprendizagem

comece antes dos seis anos de idade. Depois dessa idade, essa possibilidade começa a decrescer (Bongaerts & Planken & Schils 1995:34).

De facto, os traços fonético-fonológicos parecem ser a componente mais resistente no processo de aquisição das palavras. No entanto, a questão pode ser considerada de modo mais agudo. Martohardjono & Flynn 1995 distinguem reconhecimento de produção e observam que certos aspectos do desenvolvimento fonológico (isto é, o contraste entre fonemas) podem reflectir capacidades biológicas, enquanto que outros (por exemplo, o ajuste perfeito do sotaque) podem estar relacionados com características específicas dos dados em observação. E Lengyel 1995 acrescenta que parece haver uma interação entre factores fonológicos e semânticos, o que implica que aspectos fonológicos possam ser apercebidos e assimilados segundo critérios que estão para além do domínio estritamente fonológico. Mas adverte que é preciso mais informação sobre os diferentes elementos (traços, segmentos e traços suprasegmentais), sobre a facilidade ou dificuldade com que são adquiridos, sobre as estratégias usadas e sobre a influência da L1 ¹²³ (Lengyel 1995:134).

Mas é evidente que, na maior parte dos casos, o sotaque é só uma das muitas outras características que distinguem a produção destes falantes da dos nativos. Muitas delas são comuns ao escrito e ao oral. Outras são privativas do texto escrito e correspondem, de certo modo, ao sotaque próprio de cada falante ou de cada grupo. Essas perdem-se com a transcrição, seja ela fonética ou ortográfica, dos documentos manuscritos. Refiro-me à organização do texto na página, ao material de escrita utilizado, e às idiossincrasias caligráficas. São elas que nos permitem identificar um texto como tendo sido escrito por um sueco, um chinês, um alemão, ou um eslavo, sem precisarmos de saber o nome e a nacionalidade do seu autor. E esta é também uma forma especial de *transfer*.

¹²³ Odlin (1989:112-127) revê um grande número de trabalhos, disponíveis até então, sobre **transfer** ao nível segmental e suprasegmental.: «Although cross-linguistic influences on pronunciation frequently involve segmental contrasts, the influences are also frequently evident in suprasegmental contrasts involving stress, tone, rhythm, and other factors. Stress patterns are crucial in pronunciation since they affect syllables and the segments that constitute syllables [...] Sutch interactions have important implications not only for speech production but also for comprehension. [...] When non-native speakers do not use a stress pattern that is a norm in the target language, vowels and consonants may vary from the target pattern, and this can result in a total misperception by listeners.» (Odlin 1989:117).

O que farei na Parte II, em particular, no capítulo 6, *Vocabulário disponível e léxico deficitário*, será exactamente observar o discurso produzido, por **escrito**, por vários grupos de falantes de PE L2. Mas, embora cada um desses grupos seja constituído por indivíduos com percursos próprios, eles apresentarão, certamente, muitas características em comum, e apresentarão outras determinadas pelos conhecimentos linguísticos prévios à aprendizagem do Português. É por isso que o critério usado para a organização dos documentos foi a língua materna dos aprendentes.

2.3. Sumário

Seguindo Levelt 1989, fiz o percurso *from intention to articulation*. Pelo caminho, procurei ir reflectindo sobre os problemas que se põem a um falante bilingue quando produz discurso na sua língua desfavorecida e sobre as estratégias a que pode recorrer para os resolver. Fi-lo revendo investigação disponível realizada sobre os aspectos psico e neurolinguísticos do bilinguismo, no quadro deste modelo de produção. Procurei também acrescentar sugestões e críticas vindas de outras perspectivas teóricas. Mas é claro que a aplicação de uma teoria conexionista colocaria o problema de forma diversa daquela que aqui foi seguida. No entanto, um modelo é sempre uma metáfora da realidade ¹²⁴. E este, como disse na Introdução, tem o mérito de fornecer um quadro integrador para as questões que se põem na produção de discurso por bilingues e do modo como a investigação as tem abordado. As questões levantadas são, indubitavelmente, só algumas das muitas a ter em conta na construção de um modelo de processamento para a compreensão e para a produção em falantes bilingues. Mas a variedade de factores envolvidos, e que determinam a natureza do sistema particular de cada falante, alerta-nos para que temos de ser muito cautelosos na hora de fazer quaisquer generalizações.

¹²⁴ Swan (1997:175-176) refere-se assim ao modelo de Levelt 1989: «While this is not without value, there is a tendency for such box-and-arrow models to appear more explanatory than they are, by relabelling processes as if they were causal entities located in the brain. Saying that the brain contains a 'conceptualiser' a 'formulator' and an 'articulator' may amount to little more than using nouns instead of verbs to restate the fact that we think of things, put them into words and say them. (We do not explain what makes it rain by saying that the sky contains a 'rainer').»

Por outro lado, é oportuno realçar, nas várias áreas de conhecimento, um outro percurso, paralelo à dicotomia **vocabulário** vs. **léxico**, actualmente em foco em estudos de L2. Na Psicolinguística, os primeiros estudos sobre representação, organização e acesso ao léxico eram feitos usando como estímulo, visual ou auditivo, sobretudo palavras ou frases isoladas. Actualmente, a tendência dominante vai no sentido de estudos de processamento *on-line*. É reconhecido que esse percurso teve como motor as opções das várias abordagens linguísticas para as quais o léxico deixou de ser uma lista de itens idiossincráticos para passar a ser o repositório da maior parte da informação sintáctica. Considera-se, no entanto, que, no caso da L2, sendo a representação do léxico indissociável da sua dimensão, estudos ditos de vocabulário que observem este aspecto são indispensáveis.

Capítulo 3

Léxico e aquisição de L2

3.1. *Introdução*

Spolsky (1988 e 1989) elenca um conjunto de factores envolvidos na aprendizagem de uma L2. Diz Spolsky que o *contexto social* proporciona **oportunidades de aprendizagem formal e ou informal** que conduzem a *atitudes* de vários tipos e que se apresentam no sujeito como *motivação*. A estas, juntam-se *outras características pessoais*, tais como a *idade*, a *personalidade*, *diversas outras capacidades* e os **conhecimentos prévios**. É a interacção entre o sujeito e a situação que determina os **resultados linguísticos e não linguísticos**. O objectivo de Spolsky é com certeza chamar a atenção para a complexidade de circunstâncias sob as quais uma L2 é aprendida. Mas se tivermos também em consideração tudo aquilo que fica sugerido na palavra ‘interacção’, ou seja, o modo como a L2 é processada e adquirida e como é que ela está representada na mente de um bilingue, vemos que o quadro não é de todo simples.

Certamente por isso, a investigação em aquisição de L2 tem sido abordada de diferentes perspectivas e tem, por conseguinte, recebido contributos de várias áreas da Ciência Cognitiva. Foi por isso que, nos dois capítulos anteriores, procurei dar uma panorâmica, centrada no léxico, de propostas que têm influenciado esta área. Neste capítulo, tal como anunciei antes, procurarei, como nos capítulos anteriores, mostrar quais as principais linhas teóricas que têm orientado a **investigação em aquisição de L2** e como progressivamente ela tem vindo a centrar-se mais no léxico.

De facto, várias ciências da linguagem têm contribuído, segundo a sua vocação, com propostas que tentam responder a questões sobre o modo como um determinado factor pode influenciar o resultado. A Sociolinguística, por exemplo, ao fazer a interface entre sociedade e língua, dispõe de instrumentos apropriados para abordar o modo como o **contexto social** pode determinar a atitude do sujeito face à L2 e, conseqüentemente, a criação de mais ou menos oportunidades de exposição a *input* adequado.

A Linguística Teórica dispõe de ferramentas que permitem relacionar os **conhecimentos linguísticos prévios** do sujeito, nomeadamente a sua L1 ou outras que eventualmente conheça, com a **L2**, e que permitem também guiar a análise do **produto resultante da aprendizagem**. Ciências da cognição, como a Psicolinguística, por seu turno, estão vocacionadas para abordar **o modo como qualquer língua natural é processada, como é adquirida, como está representada na mente humana**, e como é que determinadas características e capacidades individuais (idade, motivação, atenção, inteligência) actuam em todo o processo.

Assim, conforme a área de onde vem cada um dos investigadores em L2, as teorias de aquisição têm dado mais destaque a uns ou a outros destes aspectos, deixando, por vezes, alguns deles completamente esquecidos. Daí que esta disciplina continue a sentir a necessidade de integrar todas estas colaborações, além de, como é compreensível e inevitável, continuar a usar como ponto de referência os avanços feitos na investigação em aquisição de L1. No entanto, mais de meio século depois dos primeiros estudos em L1, também os conhecimentos que esta pode disponibilizar não são suficientemente definitivos para que sobre eles se possam construir teorias incontestáveis. Procurando destacar o léxico, o que farei a seguir será dar uma panorâmica muito rápida da investigação e de como as várias áreas têm contribuído para aquilo que hoje se sabe, ou se julga saber, sobre a aquisição de L2.

3.2. *Os primeiros estudos*

Chomsky, em *Aspects of the Theory of Syntax* (1965), desenvolveu uma teoria, já esboçada em 1959, numa recensão crítica à obra *Verbal Behavior* (1957) de Skinner ¹²⁵, em que defende que o facto de todas as crianças adquirirem gramáticas, comparáveis no essencial, de grande

¹²⁵ N. M. da Silva afirma que a obra de Skinner, em geral, tem sido mal compreendida e “alvo de reacções reflexas estereotipadas”. «[**Verbal Behavior**] é uma obra cujo valor foi obscurecido pela incompreensão de Chomsky quanto aos objectivos de interpretar o comportamento verbal a partir da psicologia da aprendizagem. [...] *Verbal Behavior* pode vir a ter ainda um futuro promissor, quando os psicólogos da linguagem se derem conta que as gramáticas generativas, as teorias da informação, os modelos computacionais, são adequados para a linguística interessada na língua (enquanto sistema), mas inermes para lidar com a *fala*, enquanto comportamento humano em situação concreta.» (N. M. Silva 1991:94).

complexidade, com uma rapidez notável, sugere que os seres humanos estão de algum modo predispostos para isso, e que têm capacidade para manipular dados e formular hipóteses (Chomsky 1959:57). Corder 1967 – em *The significance of learners' errors* – põe a hipótese de o mesmo poder acontecer com os adultos a aprenderem uma L2:

I propose therefore as a working hypothesis that some at least of the strategies adopted by a learner of a second language are substantially the same as those by which a first language is acquired. Such a proposal does not imply that the course or sequence of learning is the same in both cases (Corder 1967/1974:22).

Refere o debate intenso que na altura já punha em causa a hipótese contrastiva, chama a atenção para as consequências que isto pode ter no ensino, e prevê que uma delas pode ser exactamente o desvio da ênfase **do ensino para a aprendizagem**. E, pelo menos em parte, não se enganou. A hipótese forte de *transfer* da L1 é abandonada, e o discurso dos falantes de L2 passa a ser encarado como um **sistema**, ainda que sujeito a muita variação. Corder 1967 chama-lhe '*idiosyncratic dialect*', Selinker 1972 '*interlanguage*', Nemser 1971 '*approximative system*'; ele passa a constituir material preferencial da investigação, material obtido, principalmente, em contexto de aprendizagem informal e semi-formal e, em menor escala, em contexto exclusivamente formal ¹²⁶.

Os primeiros estudos sobre aquisição de L2 foram feitos para o Inglês, no final dos anos 60. Mas é a partir de meados dos anos 70 que revistas como *IRAL (International Review of Applied Linguistics)*, *TESOL Quarterly* e *Language Learning* começam a publicar um considerável número de artigos que reflectem o interesse que a investigação em aquisição de L2 começa a despertar ¹²⁷. Nesta primeira fase, a investigação em L2 caracteriza-se por uma forte dependência dos estudos realizados

¹²⁶ Alguns investigadores chamam por vezes a atenção para certos aspectos que só em contexto de aprendizagem exclusivamente formal poderiam ser investigados. (Ver Leiria, no prelo) Daí, posições como esta: «I would like to suggest here that the FL profession needs cease being only a consumer of SLA research and theory and start becoming an active contributor, making FL learning just as important as FL teaching.» (VanPatten 1990:22). Ver Capítulo 7.

¹²⁷ Alguns destes textos foram reimpressos em Burt & Dulay (ed.)1975 e em Richards (ed.)1974.

sobre a aquisição de L1 e tem como principais pontos de referência Chomsky e Roger Brown.

Alguns dos primeiros trabalhos experimentais da década de 70 centram-se na tentativa de encontrar uma **ordem de aquisição** e de a comparar com a de L1. Susan Ervin-Tripp 1974 põe a questão: «*Is second language learning like the first?*» Analisa dados de aquisição de L2 (aprendizagem do Francês por crianças inglesas em meio natural) e conclui que, em muitos aspectos, o desenvolvimento da **sintaxe** e da **morfologia** segue a ordem encontrada nos estudos de L1. Dulay & Burt 1974 comparam a ordem de aprendizagem, em meio escolar, de onze morfemas do Inglês por crianças cuja L1 é o Chinês ou o Espanhol, e concluem que ela é sensivelmente a mesma para ambos os grupos. Intitulam o texto de «*Natural sequences in child second language acquisition*» e, reflectindo a influência das propostas de Chomsky, afirmam que a aquisição de uma L2 não pode ser encarada como um hábito, mas que a semelhança dos erros e os tipos de erros reflectem aquilo que, para a L1, tem sido referido como uma ‘*creative construction*’ (Dulay & Burt 1974:37). Bailey & Madden & Krashen 1974, «encorajados por estes resultados», dão mais um passo: «*Is there a ‘natural sequence’ in adult second language learning?*», perguntam. Observam adultos, com idades compreendidas entre os dezassete e os cinquenta e cinco anos, de diferentes línguas maternas, aprendendo em meio escolar, e concluem que as crianças e os adultos usam estratégias comuns e processam os dados linguísticos fundamentalmente da mesma maneira, apesar dos seus ‘*backgrounds*’ linguísticos diferentes (Bailey & Madden & Krashen 1974:235). Os autores comentam que, independentemente da ordem por que são apresentados os morfemas gramaticais, os adultos parecem tirar proveito da instrução. E propõem: uma hipótese interessante e testável é verificar se o ensino formal que conduz a melhores resultados é o que segue “a ordem de dificuldade observada” ou o que segue um “*natural syllabus*”; e sugerem que outra hipótese de investigação seria descobrir que aspectos da linguagem seguem uma sequência universal e compreender que factores a determinam.

Contudo, a ordem observada nestes estudos não era exactamente aquela que Roger Brown encontrara para a L1 ¹²⁸. Brown 1973 desenvolvia investigação para a L1 e seguia um caminho diferente de Chomsky. Analisou o discurso de três crianças e estudos longitudinais disponíveis sobre o desenvolvimento da linguagem infantil e concluiu que, embora a produção das crianças seja mínima, elas não só nomeiam objectos mas também usam as palavras para exprimir pensamentos que envolvem esses objectos, o que reflecte bastante capacidade de conceptualização. Observa também que a ordem das palavras usadas pelas crianças não viola, em geral, a ordem do discurso adulto. Na linha de Brown, Hakuta 1976 realiza um estudo longitudinal de uma criança japonesa a aprender Inglês como L2, e diz encontrar, na ordem de aquisição dos morfemas gramaticais, evidência empírica suficiente para a existência de *transfer*, e defende que o produto linguístico analisado resulta de factores internos e externos. Conclui dizendo que a questão de factores internos vs. factores externos no desenvolvimento é um clássico também na psicologia. E para ilustrar a sua afirmação cita um texto de 1907 de Clara e William Stern:

In terms of the contents of his speech, [the child] is neither a pure associative machine nor a sovereign constructor of concepts. Rather, his speech is based on the continuing interaction of external impressions with internal systems, which usually function unconsciously; it is thus a constant 'convergence'. The detailed investigations pertaining to the development of speech and thought should determine the relative participation of both forces and also show how they accommodate each other (Stern 1907, em Blumenthal 1970:86-7).

Apesar do crescente número de estudos, quaisquer generalizações quanto a sequências de aquisição ou quanto ao papel da L1 e a processos de *transfer* é problemática dado que, até então, a variedade das L1 envolvidas é muito restrita e a L2 é, na maior parte dos casos, o Inglês. Hyltenstam 1977 – *Implicational patterns in interlanguage syntax variation* – é um estudo experimental sobre a colocação do morfema de negação em

¹²⁸ Estes estudos misturavam **morfemas** tão diversos como artigo, possessivo, pronome pessoal complemento, 'past tense' regular e irregular, progressivo, e 'be' auxiliar. Zobl & Licerias 1994 pretendem mostrar que a heterogeneidade dos morfemas não conduz a nada de conclusivo quanto à ordem de aquisição.

Sueco como L2 ¹²⁹. Um dos seus objectivos consiste exactamente em alargar a investigação a outras línguas além do Inglês. O autor usa instrumentos da Sociolinguística para a análise dos dados: *escalas implicacionais* (DeCamp 1971), *regras variáveis* (Labov 1969) e a noção de *continuum linguístico* (DeCamp 1971 e Bickerton 1975) e conclui que 160 aprendentes, falantes de línguas muito diferentes (Polaco, Inglês, Grego, Servo-Croata e Persa), e com níveis de escolarização muito diversos, seguem uma **sequência de aquisição** das negativas bastante regular ¹³⁰.

Mas não é em contexto semi-formal de aprendizagem, como os que temos observado até aqui, que os estudos de Sociolinguística e de *pidgins* e crioulos estão mais presentes, mas numa outra linha de investigação, que começava a desenvolver-se paralelamente a esta, que privilegia os estudos longitudinais, e que observa a produção linguística de imigrantes a aprenderem a língua em **contexto informal**. Esta linha acrescenta aos factores cognitivos factores de natureza social ¹³¹, pessoal e afectiva (ver Schumann 1978a) ¹³². *The Pidginization Hypothesis* e *The Acculturation Model* de Schumann são um exemplo da importância atribuída a este tipo de factores na aquisição em contexto informal. Integrado num projecto desenvolvido em Harvard, Schumann observa durante dez meses um trabalhador falante nativo de Espanhol, natural da Costa Rica, que

¹²⁹ O autor afirma que «such regularities have previously been found in sound systems acquisition by second-language learners (L. Dickerson 1975, W. Dickerson 1976).»

¹³⁰ Em Sueco, o morfema de negação segue o verbo nas subordinantes (ex: Kalle kommer *inte* idag = Kalle vem *não* hoje) e precede-o nas subordinadas (ex: Det är shönt att Kalle *inte* kommer idag = É bom que Kalle *não* venha hoje). O teste foi aplicado duas vezes, três e cinco semanas depois do início da aprendizagem formal. Hatch (1978:55) põe duas objecções a este estudo. Uma diz respeito à escolha da tarefa, que não permite muita escolha, e a outra está relacionada com o léxico, nomeadamente, com a questão dos **chunks memorizados**.

¹³¹ «Certain **social factors** can either promote or inhibit contact between the two groups and thus affect the degree to which the 2LL group learns the target language. The first such factor involves social dominance patterns. If the 2LL group is politically, culturally, technically, or economically superior (dominant) to the TL group, it will tend not to learn the target language. [...] The second social factor affecting second language learning involves three integration strategies: assimilation, preservation, and acculturation. [...] Enclosure is the third [...] Attitude is another [...] The final is the length of residence in the target language area.» (Schumann 1978:163-166).

¹³² «Some have criticized creolists such as DeCamp and Bickerton for concentrating primarily on the internal linguistic constraints which affect implicational relationships in the creole continuum. [...] Bickerton however has argued that the basis for the distribution of lectal features is purely linguistic. He does not believe that the addition of a **social component** to implicational scales add anything to our knowledge of the patterning of linguistic structure or to an explanation of change.» (Romaine 1988:181-182).

privilegia o contacto com falantes da sua língua e de Português. Apesar de parecer ter uma atitude e motivação positiva em relação à L2, o Inglês, a sua interlíngua não passou de um estágio bastante pobre. Schumann atribui isto à distância social e psicológica em relação à L2, e argumenta que a *aculturação* é determinada pelo grau de distância social e psicológica entre o sujeito não nativo e a cultura e língua alvos.

Schumann (1978a:76) propõe que os primeiros estádios de aquisição de uma L2 são caracterizados pelos mesmos processos de simplificação responsáveis pela formação de *pidgins*. Ou seja, tal como acontece com os *pidgins*, esses primeiros estádios satisfazem a função comunicativa ou *referencial* da linguagem, mas não as funções *integrativa* e *expressiva*. Esta teoria foi pouco depois abandonada por Schumann (1978b) por considerar que o desenvolvimento de uma L2 é motivado pelo desejo do sujeito de se aproximar da variedade alvo, aproximando-se, portanto, mais de um processo de descrioulização. Este modelo foi alvo de muitas críticas ¹³³. As mais fortes estão relacionadas com o facto de um *pidgin* ¹³⁴ se desenvolver quando falantes de várias línguas estão em contacto, ao contrário do que acontece no caso de aquisição de L2 que é uma situação de contacto bilingue ¹³⁵. No entanto, Larsen-Freeman & Long (1991:265) salientam que estes estudos tiveram o mérito de chamar a atenção dos investigadores para alguns processos de simplificação presentes nos primeiros estádios de desenvolvimento da interlíngua e para a importância dos **factores sociais, individuais e psicológicos**.

No final dos anos 70, surge uma teoria, particularmente conhecida e geradora de grandes debates até aos dias de hoje, que o seu autor vinha desenvolvendo havia já algum tempo. Ela pretende responder à questão do **papel do ensino**, e vem certamente na sequência da investigação realizada

¹³³ Para críticas ao modelo, ver Romaine 1988 e Larsen-Freeman & Long 1991.

¹³⁴ «A **pidgin** is a auxiliary language that arises when speakers of several mutually unintelligible languages are in close contact; by definition, it has no native speakers» (Bickerton 1984:173).

¹³⁵ *The Nativization Model* de R W Andersen (1979 e 1983, entre outros) está relacionado com os trabalhos de Schumann. Mais tarde, Andersen constrói um modelo 'cognitivo-interaccionista' de aquisição de L2 em meio natural e semi-formal a que chama "*Authentic Language Plus Model*" (Andersen 1990). Este modelo baseia-se em investigação pessoal do autor sobre a aquisição do aspecto verbal e também nos trabalhos de Bybee e Slobin, e, por isso, dá muita importância à componente semântica do léxico. Ver Leiria 1991.

em 1974 com Bailey e Madden e de que falámos acima. Trata-se de **The Input Hypothesis** de Stephen Krashen (ver Krashen 1981, 1988 ¹³⁶ e 1994). Ela faz parte e dá o nome a um conjunto de **cinco hipóteses** inter-relacionadas que constituem uma teoria de aquisição de L2 ¹³⁷.

Perante a constatação de que em certas tarefas, nomeadamente na produção escrita, a ‘ordem natural’ parece alterada, Krashen 1976 propõe **The Acquisition-Learning Hypothesis** (ver também Larsen-Freeman 1975). Os adultos dispõem de duas formas distintas e independentes de desenvolver competência numa L2: por **aquisição**, ou seja, por um processo subconsciente, semelhante, se não igual, àquele de que as crianças dispõem no desenvolvimento de L1, e por **aprendizagem** que, em ‘termos não técnicos’, é definida pelo autor como ‘saber sobre’ uma língua, conhecido como ‘gramática’ ou ‘regras’ (Krashen 1981:10) ¹³⁸. **The Natural Order Hypothesis** defende que as regras de uma língua são adquiridas por uma ordem previsível, umas tendencialmente mais cedo e outras mais tarde. **The Monitor Hypothesis** estabelece o modo como aquisição e aprendizagem são usadas na produção. A nossa capacidade para produzir discurso resulta da competência adquirida, enquanto que aquilo que é aprendido só está disponível para monitorizar o discurso. Assim, a primeira é a única fonte de conhecimento que o sujeito pode usar na comunicação em tempo real. Quanto à segunda forma de conhecimento, pode usá-la para eventuais correcções, se tiver tempo suficiente e se estiver consciente da regra. **The Input Hypothesis** ¹³⁹ pretende explicar como é que esse conhecimento é adquirido. *Input* compreensível é o ingrediente básico na aquisição da linguagem. Ou melhor, para que a aquisição aconteça, é necessário que o *input* disponibilizado esteja um pouco acima do estágio de

¹³⁶ Krashen 1988 é a reimpressão de uma série de artigos, alguns revistos, publicados em revistas da especialidade entre 1976 e 1980.

¹³⁷ Preston 1989 – *Sociolinguistics and Second Language Acquisition* – considera o **monitor model** como uma teoria variacionista por distinguir aquisição de aprendizagem (*a dual competence model*). Inclui no mesmo capítulo outras duas de que não trataremos aqui: *A continuous competence model* (Tarone 1983), e *A variable competence model* (Ellis 1984).

¹³⁸ Krashen afirmará mais tarde, parecendo querer enquadrar-se numa perspectiva cognitivista, que o conhecimento adquirido é acidental e implícito, enquanto que o conhecimento adquirido por aprendizagem é intencional e explícito (Krashen 1994:46).

¹³⁹ Para teorias que o autor considera serem ‘**rivaais**’ da **input hypothesis**, ver Krashen 1994.

capacidade de compreensão do aprendente. A fórmula é « $i + 1$ ». É deste modo que o *input* disponibilizado na sala de aula, em particular para os principiantes, que, geralmente, têm dificuldade em obtê-lo em meio natural, pode ajudar. *The Reading Hypothesis* é um caso especial da *input hypothesis*. A leitura estimula a aquisição e é responsável por uma boa parte da nossa competência para a compreensão de material escrito. E, por último, ***The Affective Filter Hypothesis*** prevê que, se o sujeito está ansioso ou não se identifica com a comunidade de falantes da língua, este filtro actuará de modo que o *input* não atingirá o mecanismo de aquisição da linguagem, ‘*language acquisition device*’ (Krashen 1994:46).

As primeiras críticas a este modelo não se fizeram esperar e dirigiram-se sobretudo à *Acquisition-Learning Hypothesis*, ou seja, ao facto de os dois tipos de conhecimento serem encarados como completamente independentes e de a aprendizagem ser considerada como não importante para o desenvolvimento da interlíngua. McLaughlin 1978 chama a atenção para que, em termos psicológicos, a distinção consciente vs. inconsciente, em que se apoia a distinção aprendizagem vs. aquisição, não é sustentável. E propõe que seria mais adequado falar de um *continuum* entre **processos controlados** e **automáticos** ¹⁴⁰. Bialystok 1978 apresenta um modelo alternativo: um *continuum* entre **conhecimento implícito** e **conhecimento explícito** inter-relacionados por processos de inferência. Conhecimento implícito é constituído por informação que é automática e espontaneamente usada em tarefas linguísticas. Conhecimento explícito é constituído por factos que o falante sabe sobre a língua e a capacidade para os articular. Estes dois investigadores reelaboraram posteriormente estas suas propostas (Bialystok 1981 e 1982; McLaughlin *et alii* 1983, McLaughlin 1990) e, como veremos, elas continuam a ser influentes em investigação desenvolvida na abordagem cognitiva ¹⁴¹.

¹⁴⁰ «The advantage of this distinction is that it enables one to avoid disputes about ‘conscious’ or ‘subconscious’ experience, since the **controlled-automatic distinction** is based on behavioral acts, not on inner states of consciousness.» (McLaughlin 1978:318).

¹⁴¹ Por exemplo, Rod Ellis (1994) – *A Theory of Instructed Second Language Acquisition* – cruza os dois tipos de conhecimento (explícito e implícito) de Bialystok com os dois tipos de processamento (controlado e automático) de McLaughlin. Esta teoria enfatiza a importância dos conhecimentos prévios (L1, outra ou outras conhecidas pelo sujeito, conhecimento do mundo e hábitos de aprendizagem) na aprendizagem de uma L2. Referências mais detalhadas a Ellis 1994 podem encontrar-se em Leiria 1998b.

O modelo de Krashen tem continuado no centro das atenções e das críticas (ver White 1987; e Cook 1993). Referirei só algumas delas. Embora nunca o afirme, mas como se pode ver pelo que foi dito acerca de *The Affective Filter Hypothesis*, Krashen parece aceitar a existência de qualquer dispositivo inato para a aquisição. No entanto, em nenhum momento especifica quais as suas propriedades, nem estabelece uma relação entre ele e os conhecimentos da L1 e da L2. Ou seja, ela não resolve nada quanto ao modo como esse conhecimento é adquirido, quanto ao papel da L1 e quanto à eventual transferência de conhecimentos desta para a L2. Ao atribuir toda a responsabilidade de possíveis insucessos ou dificuldades na aquisição da L2 a factores sociais ou psicológicos, afasta qualquer hipótese relativa a outros aspectos, cognitivos ou linguísticos, à relação entre as duas línguas em presença, e por conseguinte, à ordem de aquisição. E, apesar de este modelo, provavelmente pela sua simplicidade e aparente clareza de formulação, ter atraído muito a atenção dos professores de línguas, ele não fornece quaisquer pistas para quais as características que deve ter o *input* disponibilizado na sala de aula, e como é que ele deve ser modificado, de modo a facilitar a aquisição ¹⁴². Vivian Cook (1993:67) diz que ele é demasiado vago e injustificado por evidência empírica o que, segundo este autor, faz dele um modelo pouco mais do que metafórico. No entanto, ele reúne as principais questões envolvidas na aquisição e desenvolvimento de uma L2 a que se refere Spolsky e daí o ter-se mantido como um marco referencial em todas as discussões até aos nossos dias.

Como veremos a seguir, muitos destes estudos e teorias dos anos 70 continuam a ser pontos de referência e traçam as linhas de investigação perseguidas até à actualidade. Ficara assente que a hipótese de Skinner 1957 – a um estímulo corresponde uma resposta – não dava conta da realidade. Mas ficavam por responder muitas outras questões: (1) **a ordem que preside à aquisição e ao desenvolvimento da linguagem deve-se a uma faculdade inata de linguagem, como propõe Chomsky, a estratégias universais de processamento da linguagem, como propõe a psicologia do desenvolvimento, ou à combinação de ambas?** (2) **qual o**

¹⁴² Trabalhos experimentais que pretendem avaliar as vantagens de input modificado foram realizados por Rod Ellis 1995 e Ellis & Tanaka & Yamazaki 1995.

papel da idade e, conseqüentemente, qual a relação entre desenvolvimento cognitivo e desenvolvimento linguístico? (3) como e em que medida é que o ensino formal interfere na aquisição? E, agora que esta fase inicial da investigação minimizou o peso do transfer, (4) qual é então o papel da L1? A investigação continua, até hoje, a procurar respostas para estas questões.

De qualquer modo, estavam lançadas as duas linhas que iriam guiar a investigação em aquisição de L2. Uma, na sequência das propostas de Chomsky, que põe a tónica em factores internos – a aquisição e desenvolvimento da linguagem é como é porque todos os seres humanos nascem equipados com um dispositivo especializado para essa função – e que usa como enquadramento as teorias desenvolvidas no âmbito da Gramática Generativa. Uma outra, na sequência de Brown, que privilegia os aspectos relacionados com capacidades de processamento idênticas de que dispõem todos os seres humanos, que se serve de enquadramentos linguísticos muito mais variados (que vão da Sociolinguística aos estudos de *pidgins* e de crioulos e à tipologia linguística), e que procura apoiar-se mais em investigação psicolinguística.

3.3. *A abordagem linguística*

Second language acquisition research has historically been influenced by the field of linguistics, not only in terms of theoretical insights, but also in terms of the emphasis placed on various topics (Gass 1988:92).

Na década de 80, a *Gramática de Princípios e Parâmetros* constituiu um marco decisivo também na investigação em L2 (Chomsky 1981, 1986 a, 1986b e 1989). O conceito central desta teoria continua a ser a Gramática Universal (GU). Independentemente de eventuais diferenças de posição dentro deste quadro, todos aceitam que **(1)** a capacidade da linguagem é inata e está representada na mente de todos os seres humanos, **(2)** ela constitui um módulo separado de outras faculdades da mente, **(3)** o conhecimento sobre a linguagem consiste numa gramática que permite ao falante distinguir frases gramaticais de não gramaticais, **(4)** apesar de a evidência empírica disponibilizada pelo *input* ser pobre e variável, o facto de todos os falantes de uma mesma língua convergirem para uma gramática semelhante sugere a existência de uma GU que impõe fortes restrições às

gramáticas possíveis. Mas agora ela deixa de ter um conteúdo misterioso, como acontecia com LAD (*'language acquisition device'*), e passa a ter um conteúdo específico:

What we 'know innately' are the principles of the various subsystems of S_0 and the manner of their interaction, and the parameters associated with these principles. What we learn are the values of the parameters and the elements of the periphery (along with the lexicon to which similar considerations apply) (Chomsky 1986:150).

O conhecimento inato de que o falante dispõe é constituído por um conjunto de princípios que aplica a todas as línguas e de um número limitado de parâmetros que variam de língua para língua. Adquirir uma língua significa aprender a aplicar esses princípios a uma língua particular e aprender qual o valor adequado para cada parâmetro. O modelo de aquisição é conceptualizado em termos de estádios da mente: um estádio inicial, geneticamente determinado, (S_0 - estádio 0) presente na mente de todas as crianças como um sistema de princípios e de parâmetros que, respondendo à evidência linguística disponibilizada à criança, cria uma gramática nuclear que fixa valores para todos os parâmetros, produzindo uma das línguas possíveis. Aquisição é um modelo idealizado, no qual o tempo e a experiência desempenham um papel mínimo. O factor crucial é a relação entre este estádio (S_0) e o estádio estabilizado (S_s - *'steady state'*). O desenvolvimento é a história dos estádios intervenientes e reflecte a complexa interacção da linguagem com outras faculdades da mente que estão a amadurecer ao mesmo tempo. O estádio estabilizado, em que os parâmetros já estão fixados, consiste numa *gramática nuclear*, numa *gramática periférica* e num *léxico* composto por itens idiossincráticos. A distinção entre nuclear e periférico não é absoluta. Trata-se de um *continuum*. Assim, **marcação** refere-se ao afastamento do que é mais neutro, mais usual. Quanto mais um item se afasta do que é mais usual mais marcado é. Um outro sentido de marcação refere-se à escolha entre possíveis fixações de um dado parâmetro. Uma pode estar mais próxima do núcleo e outra da periferia. Para decidir qual delas é a mais nuclear, usa-se informação da aquisição dos primeiros estádios de L1 (Cook 1988:54).

Além das **gramáticas, nuclear e periférica**, a criança tem de adquirir um grande número de **itens lexicais**, as suas propriedades

fonético-fonológicas, o seu significado, e as suas restrições sintácticas. E, como vimos no Capítulo 1, o princípio de projecção prevê que as propriedades das entradas lexicais se projectam na sintaxe da frase. Assim, o léxico deixa de ser uma matéria à parte, uma lista de palavras e significados e passa a desempenhar um papel importante na sintaxe. **Aprender uma língua consiste, em grande parte, em aprender, a partir dos dados disponíveis, os elementos do léxico e as suas propriedades** (Chomsky 1982a:8; Cook 1988:57). Aprender uma L2 afecta os aspectos paramétricos mas não afecta os princípios, uma vez que estes não variam de língua para língua e, portanto, não precisam de ser adquiridos. Assim, a aquisição consiste na fixação dos parâmetros e do léxico; e «perhaps parameters only belong to the lexicon». A multi-competência é possível porque ela representa um léxico alargado, com muitos mais itens lexicais especificados em termos de parâmetros (Cook 1994: 499) ¹⁴³.

Apesar de a teoria começar a prestar atenção às propriedades do léxico, a maior parte da **investigação em L2**, realizada durante os anos 80, no âmbito deste modelo, continuou a centrar-se bastante na sintaxe ¹⁴⁴. Essa investigação tinha como principal objectivo testar se os parâmetros identificados pela teoria são ou não violados nas gramáticas intermédias e, assim, discutir se a GU continua ou não disponível para a aquisição de L2 ¹⁴⁵. As diferentes posições têm ido desde a hipótese de que a GU não está disponível ('no-access position') para a aquisição de L2 (Clahsen & Muysken 1986), até à hipótese de que ela opera exactamente como em L1 ('direct-access position'). Neste caso, a aquisição de L2 é encarada como sendo o resultado da interacção pura entre a GU e os dados da L2 (Dulay & Burt 1974; Krashen 1981). Mas a maior parte dos investigadores defende uma

¹⁴³ «In one sense this fits with modern ideas of the bilingual lexicon; [...] the evidence is mostly in favour of one interdependent store rather than separate stores. The lexical parameter view of acquisition fits the view of multi-competence [...] although its full implications await the further development of the syntactic theory.» (Cook 1994:499).

¹⁴⁴ «Within linguistics, the lexicon has taken on a secondary role in terms of theory construction; within second language acquisition, the rift between studies of the **lexicon** and theory construction seems to be even greater.» (Gass 1988:92).

¹⁴⁵ Ver Flynn 1987, Flynn & O'Neil (eds.)1988.

posição intermédia ('indirect-access position'), ou seja, que o acesso à GU é mediado pela L1 ¹⁴⁶.

Quer a versão forte quer a versão fraca da acessibilidade à GU, ao prever a intervenção da L1, conduz a investigação a uma nova abordagem dos estudos contrastivos e a uma releitura do conceito de **transfer**. Assim, mais ou menos duas décadas depois de Chomsky a ter posto em causa, a análise contrastiva regressa, mas agora com um forte enquadramento teórico. A semelhança entre as línguas é explicada como reflexo da GU, na medida em que ela constitui uma base comum para todas elas. Adquirir uma L2 significa, em boa parte, refixar o valor de parâmetros diferentes dos da L1. Mas, tal como acontece em relação à L1 ¹⁴⁷, mesmo entre aqueles que defendem que o desenvolvimento da L2 é de algum modo guiado ou restringido pelos princípios da GU, não tem havido uma posição assumida quanto às propriedades linguísticas do **estado inicial** que será modificado ou elaborado pela interacção do *input* com a GU durante o desenvolvimento da L2. Para responder a essa falta, em 1996, no n.º 1 do volume 12 da revista *Second Language Research*, surgem três teorias que abordam a questão da GU e do *transfer*. Elas têm como objectivo caracterizar o estado inicial de L2, ou seja, **de que conhecimentos gramaticais é que se parte quando se começa a aprender uma L2**. As três propostas aceitam que o acesso à GU é mediado pela L1, mas divergem quanto àquilo que é transferível, isto é, quanto àquilo que consideram constituir o estado inicial. Vainikka e Young-Scholten 1996 defendem que um parâmetro pode ser refixado; que o que se transfere da L1 é só a projecção das categorias lexicais e que, por conseguinte, o desenvolvimento consiste na aquisição das categorias funcionais. Eubank 1996 defende que o estado inicial de L2 compreende toda a gramática de L1, excepto a morfologia flexional e os valores dos parâmetros que lhe correspondem. Para Schwartz e Sprouse 1996 o estado inicial da aquisição de L2 é o estado final de L1 (excluindo a

¹⁴⁶ Ver White 1989 e Cook 1993.

¹⁴⁷ «One of the interesting questions for the UG model concerns the **initial setting for parameters**. [...] A further controversy surrounds the issue of whether all the principles and parameters are present in the mind to start with, or whether they come into being over time. In other words, are the principles and parameters like the heart, which is structurally complete at birth, or like the teeth, which grow and are replaced over many years?» (Cook 1993:203). Ver Hyams 1986.

matriz fonética dos itens lexicais) e assumem que, sempre que se verifique incapacidade para encontrar uma representação, se recorrerá à GU. Fazem notar que, deste modo, o estado inicial de L2 é muito diferente de L1, porque todos os princípios e valores de parâmetros são transferidos para o estado inicial de L2 como foram instanciados em L1, logo na primeira exposição ao *input* de L2. O desenvolvimento de L2 é determinado pelo estado inicial, pelo *input*, pela GU e por outras condições relacionadas com ‘aprendibilidade’ (*learnability*)¹⁴⁸. E, exactamente porque o estado inicial da L2 é diferente do da L1, o estado final da L2 não tem necessariamente que ser uma réplica da L1. Dito de outro modo, **o que explica a possibilidade de um falante de L2 nunca chegar a atingir uma competência igual à do falante de L1 é o facto de partirem de um estado inicial diferente.**

Estas propostas, que reexaminam estudos da primeira fase da investigação (entre outros, Dulay & Burt 1974, Bailey & Madden & Krashen 1974, Gass 1983, e Bley-Vroman 1989¹⁴⁹), constituem, segundo os seus autores, uma tentativa de encontrar uma explicação para o *transfer* e não, como tem acontecido, uma mera descrição. Por isso mesmo, elas apresentam-se como muito promissoras quanto ao desenvolvimento da investigação neste quadro teórico (ver Juffs 1996).

3.4. A abordagem cognitiva

Como vimos no capítulo 1, as propostas de Chomsky, e nomeadamente a sua Gramática Universal, têm funcionado como detonador de uma boa parte da investigação desenvolvida sobre a linguagem, não só no âmbito da linguística mas também em todos os domínios da Ciência Cognitiva interessados na representação, processamento e aquisição da linguagem natural. Mas como vimos também, a propósito das abordagens funcional e cognitiva, muitos investigadores têm posto em causa o estudo

¹⁴⁸ «The claim that many properties of language are not available directly from input is known variously as ‘the poverty of the stimulus problem’, ‘the projection problem’, and ‘the logical problem of language acquisition’. The general approach to language acquisition from this perspective is known as ‘the **learnability**’ approach.» (Juffs 1996:12).

¹⁴⁹ «Thus, the picture of the difference between child language development and foreign language learning as advocated here is the following:

<i>Child language development</i>	<i>Adult foreign language learning</i>
A. Universal Grammar	A. Native language knowledge
B. Domain-specific learning procedures	B. General problem-solving systems»

(Bley-Vroman 1989: 51)

da linguagem que não tenha em consideração a sua função e outros aspectos sociais, biológicos e cognitivos. Eles argumentam que as descrições desenvolvidas no âmbito da Gramática Gerativa são necessárias para uma teoria de aquisição mas não são suficientes, e que o formato das representações propostas são, com certeza, muito diferentes da nossa representação mental da linguagem. Assim, há alguns linguistas e muitos psicolinguistas que, desde os anos 70, vêem a linguagem não como um fenómeno peculiar, mas como podendo ser integrada numa estrutura mais abrangente, ou seja, como é que o ser humano adquire e armazena o conhecimento. Neste quadro, a semelhança entre as sequências de aquisição de diferentes crianças, crescendo em diferentes comunidades e aprendendo línguas diferentes, é atribuída a princípios funcionais e a restrições impostas por capacidades de processamento semelhantes (Slobin 1985 e MacWhinney 1986), e não necessariamente à presença, na mente do ser humano, de uma gramática inata.

A psicolinguística do desenvolvimento, continuando os estudos dos anos 70 de Bowerman, de Brown e de Slobin, parte não da 'I-language', como fazem os generativistas, mas da 'E-language'¹⁵⁰, para desenvolverem um *método interlinguístico* que lhes permita estabelecer a relação entre *universais linguísticos* e *universais cognitivos* (Slobin 1973). Para construir uma gramática ('*basic child language*') qualquer criança actua sobre o *input* disponível, usando um conjunto de *princípios operatórios* semelhantes, que constituem '*the language making capacity*' (LMC). Esse conjunto de princípios operatórios, que na sua forma inicial são prévios à experiência da criança face à linguagem, subdividem-se em dois grandes grupos: ***princípios de atenção*** e ***princípios de armazenamento***. Os primeiros actua no sentido de extrair do *input* e segmentar pequenos *chunks*. A extracção «é um processo rudimentar, guiado por princípios quase puramente fonológicos» (Peters 1985:1030) e envolvem *frequência*, e

¹⁵⁰ «I-language approach [...] sees language as 'a system represented in the mind/brain of a particular individual' (Chomsky 1988); I-language research aims to represent this mental state; a grammar describes the speaker's knowledge of the language, not the sentences that have been produced. [...] E-language aims to collect samples of language and then to describe their properties. [...] E-language research constructs a grammar to describe the regularities found in such a sample. [...] The E-language approach includes not only theories that emphasize the physical manifestations of language but also those that treat language as a social phenomenon.» (Cook 1988:12-13)

saliência fonológica e psicolinguística. A segmentação, por sua vez, consiste na descoberta de sub-unidades que compõem cada um dos *chunks*. Estas vão sendo armazenadas, conforme as categorias a que pertencem, em duas grandes áreas: a *semântica*, que não é estática, visto que, com o desenvolvimento, vai sofrendo constantes reorganizações, e a dos *functores*, cuja arrumação obedece a princípios de *saliência*, de *relevância*, de *proximidade sintáctica* e de *objectivo*. À medida que a criança vai processando mais informação linguística, a sua gramática vai-se aproximando mais da gramática dos adultos falantes da língua em aquisição. Neste percurso, as crianças caminham de uma gramática universal para as gramáticas divergentes de cada uma das línguas (Slobin 1985:1160).

Com o **método interlinguístico**, que observa o comportamento linguístico de várias crianças de diferentes línguas, pretende-se comparar as sequências de aquisição de modo a encontrar *universais e particulares* que permitam uma imagem mais nítida da linguagem infantil (Slobin 1985b). O método pressupõe um plano de investigação com duas vertentes, que resultam de duas hipóteses complementares: (1) **a hipótese nula: o desenvolvimento da linguagem é igual em todos os casos; e (2) a hipótese dos efeitos específicos de cada língua: factores relacionados com a acessibilidade das formas linguísticas em cada língua desempenham um papel em cada caso**. Aplicado o método interlinguístico¹⁵¹, encontram-se padrões de desenvolvimento muito semelhantes, por isso, é lícito concluir que, em todas as situações, operam universais de desenvolvimento. O que determina a ordem de emergência das formas linguísticas é o desenvolvimento conceptual, daí que se encontre uma ordem de desenvolvimento comum intra e interlinguisticamente e que, a nível formal, todas as crianças revelem preferências semelhantes no que respeita a selecção e ordenação das formas¹⁵². Contudo, e como prevê a hipótese dos efeitos específicos de cada língua, o curso da aquisição não é

¹⁵¹ Os volumes editados por Slobin, *The Crosslinguistic Study of Language Acquisition* (1985a), compreendem informação de várias línguas indo-europeias (Alemão, Espanhol, Francês, Inglês, Italiano, Polaco, Português, e Romeno) e de Hebraico, Húngaro, Japonês, Kaluli, Samoano, Turco, e de ASL (American Sign Language) (Slobin 1985:20).

¹⁵² Ver Gívon : modo pragmático – modo sintáctico; Chomsky: inviolabilidade de princípios.

exactamente o mesmo nos seus detalhes, já que a interação entre a forma e a função influencia a velocidade e a sequência do desenvolvimento. Diferenças interlinguísticas, relacionadas com clareza, saliência, relevância, frequência, e maior ou menor elaboração de um dado paradigma, devem ser tidas em conta. Ou seja, porque só estudando os particulares se podem descobrir os universais, é preciso analisar os dados no sentido de verificar os efeitos específicos de cada língua.

Segundo o *Competition Model* de MacWhinney e colaboradores, que está associado à teoria de Slobin, o que explica os diferentes tipos de línguas é o facto da capacidade de processamento ser limitada (Bates & MacWhinney 1982; MacWhinney 1986). O modelo baseia-se em três tipos de *princípios: de representação*, que destacam **a importância do léxico** ¹⁵³, *de processamento* ¹⁵⁴, que destacam o modo como os itens entram em competição uns com os outros, e *de aprendizagem*, que garantem que as representações da criança se aproximam continuamente das representações dos adultos. As línguas podem usar quatro tipos de sinais para codificar o sentido: a ordem das palavras, o vocabulário, a morfologia e a entoação. Porque o sistema de **processamento de informação** de que o ser humano dispõe só é capaz de utilizar uma parte da informação, os diferentes tipos de sinais competem entre si para preencher esse espaço limitado de processamento, e, por isso, os diferentes tipos de línguas chegaram a diferentes modos de combinação destes quatro tipos de sinais. Assim, há línguas que se apoiam mais na ordem das palavras, como o Inglês, na morfologia, como o Latim, ou na entoação, como o Chinês (ver Cook 1993:257). A criança aprende não só em que tipos de sinais é que a sua língua se apoia para codificar o sentido mas também a relação de forças entre eles para saber a qual deve prestar mais atenção no acto do processamento. Daí que a frequência de determinadas ocorrências seja determinante para a aquisição. Quanto mais vezes um determinado item é

¹⁵³ «The core representational structure in the Competition Model is the **lexical item** – an association between external and internal function.» (MacWhinney 1986:251).

¹⁵⁴ « **[L]anguage processing** should be governed by many of the same basic principles that govern other aspects of cognitive processing and that the acquisition of language can be explained in terms of general learning principles placed at the service of communicative intentions.» (MacWhinney 1986:250).

ativado mais fácil será o seu acesso e, conseqüentemente, a sua automatização no processamento e na produção de discurso.

Este modelo e todos os outros **modelos de armazenamento e aquisição de conhecimento** têm na base o conceito de ‘*spreading activation*’: a mente é encarada como uma rede altamente complexa através da qual a activação se difunde. Aprender significa fortalecer partes da rede graças à frequência de ocorrências (Anderson 1983b; ver Cook 1993:265). A **memória** é fundamental para essa aprendizagem. O *ACT Model (Adaptive Control of Thought)* de Anderson (1983a) estabelece uma distinção clara entre *conhecimento declarativo* e *conhecimento procedimental*, e distingue três formas de *memória*: *de trabalho*, *procedimental* e *declarativa*. A memória de trabalho é usada na produção de regras e estabelece a ponte entre a procedimental e a declarativa. Ao aprender, qualquer ser humano se move do conhecimento declarativo para o procedimental. Ao contrário do *ACT*, as teorias conexionistas, de que o *PDP (Parallel Distributed Processing)* é a mais conhecida, não assentam num sistema de produção de regras e não aceitam a separação entre memória declarativa e procedimental ¹⁵⁵. A designação ‘**conexionismo**’ resulta do facto de esta abordagem se inspirar na actividade neurofisiológica do cérebro. Os conexionistas sustentam que as diferentes componentes da informação são processadas *em paralelo*, independentemente umas das outras, e *distribuídas* por diferentes níveis ¹⁵⁶. Por isso, o conexionismo é também conhecido pela designação de ‘*parallel distributed processing*’ (Singleton 1999:121). O aspecto central de um *PDP* é uma rede neuronal formada por um conjunto de unidades ligadas umas às outras por uma série de caminhos. Essas unidades comunicam umas com as outras transmitindo o seu conteúdo àquelas com que estão

¹⁵⁵ «We propose an alternative to explicit inaccessible **rules**. We suggest that lawful behavior and judgments may be produced by a mechanism in which there is no explicit representation of the rule. Instead, we suggest that the mechanisms that process language and make judgments of grammaticality are constructed in such a way that their performance is characterizable by rules, but that the rules themselves are not written in explicit form anywhere in the mechanism.» (Rumelhart & McClelland 1986:217).

¹⁵⁶ O conexionismo enquadra-se, assim, numa perspectiva – o **processamento paralelo** – que, de certo modo, se opõe ao **processamento serial** (ver Levelt 1998, cap 2). A diferença essencial entre as duas consiste em a primeira privilegiar a noção de não dependência, umas das outras, das diferentes operações do processamento, as quais são desencadeadas por determinados eventos e estímulos, enquanto que a segunda representa o processamento organizado em série, dependendo cada estado do *output* do anterior.

interligadas. As conexões são sempre uni-direccionais e as unidades podem excitar-se ou inibir-se umas às outras. Este modelo assume ainda que o conhecimento de eventos, conceitos e linguagem está representado de forma muito difusa no sistema cognitivo (Rumelhart & McClelland 1986) ¹⁵⁷.

No início dos anos 80, um grupo de investigadores de **L2** começou a usar o método interlinguístico (ver volume de 1984, editado por Andersen). O seu trabalho baseia-se na premissa de que a compreensão sólida da aquisição e do uso de uma L2 requer evidência de um grande número de diferentes línguas segundas (e primeiras) adquiridas e usadas numa grande variedade de cenários (Andersen 1984a:1) ¹⁵⁸. Mas se, no que respeita a linguagem infantil, o que determina a emergência das formas é o desenvolvimento cognitivo, são esperáveis diferenças, no que respeita as sequências de aquisição, entre crianças e adultos a aprenderem uma L2. Sendo assim, os resultados obtidos na investigação em L1 não são, pelo menos em grande parte, aplicáveis à L2. Esta constatação conduz os investigadores a procurarem outros apoios. **A tipologia linguística**, cujo objectivo é, a partir de dados linguísticos estabilizados, analisar padrões interlinguísticos, e que afirma que **as línguas são semelhantes porque estão sujeitas às mesmas restrições impostas pelas capacidades de processamento**, surge como tendo condições para servir como termo de comparação e assim validar, e ser validada, pelas sequências encontradas para a L2. A pergunta que põem e que desejam ver respondida pela investigação é: (1) **em que medida é que a noção de marcação** ¹⁵⁹ **pode**

¹⁵⁷ Broeder & Pkunkett 1994 dizem que o **PDP** não tem merecido muita atenção em L2, mas descrevem alguns trabalhos feitos, até então. Por exemplo, a semelhança da célebre simulação de aprendizagem do *past tense*, feita por Rumelhart & McLelland 1986, Sokolik & Smith 1992 simulam, em computador, uma rede que é treinada para a aquisição de um léxico de 225 nomes masculinos e outros tantos femininos. Em seguida, a rede é capaz de classificar correctamente, quanto ao género, outros 150 nomes. Para crítica a esta experiência, ver Carroll 1995 ("Sokolik & Smith 1992 experiments greatly oversimplify the complexity of 'gender learning': pp. 203).

¹⁵⁸ O volume editado por Andersen (1984) é um contributo fundamental para a investigação de universais em aquisição de L2, numa perspectiva interlinguística. Contém artigos sobre as seis línguas que na época já tinham sido objecto de investigação enquanto L2 (Alemão, Espanhol, Francês, Inglês, Neerlandês e Sueco). Vários artigos concentram-se nos princípios operatórios de Slobin e Peters.

¹⁵⁹ «[L]inguistic phenomena that are common in the world's languages, that seem easier for linguistic processing, and that are more 'natural' than others, are unmarked as opposed to their marked variants.» (Hyltenstam 1986:57). Ver mais acima o conceito de marcação em Chomsky.

prever os problemas de desenvolvimento e de aprendizagem que um determinado contexto de L1/L2 põe? Haverá paralelismo entre hierarquias de marcação e sequências de aquisição? (Hyltenstam 1984:44). Poder-se-á pôr como hipótese que elementos e estruturas interlinguisticamente não marcadas surgirão em fases iniciais de aquisição, independentemente das condições de marcação da L1 e da L2 (o que, de certo modo, corresponde à hipótese nula de Slobin). Poder-se-á ainda pôr como hipótese que as condições de marcação de L1/L2 desempenharão um papel importante nas fases seguintes de desenvolvimento: os padrões de *transfer* estarão dependentes da marcação em L1 e em L2 e das relações entre as duas línguas (Hyltenstam 1986:75).

A investigação realizada durante os anos 80 aponta para que o conceito de marcação pode oferecer uma base teórica segura para a formulação de hipóteses em investigação de L2. Mas, para que um dado fenómeno linguístico seja estudado a partir de hipóteses formuladas com base na tipologia linguística, é necessário que esta o tenha estudado. Daí que os trabalhos de Jakobson 1941 sobre fonologia segmental e a hierarquia de acessibilidade de Keenan & Comrie 1977 ¹⁶⁰ tivessem sido alvo de particular atenção ¹⁶¹. Outros investigadores, particularmente interessados no fenómeno da aquisição da linguagem, começam a desenvolver estudos tipológicos que lhes permitam apoiar a sua investigação. É o caso de Ake Viberg que, desde o início dos anos 80, vem desenvolvendo investigação centrada no **léxico** ¹⁶². Como vimos no Capítulo 1, a partir de estudos de frequência em onze línguas da Europa, concluiu que um pequeno número de campos semânticos cobre a maior parte das palavras mais frequentes dessas línguas, e que, além disso, muitos desses significados são tipologicamente não marcados e ocupam lugares cimeiros nas hierarquias universais de lexicalização (Viberg 1993a:341-2).

¹⁶⁰ Se um dado sintagma nominal é passível de relativização numa dada língua, então qualquer outro que esteja situado acima na hierarquia também é passível de relativização. A hierarquia é como se segue: sujeito < objecto directo < objecto indirecto < objecto obliquo < genitivo < objecto de comparação. Ver, entre outros, Eckman 1977, e Gass 1979.

¹⁶¹ Para revisão de trabalhos sobre marcação e L2, ver Hyltenstam 1986.

¹⁶² Ver Viberg 1984 'The verbs of perception: a typological study' (*Linguistics* 21:123-162). Este texto foi apresentado pela primeira vez no *Linguistics Universals Workshop*, em Cascais, em 1982.

Viberg está particularmente interessado nos **verbos**. Como vimos também no capítulo 1, Viberg isola um pequeno número de verbos (13) que considera serem os **verbos básicos** das onze línguas analisadas. Esse conjunto de verbos básicos é constituído **por verbos nucleares** (os que reflectem tendências universais e que são os tipologicamente menos marcados dentro do respectivo campo semântico) mais alguns **verbos específicos** dessa língua ou do grupo de línguas de que ela faz parte.

Partindo deste quadro, estudou dados de crianças de seis anos falantes de Sueco como L2. Os **verbos nucleares**, que representam os traços não marcados, tendem a ser mais usados por estas crianças do que pelas do grupo de controle constituído por crianças suecas da mesma idade; enquanto que **outros verbos básicos** que, do ponto de vista semântico, representam padrões linguísticos específicos, tendem a ser evitados ou são usados de tal modo que neutralizam essa especificidade. Se tivermos em consideração que quanto maior é a abrangência semântica de um lexema maior será o número de construções em que ele entrará, compreenderemos que ele defenda que os verbos nucleares tendem a funcionar como protótipos sintácticos na aquisição de L1 e de L2 (Viberg 1993a:351). E que ponha a hipótese de, uma vez adquiridos os verbos básicos nos seus usos mais típicos, o falante ter atingido uma fluência básica e um domínio dos aspectos fundamentais da estrutura das frases. E mais:

The great utility of a small number of verbs in terms of semantic coverage, syntactic prototypicality and morphological accuracy may contribute to the explanation for the experience of 'plateaux' in second language acquisition (Viberg 1993a:368).

A maior parte dos trabalhos de investigação referidos até aqui que seguiram um método interlinguístico foram realizados com evidência empírica de aprendizagem em contexto semi-formal. Viberg parte da tipologia linguística para encontrar um léxico básico e põe a hipótese de sobre ele o falante de L2 construir uma gramática básica. No que respeita a investigação em aquisição, os seus informantes dispõem de um ambiente bastante rico no que respeita a L2 o que justifica a expectativa, por parte do investigador, de os aprendentes atingirem um nível de desenvolvimento linguístico, em fluência e correcção, próximo da variedade alvo. Wolfgang

Klein & Clive Perdue 1997 usam o mesmo enquadramento, interlinguístico e tipológico, mas abordam a questão de uma competência básica de um ponto de vista completamente diferente: partem da análise do discurso de primeiros estádios de aquisição de uma L2 em contexto informal para encontrar aquilo a que chamam **The Basic Variety** ¹⁶³. Os autores dizem não estar interessados numa abordagem *erro / não-erro*, que acham legítima quando o ensino está envolvido, mas numa abordagem que faça luz sobre quais as propriedades necessárias e quais as acidentais na linguagem humana e como é que ela se desenvolve (Klein & Perdue 1997:303). E acentuam: «It should be clear, however, that we are talking here about the acquisition of I-language. I-language is a small fraction of the knowledge, which is required to be a fluent speaker of a ‘language’», porque «almost everything someone has to learn, when he or she wants to become a speaker of the target language, is irrelevant for this subpart of language acquisition» (Klein & Perdue 1997:342).

Os autores dizem que, de qualquer modo, a ideia de uma variedade básica não é nova. Para caracterizarem *The Basic Variety* do aprendente de L2 em contexto informal, usam uma metodologia interlinguística semelhante àquela que Slobin 1985 usara para chegar à sua *basic child language*. E referem, além deste, trabalhos de outros investigadores já nossos conhecidos: Schumann 1978, o contraste entre modo pragmático e modo sintáctico de Givón 1979, assim como Bickerton 1984 e a sua *natural semantax* ¹⁶⁴.

Klein e Perdue (1997) constataam que adultos a aprenderem uma L2 em contexto informal, a partir do momento em que sabem algumas palavras, começam imediatamente a ser capazes de comunicar e de as usar apropriadamente no discurso. Eles produzem enunciados simples, com poucos constituintes à volta do verbo, mas são capazes de exprimir relações espaciais e temporais (‘antes’, ‘depois’, simultaneidade, duração,

¹⁶³ **Evidência empírica para a BV** é constituída pelos materiais recolhidos no âmbito do ‘European Science Foundation Project on Adult Acquisition’ (ver Perdue (ed) 1993). O projecto envolve 5 línguas alvo (Inglês, Alemão, Neerlandês, Francês e Sueco) e 6 línguas fonte (Punjabi, Italiano, Turco, Árabe, Espanhol e Finlandês). O n.º 4 do volume 13 (1997) de *Second Language Research* organiza-se à volta do texto de Klein & Perdue 1997. Além deste, contém outros textos que o comentam.

¹⁶⁴ Ver Bickerton 1984 e “the language bioprogram hypothesis” (LBH).

habitualidade, iteração, etc). São capazes de distinguir estados e eventos, e localização e mudança de localização. É a este sistema muito simples, no que respeita os seus princípios organizativos, que chamam *The Basic Variety* (BV). Estes sistemas básicos têm a característica de ser muito semelhantes, independentemente do aprendente e da sua L1.

Os itens lexicais ocorrem geralmente numa forma invariante. Ou seja, na BV não há flexão, marcas de caso, número, género, tempo, aspecto, ou quaisquer outras concordâncias morfológicas. A forma fonológica é muitas vezes fortemente influenciada pela L1. A dimensão do **léxico** e a sua origem variam bastante em função do falante. O que não varia é a sua composição. Ele consiste, essencialmente, num conjunto de palavras tipo substantivo ('noun-like') e tipo verbo ('verb-like'), mais alguns adjectivos e alguns advérbios, uns poucos quantificadores, alguns determinantes (sobretudo demonstrativos), uma palavra para exprimir a negação, e algumas preposições com significado generalizável. O sistema pronominal dispõe só de meios mínimos para se referir ao falante, ao ouvinte e a uma terceira pessoa. Não existe referência pronominal anafórica a entidades inanimadas. Em resumo, o léxico é constituído por uma lista de itens pertencendo a classes abertas e uma pequena lista de itens de classes fechadas com significado lexical. De qualquer modo, ele é suficiente para estabelecer um sistema de comunicação altamente eficiente e, neste sentido, a variedade básica é uma verdadeira língua ('a real language') (Klein & Perdue 1997:333).

Klein & Perdue não fazem esse exercício, mas, se as teorias cognitivas e tipológicas estão certas, é provável que o léxico desta variedade básica tenha uma composição muito semelhante ao da proposta de Viberg encontrado pelo método interlinguístico. O que falta nesta variedade básica são todos os aspectos 'acidentais', tudo aquilo que é preciso para atingir níveis de fluência, correcção e complexidade próximos dos de um falante de L1.

Nos últimos anos, os investigadores de L2 que aceitam como enquadramento teórico os modelos de armazenamento e **aquisição de conhecimento** de que falámos antes têm insistido bastante em que aprender uma língua, ou aprender outra coisa qualquer, demora muito tempo. No caso de uma língua, esse tempo é gasto a extrair e a armazenar

informação a partir do *input* disponível o que exige atenção. Mas a questão que a psicologia se vem pondo há já bastante tempo é a de determinar que aspectos do conhecimento são obtidos conscientemente e que aspectos são obtidos inconscientemente ¹⁶⁵. E é assim que continua actual a questão levantada por Krashen, nos anos 70, quanto à distinção entre **aquisição** e **aprendizagem**, quanto à transferência ou não de **conhecimento explícito** para **conhecimento implícito**, e quanto ao papel do ensino na aprendizagem. Dito de outro modo, **o conhecimento metalinguístico, disponibilizado na sala de aula, pode ou não tornar-se disponível para uso 'on-line' ?**

Michel Paradis 1994, em texto intitulado *Neurolinguistic aspects of implicit and explicit memory: implications for bilingualism and SLA*, defende que a distinção de Krashen entre competência linguística implícita (aquisição) e conhecimento metalinguístico (aprendizagem) pode ser sustentada pela neuro-anatomia. Pacientes amnésicos apresentam um déficite considerável na memória explícita mas mantêm intacta a memória implícita. A memória implícita e a memória explícita dependem de diferentes sistemas cerebrais e estão, por conseguinte, diferentemente envolvidas no processo de aquisição/aprendizagem de uma L2. A memória declarativa, que serve a aprendizagem formal de uma L2, é, do ponto de vista anatómico e neuro-funcional, diferente da memória procedimental que serve a aquisição de uma língua, L1 ou L2, em contexto conversacional. Defende que o conhecimento metalinguístico aprendido formalmente na escola não é integrado na competência linguística e não se torna disponível para uso automático (Paradis 1994:393). Conhecimento implícito e conhecimento explícito são servidos por diferentes sistemas da memória, têm diferentes conteúdos, e por isso um não se pode transformar no outro, ser convertido ou transferido. Ou seja, **o conhecimento metalinguístico não se transforma em conhecimento procedimental graças a**

¹⁶⁵ «In psychology the **role of consciousness in learning** is discussed under such diverse rubrics as 'implicit' or 'explicit' learning; 'incidental' or 'intentional' learning; 'conscious', 'unconscious', or subconscious processes; 'implicit' or 'explicit' memory; 'metacognition'; 'automatic' and 'controlled' processing; 'subliminal learning'; 'attended vs. unattended learning'; memory of 'how', also called 'procedural', 'skill', and 'habit' memory, *vs.* memory of 'that', also called 'declarative' memory; 'practical' vs. 'academic' intelligence, etc. (Squire 1987).» (N Ellis 1994:213).

automatização. O que é automatizado é a capacidade para produzir sequências correctas de palavras, independentemente dos processos usados para atingir esse resultado (Paradis 1994:402). E acrescenta:

*This those not mean that metalinguistic knowledge cannot be useful in the process of learning another language, whether by focusing attention on some aspect of the linguistic data that would otherwise have gone unnoticed, or by allowing one to check one's output, or to deduce who does what to whom through a conscious identification of case markers, and thereby improving one's practice. But **it is the practice, not the metalinguistic knowledge, which improves automatic performance (and by implication, linguistic competence)** (Paradis 1994:405).*

Assim, um falante de L2 que teve acesso a ensino formal dispõe de duas fontes de conhecimento. Um enunciado é geralmente produzido recorrendo a competência implícita; a competência explícita só pode ser usada para construir frases lentamente ou para verificar a gramaticalidade do que foi produzido automaticamente (Paradis 1994:400).

Os falantes de L2 distribuem-se ao longo de um *continuum*: num extremo, aqueles que, não tendo tido oportunidade de usar a língua em contextos conversacionais, constróem laboriosamente os enunciados fazendo pausas muito longas e que são incapazes de compreender discurso produzido a velocidade normal; no outro extremo, as crianças de três ou quatro anos que não dispõem de nenhum conhecimento gramatical explícito. No caso da produção controlada, é muito difícil prestar atenção, ao mesmo tempo, aos diferentes sistemas envolvidos. Se a atenção está selectivamente concentrada em um deles (fonologia, morfossintaxe, ou léxico), os outros têm de esperar. Os processos automáticos, pelo contrário, porque não dependem de atenção, não interferem uns com os outros e por isso podem operar em paralelo (Paradis 1994:400).

The advantage of incidental acquisition is that it leads to implicit internalisation and automatic use. Its great disadvantage is that it requires a considerable amount of time (Paradis 1994:413).

Todos os aspectos da aquisição e uso da morfossintaxe e da fonologia são compatíveis com as características da memória procedimental, enquanto que pelo menos alguns aspectos do léxico caem no domínio da memória declarativa, afirma. Há evidência crescente de que a plasticidade do cérebro para a aquisição implícita da fonologia e da

morfossintaxe diminui com a idade. Assim, a morfossintaxe, ao contrário do vocabulário, parece ser afectada por aspectos que a idade impõe à memória procedimental (*maturational constraints imposed on procedural memory*) (Paradis 1994:398). No que diz respeito ao uso do léxico, e porque o seu conhecimento é, em certa medida, explícito, a selecção de um item lexical pode por vezes ser controlado, até mesmo sem abrandamento do ritmo do discurso (Paradis 1994:399).

Para concluir, Paradis considera que tanto a L1 como a L2 se apoiam, em medidas diferentes, na memória implícita e na memória explícita; **quanto maior for a ênfase no conhecimento explícito, maior diferença haverá no processo de aquisição entre L2 e L1**; a idade interferirá na aquisição da fonologia e da morfossintaxe; se a L2 é processada de modo diferente ou não da L1, isso depende da competência linguística de que o falante dispõe; quanto mais fraca é a competência, mais o falante se apoiará em conhecimento metalinguístico e em deixas pragmáticas; o falante de L1, em situações formais, também pode recorrer a estes tipos de informação; assim, o processamento de L2 não tem necessariamente de ser diferente do ponto de vista qualitativo; o que varia é a quantidade de suporte fornecido pelo conhecimento metalinguístico e ou por aspectos pragmáticos (Paradis 1994:413-4). Embora a distinção entre aquisição e aprendizagem continue a merecer a atenção dos investigadores, esta é, com certeza, a razão pela qual, actualmente, muitos a evitam e preferem usar os dois termos indiferentemente. Eles assumem que, sendo qualquer processo de aprendizagem em certa medida controlado, a distinção entre processos conscientes e inconscientes, ou entre aprendizagem e aquisição, não é qualitativa mas quantitativa ¹⁶⁶.

Escolhi o texto de Paradis porque ele faz a ponte entre evidência neurolinguística e aquisição. Mas, na verdade, ele faz parte de um relativamente grande grupo de investigadores que defende que competência linguística resulta da abstracção de regularidades a partir de exemplos fornecidos pelo input ¹⁶⁷. Esta ideia de que o que é automatizado é a capacidade para produzir **sequências correctas de palavras**, e que será a

¹⁶⁶ Ver, por exemplo, Laufer 1997:155.

¹⁶⁷ Vários textos que representam esta posição relativamente recente e bem fundamentada fazem parte do volume editado por Ellis (1994) *Implicit and Explicit Learning of Language*.

partir destas que o aprendiz vai construindo a competência linguística, é desenvolvida por Nick Ellis 1996:

[M]uch of language acquisition is in fact sequence learning (for vocabulary, the phonological units of language and their phonotactic sequences; for discourse, the lexical units of language and their sequences in clauses and collocations) (Ellis 1996:91).

Aprender uma língua consiste em analisar e aprender sequências. Aprender sequências de sons em palavras e sequências de palavras em frases. Estas sequências formam uma base de dados a partir da qual o sujeito pode abstrair uma gramática ¹⁶⁸. **Chunking** surge, assim, como um processo geral que preside a aquisição de uma L2. Aprender a **estrutura de uma palavra** implica identificar unidades, a sua sequência particular em palavras, e as probabilidades de sequencialização geral numa língua particular. Porque o **princípio idiomático** preside a uma boa parte do discurso fluente, aprender a **estrutura discursiva** implica, em grande parte, aprender sequências particulares de palavras em combinatórias e outras sequências habituais. Aprender a **classe gramatical** de uma palavra e aprender as **estruturas gramaticais**, em geral, implica a análise implícita e automática da posição de palavras em relação a outras.

Os sujeitos diferem na capacidade de repetir sequências fonológicas, ou seja, diferem quanto à **memória fonológica a curto prazo** (MFCP) ¹⁶⁹. Diferenças individuais na MFCP são responsáveis pela maior ou menor facilidade de aprender o **vocabulário** de uma L2. E, conseqüentemente, a **gramática** da L2. A **repetição** de sequências permite a sua consolidação na memória fonológica a longo prazo (MFLP). Sendo assim, a experiência quanto a sequências regulares, porque mais frequentes, é maior do que quanto a sequências irregulares. À medida que uma rede de associações

¹⁶⁸ «Grammars are systems of rules that operate on a lexical database.» (Ellis 1994:119).

¹⁶⁹ Sevice 1992 e Service & Kohonen 1995 são dois interessantes estudos que relacionam a capacidade de repetir correctamente pseudo-palavras com a capacidade geral para aprender uma L2, ou seja, investigam o papel da **memória fonológica** na aquisição de uma L2. «[T]he data we analysed are consistent with the hypothesis that a special relationship exists between phonological memory and foreign language word learning. More specifically, the ability to create the kind of phonological or phonetic representations in the phonological store that are required for accurate repetition of unfamiliar-sounding pseudowords predicts the learning of foreign words during the early stages of foreign language acquisition. The issue is important, as vocabulary learning may be one of the stumbling blocks for individuals who fail to learn foreign languages.» (Service & Kohonen 1995:171).

fonológicas vai sendo estabelecida, e a **capacidade de generalização** vai aumentando, o sistema vai-se aproximando do dos falantes nativos.

Levando esta ideia um pouco mais longe, Skehan 1998 propõe que o falante, de L1 ou de L2, dispõe de um “dual-mode system”: um sistema baseado em regras (*rule-based system*) e um sistema baseado em elementos soltos (*exemplar-based system*). O primeiro é económica e elegantemente organizado, e apoia-se num léxico também ele bem organizado; este sistema de regras permite gerar novas aplicações. O segundo, pelo contrário, apoia-se principalmente na memória, e é redundante porque contém múltiplas representações de um mesmo item lexical, item esse que faz parte de uma sequência maior; além disso, este sistema, se por um lado é muito limitado quanto ao seu potencial generativo, por outro permite uma grande rapidez no processamento (Skehan 1998:88-89)¹⁷⁰. O falante pode assim mobilizar um ou outro dos sistemas em conformidade com os objectivos e o contexto comunicativo. Mas o autor chama a atenção para um facto que parece estar subjacente a todas estas mais recentes abordagens: reconhecer a existência deste duplo código implica uma mudança profunda no nosso modo de encarar **a relação entre competência e performance** (Skehan 1998:4).

A teoria de Paradis e de outros investigadores segundo a qual, de facto, não é o conhecimento metalinguístico mas a **prática** que desenvolve a competência linguística tem implicações a vários níveis. Por um lado, e como vimos, dá razão a Krashen: conhecimento metalinguístico não se transforma em conhecimento implícito; mas, por outro, e ao contrário deste, atribui ao **ensino** um papel relevante: o trabalho na sala de aula pode proporcionar input adequado e oportunidades de praticar a língua, e serve para chamar a atenção para certas regularidades e para outros aspectos que, em contexto informal de aprendizagem, poderiam não ser notados pelo aprendente. Mas, apesar de em muitos períodos da história do ensino o

¹⁷⁰ «The **rule-based system** is likely to be parsimoniously and elegantly organized, with rules being compactly structured.[...] They will draw in turn upon lexical elements (themselves well organized in a lexicon) as necessary. Such a rule-based system is likely to be generative, with rules being creative in their application. [...] An **exemplar-based system** operates in more or less the opposite manner. It is heavily based on the operation of a redundant memory system in which there are multiple representations of the same lexical elements, because in each case the element functions as part of a unit longer than the word. In consequence, the system lacks parsimony, and has only a limited generative potential. In addition, given that relatively fixed phrases are involved, the potential for expressing new and precise meanings is more limited.» (Skehan 1998:88-89).

léxico ter sido negligenciado ¹⁷¹, de facto, a maior parte da investigação com ele relacionada (melhor dizendo, relacionada com o vocabulário) tem-se ocupado muito mais com o modo como ele deve ser ensinado do que propriamente com os processos envolvidos na sua aprendizagem. Por isso mesmo, falaremos dessa investigação no capítulo 4, *Léxico e ensino de L2*.

3.4.1. *Aprender / saber uma palavra*

Seguindo esta linha de pensamento, nos últimos anos, alguns investigadores têm procurado abordar a aquisição do léxico de um modo a que Meara chamaria «microbiologista», olhando cada palavra como a representação de um conjunto de diferentes tipos de informação, cada um deles regulado por diferentes processos psicolinguísticos. Deste modo, passou-se de uma dicotomia *vocabulário passivo vs. activo* ou *receptivo vs. produtivo* para a consideração de que o conhecimento de uma palavra pode ser parcial e que o falante, de L1 ou de L2, pode ter conhecimento só de uma parte da informação que uma palavra representa.

Na Introdução a este trabalho, salientei que **aprender uma palavra** é muito mais do que aprender o seu significado e que há diversos léxicos especializados para diferentes canais de *input* e de *output*. Batia Laufer (1997:141) resume assim aquilo que considera ser necessário para saber uma palavra:

1. a sua forma oral e escrita, isto é, é preciso saber como se pronuncia e como se escreve;
2. a sua estrutura de base, as derivações mais comuns e a sua flexão;
3. as suas propriedades sintácticas e o seu comportamento numa frase ou num enunciado;
4. as suas propriedades semânticas, o que envolve o seu significado referencial, extensões metafóricas, e valores afectivos, e a sua adequação pragmática;

¹⁷¹ «One of the reasons for this neglect is the fact that vocabulary is not a close, rule governed system, but an open set and as such has probably been less attractive as a research topic.» (Laufer 1990:294).

5. as suas relações paradigmáticas com outras, nomeadamente com eventuais sinónimos, antónimos, e hipónimos; e
6. as suas relações sintagmáticas, ou seja, as suas mais frequentes combinatórias ¹⁷².

É claro que as palavras conhecidas afectam aquelas que se aprendem. Ou seja, **factores interlexicais**, tais como distância entre a L1 e a L2, e **intralexicais**, tais como a dimensão e o grau de organização do léxico da L2, e a frequência com que uma determinada palavra ocorre no *input*, determinam certamente a aprendizagem de novas palavras. Laufer (1990b e 1997) revê uma série de trabalhos experimentais disponíveis e ocupa-se de um conjunto de **factores intralexicais** que afectam a 'aprendibilidade' (*learnability*) de uma palavra: a forma falada e escrita, a extensão, a estrutura morfológica, a classe gramatical, e as propriedades semânticas ¹⁷³.

No que respeita a **forma**, na modalidade oral, quer para a compreensão quer para a produção, e embora não sejam de menosprezar características específicas da palavra e da L2, o que torna certas palavras mais difíceis do que outras é em boa parte determinado pelo sistema fonológico da L1 e pelo apoio que este fornece ou não para a discriminação e produção de certos fonemas. Na modalidade escrita, o grau de correspondência fonema-grafema é um factor a ter em consideração; quer para a leitura quer para a escrita, irregularidades e incongruências a este nível são certamente fonte de dificuldades. A extensão de uma palavra também costuma ser apontado como podendo afectar a sua aprendibilidade. No entanto, diz Laufer, parece haver evidência de que este factor só actua quando combinado com outro ou outros; por exemplo, se uma palavra longa é composta por vários morfemas familiares, ela não

¹⁷² Para o que se entende por saber uma palavra, ver também, entre outros, Clark 1993, Aitchinson & Lewis 1995 e Ellis 1997.

¹⁷³ «Section 2 examined **several features inherent in the word** itself which might affect the ease or difficulty with which the word is learnt. These were the following: phonological: pronounceability and length; grammatical: part of speech, inflexional and derivational complexity; semantic: abstractness, specificity, idiomaticity; register restrictions: multiplicity of meaning. Most of the evidence presented in the section suggests that, except in the case of length, part of speech and abstractness, the above mentioned factors seem to affect the ease of learning the word.» (Laufer 1990b: 303-304).

parece apresentar dificuldades para a compreensão ou para a memorização. A semelhança formal entre palavras também é apontada como um factor que interfere na aquisição; Laufer chama-lhes ‘synforms’. Desenvolveu trabalho experimental sobre elas, e concluiu que, no que respeita aprender a distinguir estas formas semelhantes, embora a ordem possa não ser a mesma, há uma sequência igual tanto para falantes de Inglês L1 como de L2 (Laufer 1990a:289). Primeiro aprendem a distinguir formas que se distinguem por sons consonânticos (*price/prize*) e por diferentes prefixos (*compress/supress*), em seguida as que se distinguem por sons vocálicos (*staff/stuff*), e finalmente as que se distinguem por diferentes sufixos (*economic/economical*)¹⁷⁴.

No que respeita a **estrutura** da palavra, irregularidades no plural, o género dos nomes inanimados, ou outras complexidades de natureza flexional são características que perturbam a sua aquisição. No caso da derivação, o grau de transparência dos morfemas que a compõem também pode interferir no seu reconhecimento e produção¹⁷⁵. Schmitt & Meara 1997, um estudo que já referi em 0.2. *O estado dos conhecimentos em aquisição do léxico em L2*, tendo como informantes um grupo de noventa e cinco japoneses aprendentes de Inglês, investigam a recepção e a produção de palavras com afixos. Apesar de terem verificado existir uma grande variação individual e um fraco desempenho como grupo, puderam concluir que há uma ordem de aquisição muito nítida: primeiro todos os sufixos flexionais e depois todos os derivacionais. Atribuem este resultado ao facto de os sufixos flexionais poderem ser generalizados, uma vez que o seu conhecimento é baseado em regras (*rule-based*), enquanto que a relação entre os sufixos derivativos e cada uma das palavras é maioritariamente idiossincrática (Schmitt & Meara 1997:26).

Na bibliografia, aparecem muitas referências a que certas **classes de palavras** são mais difíceis do que outras. Os nomes são geralmente

¹⁷⁴ Tenho algumas reticências quanto a esta forma de abordar a questão: 1. mistura-se aquisição da fonologia e da morfologia; 2. nada nos garante que o aprendente de L2 conhece as duas formas e que não é só por acaso que produz a forma semelhante. É provável que Laufer tivesse sido levado a esta abordagem pelo facto de os seus informantes serem falantes de línguas semíticas. Ver Capítulo 2, ‘vowel blindness’.

¹⁷⁵ «A special case of morphological difficulty in comprehension is what could be called ‘**deceptive transparency**’. Deceptively transparent words are words that look as if they were combined of meaningful morphemes.»(Laufer 1997:146). Ver Bauer & Nation 1993.

referidos como sendo os mais fáceis e os advérbios os mais difíceis, ficando no meio os verbos e os adjetivos (Laufer 1997:148). Mas é claro que cada classe alberga um número muito diferente de elementos em relação às outras e que cada uma delas é composta por elementos com características muito diversas. É verdade que, nos primeiros estádios de aquisição, o adulto, tal como a criança na sua L1, usa mais **nomes** do que verbos, mas isso deve-se com certeza ao carácter mais referencial dos primeiros ¹⁷⁶. A esse propósito, Dietrich (1990:21) chama a atenção para que é possível contar uma história só com nomes mas não só com verbos, enfatizando assim o valor pragmático da referência nominal na comunicação básica. No entanto, os **verbos** são muito mais centrais no desenvolvimento lexical. Vimos isso a propósito dos estudos de Viberg. Também Broeder, Extra, von Hout & Voionmaa 1993, que estudaram o desenvolvimento lexical da L2 de imigrantes, concluíram que quanto mais alta é a proporção de nomes no léxico do aprendente mais baixa é a riqueza lexical. Além disso, sendo a variação morfossintáctica dos verbos tão grande, que critérios são usados para afirmar que se sabe ou não um verbo por oposição a um nome? Só porque surgiu uma das suas formas no discurso? E como lidar com a dificuldade que constitui incluir ou não uma palavra na classe dos advérbios (ver Ramat & Ricca 1994) ¹⁷⁷? E com o facto de o número de adjetivos existentes em diferentes famílias de línguas poder variar tanto (ver Dixon 1977) ¹⁷⁸? Aitchinson (1987/1994:105-107) afirma que **há evidência suficiente para considerar os nomes, os verbos e os adjetivos como classes relativamente separadas no léxico mental,**

¹⁷⁶ «The preference for **nouns** shown by **aphasics** is not just because there are more nouns than verbs. It is more likely because nouns are relatively free of syntactic restrictions. **Verbs**, on the other hand, are somewhat more tricky, perhaps because they are inextricably entangled with the syntax of the sentence.» (Aitchison 1987/1994:102).

¹⁷⁷ «In the case of **adverbs**, words which behave somewhat differently are lumped together under the same label – though it is possible to identify some ‘core’ adverbs, often regarded as manner adverbs [...] However, the whole set of words labelled adverbs shows, first, that care should be taken in assigning words to words classes; and second, that we are unlikely to be able to work out with any accuracy how many different word classes are distinguished by the mental lexicon.» (Aitchison 1987/1994:106).

¹⁷⁸ «[N]ot all languages have the major word class **Adjective**. Either they have no Adjective class at all, or else there is a small non-productive minor class that can be called Adjective.» (Dixon 1977:20). Quando comparados com nomes, «Adjectives are rather different. They are less independent, and often rely for their interpretation on the noun to which they are attached.» (Aitchison 1987/1994:104).

enquanto que os advérbios têm comportamentos muito diversificados. Quanto às preposições, elas diferem muito entre si no que respeita o seu conteúdo semântico, e cada uma delas também pode variar bastante consoante o contexto linguístico ¹⁷⁹. E esta é uma das razões por que é impossível estabelecer uma divisão entre palavras de conteúdo e palavras funcionais.

Como vimos, também as várias **propriedades semânticas** das palavras costumam ser apontadas como podendo afectar a sua aprendizagem. Laufer (1990b e 1997) refere serem abstractas, serem específicas ou restringirem-se a um determinado registo, terem múltiplos sentidos e terem valores idiomáticos. No entanto, o facto de serem **concretas ou abstractas** é com certeza muito mais válido no caso da aquisição de L1, quando o desenvolvimento cognitivo se vai fazendo em paralelo com o desenvolvimento linguístico. Aquilo que, segundo Ellis & Beaton 1993 parece influenciar a aprendizagem não é serem concretas ou abstractas mas sim a existência ou não de uma representação imagética associada ao conceito a ser aprendido. Em relação a serem ou não **específicas**, vimos também, a propósito das crianças a aprenderem Sueco como L1 ou L2, que as primeiras usam verbos mais específicos enquanto que as segundas usam tendencialmente termos mais genéricos (verbos nucleares). Blum & Levenston 1978 chamam a este processo simplificação lexical (*lexical simplification*), dizem que ele opera de acordo com princípios universais ¹⁸⁰ e que ele pode encontrar-se no discurso de falantes de L2 mas também numa série de outras situações: em tradução, no discurso dos professores, quando falamos com estrangeiros, nos *pidgins*, e em textos simplificados para aprendizagem de línguas. Não é difícil prever que esta preferência por termos genéricos, por termos menos marcados, atinja o léxico específico de certos **registos** e que também aí serão preferidas palavras mais neutras em vez daquelas que exigem maior contacto com a

¹⁷⁹ Pacientes afásicos fornecem excelente evidência empírica para que é mais fácil produzir preposições se elas têm um conteúdo semântico intrínseco. (Aitchison 1987/1994:107).

¹⁸⁰ «The universal principles involved are probably based on systematic relationships between lexical items such as hyponymy, synonymy, antonymy, and converseness. The awareness of these relationships, together with the ability to use circumlocution and paraphrase, is part of every speaker's semantic competence and enables him, when the need arises, to express complex meanings by indirect means.» (Blum & Levenston 1978:400).

comunidade de falantes (Laufer 1997:151) ¹⁸¹. E depois do que dissemos sobre **idiomaticidade e polissemia** no Capítulo 1, também não é difícil prever que quanto menos prototípico é o uso de um determinado item mais difícil será a sua aquisição (ver, entre outros, Ijaz 1986, Ellis 1997) ¹⁸².

Assim, perante o obstáculo que constitui estudar a dificuldade de aquisição de uma classe em relação a outra, o que parece mais promissor é **estudar a dificuldade relativa de palavras dentro de uma mesma classe e, ao mesmo tempo, avaliar como é que estes diferentes factores intralexicaes afectam a aquisição quando interagem entre si** ¹⁸³. No entanto, ao fazer esta avaliação, é indispensável ter em consideração aquilo que foi dito acima: os diversos aspectos da aquisição de uma palavra dependem de diferentes mecanismos de aprendizagem. Segundo Nick Ellis (1994, 1996, 1997), a aquisição dos aspectos formais e sintácticos resulta sobretudo de processos de análise inconsciente (ou implícita) das sequências, enquanto que a aquisição das propriedades semânticas e conceptuais resulta de processos de aprendizagem consciente (ou explícita).

3.5. *Sumário*

Na Introdução a esta obra apresentei, em breve apontamento, o estado dos conhecimentos em aquisição do léxico em L2. Mas aí, transpareceu mais o que falta fazer do que aquilo que tem sido feito. Aqui, dei conta, ainda que de forma abreviada, porque a extensão da bibliografia sobre o assunto a isso obriga, de uma parte da investigação realizada e que pode ser relevante para esse trabalho. Infelizmente, a extensão a que me refiro não diz respeito ao Português mas a outras línguas, muito em especial ao Inglês, ao Alemão e ao Sueco. Procurei que ficasse claro, nessa Introdução, que o trabalho realizado tem sido mais no sentido de investigar

¹⁸¹ «L2 learners will also [like L1] have different mastery of the various kinds of word knowledge, with formal, grammatical, and meaning aspects probably learned first, and some other aspects, such as collocational behavior and **register**, perhaps never being mastered at all.» (Schmitt & Meara 1997:18).

¹⁸² «Speaking natively is speaking idiomatically. (...) An important index of nativelike competence is that the learner uses idioms fluently.» (N Ellis 1997:129-130).

¹⁸³ «It would be interesting and useful to investigate how these features affect the learning difficulty when they interact with each other in different ways. For example, are specific words with ease pronunciation easier or more difficult to learn than general words with difficult pronunciation? To my knowledge, no such studies seem to have been carried out yet.» (Laufer 1990b:304).

vocabulário do que **léxico**. De facto, como veremos no capítulo seguinte, *Léxico e ensino de L2*, porque os estudos estavam particularmente preocupados com o ensino, o trabalho desenvolvido numa primeira fase centrou-se bastante em estudar estratégias de inferência e de retenção de palavras (para listagem ver Schmitt 1997:207-8), ou seja, na pesquisa de como se vai constituindo um vocabulário no sentido de expandir a sua dimensão, em particular no que respeita o significado das palavras, mas não as suas conexões; mais recentemente, e tendo por detrás os últimos desenvolvimentos em linguística, tem-se caminhado mais no sentido de estabelecer o que se entende por **saber uma palavra** e o que é que torna umas mais difíceis do que outras, ou seja, quais os factores que intervêm na sua '**aprendibilidade**'. No entanto, e embora não possamos esquecer que muita investigação feita sobre as diferentes componentes da gramática interessa ao estudo do léxico, essa investigação, como disse, salvo algumas excepções, não tem relacionado a dimensão do vocabulário conhecido com a profundidade do conhecimento de cada uma das palavras que o constituem.

Um dos objectivos que tem dominado a investigação em aquisição de L2 tem sido encontrar regularidades que permitam estabelecer sequências de aquisição. A escolha de uma ou de outra componente das línguas para o estabelecimento dessas sequências tem sido fortemente influenciada pela atenção que, no momento, a teoria linguística lhes tem prestado. Sendo, actualmente, o léxico considerado central tanto para as teorias linguísticas como para a comunicação e para a aquisição, encontrar regularidades neste âmbito pode constituir mais um passo para essa tarefa.

Capítulo 4

Léxico e ensino de L2

4.1. *Introdução*

Language teaching has always been a discipline dependent on the theoretical capacities of ‘parent disciplines’ such as psychology, linguistics, pedagogical ‘methodology’. The discontinuity that characterizes the history of language teaching – the succession of ‘new waves’ in teaching methods – is due to misapplication of findings from parent disciplines, and to the lack of an independent theoretical basis for teaching itself (Pienemann 1989:52).

Chegados aqui, afirmar que a investigação em aquisição deve informar o ensino e que o léxico deve ser central no ensino de uma língua afigura-se-nos como tão óbvio que mais parecem lugares-comuns. No entanto, afirmações que neste momento nos soam a banalidades não têm sido levadas tão a sério no âmbito do ensino de L2 como seria desejável, até mesmo recentemente.

No espaço deste capítulo, começarei por fazer uma breve história do lugar do léxico (a maior parte do tempo, do vocabulário) no ensino, procurando relacioná-la com as teorias linguísticas dominantes na altura e com as metodologias de ensino que elas foram motivando. Referirei, em seguida, investigação desenvolvida em léxico para o ensino, nomeadamente o estabelecimento de vocabulários fundamentais, e o movimento pendular entre as posições daqueles que defendem que o vocabulário deve ser ensinado explicitamente e aqueles que consideram que a sua aprendizagem deve ser implícita. Finalmente, assumindo que o ensino deve, na medida do possível, respeitar os estádios de desenvolvimento, retomarei as propostas da abordagem lexical e da abordagem cognitiva relevantes para o ensino de L2.

4.2. *Das listas de palavras às listas de actos de fala*

Apesar de o léxico ser central na aprendizagem de uma língua, e de ser sentido como tal pelos aprendentes, a investigação e o ensino nem sempre lhe prestaram a atenção que merecia. O léxico, melhor dizendo o vocabulário, foi geralmente visto como um suporte, tanto da gramática como da competência comunicativa, de tal modo que, especialmente até aos anos 80, esses termos raramente aparecem mencionados nos índices das lições ou no livro do professor.

O **Método Gramática-Tradução**, que vem de finais do século XVIII, preparava os estudantes para ler textos clássicos, para escrever e para os exames, sabendo-se que raramente teriam ocasião de falar a língua que aprendiam. Aos estudantes eram dadas explicações de gramática, geralmente na sua L1, e listas bilingues de vocabulário para memorizarem. Muito desse vocabulário era literário, em muitos casos fora de uso, muitas vezes descontextualizado e organizado por ordem alfabética. Era dito aos estudantes que deviam aprender as palavras, mas não lhes eram fornecidas quaisquer estratégias para isso.

Este método, ainda que tivesse sido muito criticado, foi usado até aos anos 20 do nosso século tanto nos Estados Unidos como na Europa. Algumas dessas críticas, que vinham sendo feitas já desde meados do século XVIII, centravam-se na falta de realismo, nomeadamente, no que respeita a oralidade e no facto de uma parte do vocabulário ser arcaico. François Gouin 1892 reconhecia a necessidade de ensinar termos gerais, sem os quais a língua não pode passar. E Thomas Prendergast inclui no seu manual de 1864 uma lista das palavras que intuitivamente considerou como sendo mais frequentes em Inglês. Infelizmente nenhum destes procedimentos inovadores teve grande impacto na metodologia de ensino adoptada na época (Zimmerman 1997:6).

Em finais do século XIX, surge em Inglaterra '*The Reform Movement*' liderado por Henry Sweet, que tenta ancorar as suas propostas nos estudos linguísticos. Este movimento chama a atenção para o primado do oral e para a necessidade de treino fonético. O estudo da fonética e da transcrição constituía o primeiro estágio do *curriculum* (*mechanical stage*), que continuava com o estudo da gramática e do vocabulário básico (*grammatical stage*), a que se seguia o estudo do vocabulário em profundidade (*idiomatic*

stage). Os estádios seguintes (*literary* e *archaic*), de nível universitário, consistiam em estudos filológicos. Segundo Zimmerman (1997:7), nas lições de Sweet, só depois de o texto completo ter sido cuidadosamente analisado, eram destacados itens vocabulares ou aspectos gramaticais como objecto de estudo ¹⁸⁴; a língua falada era cuidadosamente controlada; eram evitadas as listas de palavras isoladas e até mesmo de frases separadas; e, na selecção do vocabulário, era tida em consideração a simplicidade e utilidade de cada palavra.

O **Método Directo**, cujo nome vem da tentativa de associar directamente o significado à língua alvo sem passar pela tradução, desenvolveu-se nos Estados Unidos, a partir de finais do século XIX, e considera a interacção fundamental na aprendizagem. A língua em estudo é usada como língua de instrução e os diálogos são cuidadosamente escolhidos, em função de uma sequencialização pré-estabelecida. Este método teve muito êxito em escolas privadas, nomeadamente nas *Berlitz Schools*, mas raramente foi adoptado pelas escolas oficiais tanto na Europa como nos Estados Unidos da América.

O vocabulário frequente constitui a base das lições; o vocabulário concreto é apresentado com recurso a imagens e o abstracto por associação de ideias. No livro do professor de *An Introduction to the Teaching of Living Language without Grammar or Dictionary*, Sauveur 1874 propõe dois princípios básicos para a instrução e para a selecção do vocabulário: **1.** o professor, ao fazer perguntas, deve parecer genuinamente interessado na resposta; **2.** as perguntas devem ter um encadeamento natural, de modo que o estudante possa aprender apoiado no contexto. Sendo assim, o vocabulário será necessariamente simples e familiar (Zimmerman 1997:9).

Por volta de 1920, nasce nos EUA o *Reading Method* e em Inglaterra o *Situational Language Teaching*. Estes movimentos criticam o facto de o Método Directo acentuar a importância da oralidade sem, no entanto, fornecer linhas de orientação para a selecção dos conteúdos. Michael West, em Inglaterra, chama a atenção para que, depois de três anos de estudo, os

¹⁸⁴ «Although language is made up of words, we do not speak in words, but in sentences. From a practical, as well as a scientific, point of view, the sentence is the unit of language, not the word. From a purely phonetic point of view words do not exist.» (Sweet 1899/1964:97 – citado por Zimmerman 1997:7).

estudantes de Inglês como língua estrangeira não dispõem de um vocabulário básico porque muito tempo foi gasto com tarefas que não os ajudam a falar a língua, porque aprenderam palavras que não lhes são necessárias, e porque não as dominam completamente (West 1930, cit. por Zimmerman 1997:9). Michael West recomenda o uso de listas de frequência. Para isso propõe o uso do dicionário de frequências de Thorndike, publicado em 1921. Este foi elaborado a partir de um *corpus* constituído por textos escritos em Inglês e contém uma lista de 10.000 formas, que foi alargada para 20.000 em 1932, e para 30.000 em 1944 por Thorndike e Lorge. Embora não tivesse sido elaborada para ser usada no ensino, esta lista, publicada por West em 1953, é conhecida como *A General Service List of English Words* e tem sido usada em muitos países até mesmo como base para a preparação de outras (Van Els *et alii* 1984:207).

Paralelamente, H. E. Palmer e A. S. Hornby, integrados neste movimento e particularmente influentes tanto em Inglaterra como nos EUA, também reagem contra a falta de gradação que o Método Directo favorecia. Defendem que o ensino deve praticar estruturas básicas em actividades com sentido; consideram o vocabulário como um dos aspectos mais importantes na aprendizagem de uma L2, e trabalham no sentido de desenvolver uma base racional e científica para a selecção dos conteúdos vocabulares dos cursos de línguas (Zimmerman 1997:10). Ronald Carter e Michael McCarthy, dois dos primeiros investigadores a dedicarem particular atenção ao léxico nos anos mais recentes, afirmam que este trabalho em lexicometria, iniciado nos anos 30, continua a ser ainda hoje uma fonte de inspiração e que muitas das questões levantadas estão actualmente a ser investigadas de novo (Carter & McCarthy 1988:11). J R Firth 1935/1957, por exemplo, foi o responsável pela proeminência dada à palavra '*collocation*'. Ele considera que o significado opera a diferentes níveis e as combinatórias (*collocations*) são exactamente um deles. Assim, tão importante como o significado de uma palavra são aquelas com as quais ela coocorre. McIntosh 1961 recupera esta ideia e observa que as palavras têm só uma certa tolerância de compatibilidade e que há certas combinações de palavras que são independentes de considerações de ordem gramatical e que, por conseguinte, necessitam de uma abordagem lexical (Carter & McCarthy 1988:33).

Desenvolvido durante a II Guerra Mundial, o **método audio-oral** ou abordagem estrutural, que combina a teoria do comportamento de Skinner com a linguística estrutural, sugeria que as dificuldades de aprendizagem resultavam do conflito entre diferentes estruturas. Charles Fries (1945:9), apoiado em princípios estruturalistas, postula que os materiais de ensino mais eficazes são aqueles que se baseiam numa descrição científica da língua a ser ensinada, cuidadosamente comparada com uma descrição da língua materna daquele que a aprende. Utilizando noções behaviorísticas como *imitação*, *interferência positiva* e *interferência negativa*, a Linguística Aplicada estabelece uma relação muito forte entre as propriedades estruturais da L1 e da L2 e o grau de dificuldade que o falante da primeira teria em relação à segunda: os elementos semelhantes seriam fáceis para ele e os elementos diferentes seriam difíceis (Lado 1957). A aprendizagem era encarada como se se tratasse da formação de um hábito; era dada atenção sistemática à pronúncia e às estruturas gramaticais. Exercícios estruturais orais ocupavam a maior parte dos tempos lectivos. Os estudantes eram encorajados a imitar os sons produzidos por falantes nativos (recorria-se bastante a laboratórios de língua) e a memorizar diálogos, na esperança de que o estudante adquirisse consciência das estruturas gramaticais neles exemplificadas. O vocabulário era visto como um problema que era preciso controlar, e só era introduzido na medida em que servia as estruturas a serem aprendidas. Segundo Zimmerman (1997:11), Charles Fries (1945:38) suspeitava que os estudantes assumiam que as palavras têm equivalentes exactos em diferentes línguas o que lhes dá uma falsa sensação de segurança; para evitar essa possível simplificação abusiva era recomendada a prática da morfologia e das estruturas sintácticas usando vocabulário muito restrito e, na medida do possível, já conhecido.

Enquanto Chomsky (1957, 1959) previa a existência de um dispositivo inato (*language acquisition device* – LAD) que facilitaria a aquisição das regras gramaticais abstractas, e estabelecia uma distinção forte entre competência e *performance*, reforçando a ideia defendida pelo estruturalismo e pela análise contrastiva de que o vocabulário é de certo modo periférico, por seu turno, a psicologia do desenvolvimento (Brown 1973) analisava o discurso de crianças e concluía que, apesar de a

produção ser mínima, ela é suficiente para reflectir bastante criatividade e capacidade de conceptualização. Observava que a criança não só usa as palavras para nomear objectos mas também para exprimir pensamentos mais complexos que envolvem esses objectos; observava ainda que as crianças preferem palavras plenas e que há uma baixa ocorrência de palavras gramaticais (Steinberg 1982:153).

Na linha destes estudos que prestam particular atenção àquilo a que Chomsky 1987 chamará *E-language*, e como reacção contra a autonomia da *competência linguística*, Dell Hymes 1972 enfatiza os factores sociolinguísticos e pragmáticos da linguagem e introduz o conceito de *competência comunicativa*. Pondo a tónica na fluência, em detrimento da correcção, e na proficiência comunicativa, desviando-se da preocupação com a estrutura gramatical, o ensino tinha como principal objectivo pôr o estudante em contacto com a língua alvo tal qual ela é usada pelos falantes nativos. De facto, nos anos 70, o método gramática-tradução e a abordagem estrutural continuavam vivos e de boa saúde, preparando os estudantes para os exames, mas não os preparando para tarefas comunicativas tão simples como comprar um bilhete de comboio ou encomendar um prato num restaurante.

David Wilkins 1976 faz notar que há duas categorias de sentido envolvidas na comunicação: *categorias nocionais* (conceitos como tempo, quantidade, espaço...) e *categorias funcionais* (tais como afirmar, pedir, argumentar...). Esta abordagem, baseada em critérios temáticos e situacionais, põe necessariamente em destaque a semântica e não a sintaxe; no entanto, embora não negando a sua importância – é o próprio Wilkins (1972:111) quem afirma: «*Without grammar very little can be conveyed, without vocabulary nothing can be conveyed*» –, os metodólogos não prestam ao vocabulário nenhuma atenção especial. Ele é, no dizer de Felicity O' Dell (1997:259), subserviente da gramática e dos objectivos comunicativos.

De facto, o que aconteceu durante os últimos 30 anos, com a chamada **abordagem comunicativa**, foi o abandono sistemático de métodos mais tradicionais que punham a ênfase na forma (na gramática e no vocabulário) para se prestar atenção à linguagem enquanto meio de comunicação, pondo a tónica na fluência, em detrimento da correcção, mas

sobretudo na adequação. O locutor deve ser capaz de escolher, em função das suas intenções comunicativas, entre várias possibilidades de expressão linguística, tendo em conta os seus interlocutores e o tipo de situação em que se encontra (Coste *et alii* 1976:37). Mas, os autores de *Un niveau-seuil* advertem que ele não é «une liste de mots ni un répertoire de phrases utiles». No entanto, parecem não convencer Sinclair e Renouf

In recent years, the specification of syllabus has changed. [...] Language is viewed as a list of potential acts, and language behaviour is a succession of such acts. [...] There is no comprehensive theory of language in these terms available as yet, so such syllabuses rest on shaky ground. [...] In the new syllabuses, the structural frameworks are largely ignored, and no criteria are provided for distinguishing functions and other features which appear to overlap (Sinclair & Renouf 1988: 141).

É o domínio, por vezes excessivo, das listas de actos de fala. Apesar de os autores insistirem que estes materiais devem ser usados com ‘*souplesse*’ e em função do público, na prática, e porque eles tinham potencialidades para isso, nem sempre assim aconteceu e, muitas vezes, passou-se das listas de palavras para as listas de actos de fala.

Mas é a partir do início dos anos 90 que os investigadores começam a chamar sistematicamente a atenção para as limitações desta abordagem. Entre outros, Amrhein 1992 refere que a compreensão de actos de fala é diferente da de outros aspectos da compreensão linguística, porque depende de outros mecanismos de reconhecimento do ouvinte e das intenções ou crenças do falante. E Laufer (1990:293) que, como vimos, tem trabalhado no sentido de apurar como é que certos factores intralexicais afectam a aprendizagem das palavras, concorda com Rivers (1983:120):

Rivers says that the time has come to “consider carefully how we can provide even our elementary learners, and much more so our advanced learners, with the means to ‘get across meaning’, even before they can express discriminately fifteen ways to ask that the door be opened (Laufer 1990:293).

4.3. *Investigação em léxico para o ensino de L2*

Apesar de ser mais conhecida a chamada de atenção de Paul Meara, em 1980, para o facto de o vocabulário ser negligenciado no ensino, foi Lord 1974 o primeiro a levantar essa questão:

*Yet, vocabulary goes on being treated as the cinderella. In the applied linguistic literature discussion of methods of teaching vocabulary are rare indeed; and published FL courses have invariably one thing in common: an apparent lack of awareness of the nature and function of vocabulary, or at any rate a reluctance to face up to it. With sometimes barely more than passing lip-service to the importance of situation and context, **vocabulary tends to be reduced to an appendage, uncomfortably marshalled into words lists, and not seldom reduced to the ultimate absurdity of substitution drills** (Lord 1974:239).*

A partir de então, sucessivas referências à importância do léxico e à sua situação de abandono começam a surgir. Widdowson (1978) declara que é mais fácil, para um falante nativo, perceber uma frase agramatical com vocabulário correcto, do que uma frase gramatical mas com vocabulário inadequado (Zimmerman 1997:13). Levenston, em 1979, diz que o léxico tem sido uma vítima de discriminação. Segundo ele, há uma série de questões que precisam de ser investigadas: em que medida a aprendizagem do vocabulário de L1 e de L2 são processos paralelos? Em que medida características pessoais afectam a aprendizagem e uso do vocabulário? Como cresce o vocabulário de um aprendente de L2, como se expande, por que estádios passa, e quais os factores que influenciam este processo? Sob que circunstâncias e influenciado por que factores progride o significado das palavras desde ser vagamente reconhecido até ser identificado todo o seu alcance, desde ser usado em certos contextos até ser produzido sempre que necessário? Todas as palavras passam pelos mesmos estádios? (Levenston 1979:147-148).

Uma das razões que aponta para esta falta de interesse é que os problemas mais interessantes do desenvolvimento lexical só emergem relativamente tarde no desenvolvimento da L2. Observações como esta mostram claramente uma visão do léxico completamente independente da gramática. No entanto, já em 1976, Richards, ao levantar a questão 'o que significa exactamente saber uma palavra?', começara a caminhar no sentido de abandonar a ideia de que o léxico consiste em listas de palavras de que é preciso aprender o significado e que essas palavras nada têm a ver com a gramática.

Mas, só a partir do início dos anos 80, a investigação e o ensino começam, de facto, a prestar-lhe atenção. Ainda que Krashen considere que

o significado de uma palavra é adquirido de forma totalmente inconsciente, a sua *Natural Approach* (Krashen 1981), e em particular a *Input Hypothesis*, ao afirmar que a aquisição não terá lugar se o *input* não for compreensível, interessante e relevante, contribuiu bastante para a tomada de consciência de que o vocabulário deve ser planeado, sobretudo nos primeiros níveis. Para níveis mais avançados, a leitura é vivamente recomendada. No entanto, por esta altura, o vocabulário ainda continua a ser visto como um concorrente da gramática e a principal preocupação consiste em saber como dosear léxico e gramática nos materiais de ensino:

There is no conclusive evidence for the view that communication may well be more efficient with good vocabulary accompanied by minimal structural competence; nor for the opposing view that massive drilling of structure at the earliest stages achieves a high degree of grammatical competence which only needs to be supplemented by vocabulary expansion. *A survival-level course might well wish to place emphasis on building up a large vocabulary from the beginning, even at the expense of structural competence. A larger vocabulary than at present offered would make comprehension easier for the learner, would equip him with at least a phrase-book style set of words to enable a basic communication to take place (assisted by immediate environment, gesture, mime, etc.) and might extend listener tolerance (McCarthy 1984:14).*

De facto, subjacente à história dos métodos de ensino, há uma discussão silenciosa e permanente entre aqueles que defendem que o ensino do vocabulário deve ser explícito e aqueles que defendem que cada palavra é aprendida inconscientemente, resultando a sua aprendizagem do contacto repetido com ela e que, portanto, o seu ensino deve ser implícito. Segundo Nick Ellis 1995 podemos distinguir quatro posições:

- 1.** *a hipótese forte de aprendizagem implícita* defende que o significado de uma palavra nova é adquirido de modo totalmente inconsciente como resultado da exposição repetida a uma série de contextos em que ela é activada; Krashen e a sua *input hypothesis* ilustram esta posição;
- 2.** *a hipótese fraca de aprendizagem implícita* propõe que, provavelmente, a aquisição não será possível sem, pelo menos, se tomar consciência de que se trata de uma palavra nova;

3. *a hipótese fraca de aprendizagem explícita* defende que, sendo o ser humano um processador de informação, o facto de a maioria das palavras não nos terem sido ensinadas não significa que não tenhamos usado certas estratégias para as aprendermos. Assim, parece podermos tirar proveito de prestar atenção a um item novo; e também será proveitoso usarmos uma variedade de estratégias tanto para inferirmos o significado a partir do contexto como para a retenção desse mesmo significado;

4. *a hipótese forte de aprendizagem explícita* preconiza que a aplicação de um conjunto de estratégias metacognitivas facilita fortemente a aquisição de vocabulário novo, verificando que se trata de uma palavra nova, tentando inferir o seu significado a partir do contexto ou consultando um dicionário, e procurando consolidar este novo conhecimento através da repetição e de outras estratégias de associação semântica ou imagética.

Tendo em consideração que aprender uma palavra não consiste exclusivamente em aprender o seu significado, Nick Ellis, neste e em outros trabalhos (Ellis 1994b, 1995a, 1995b, 1996, 1997, 1999), procura determinar a que diferentes componentes se aplica cada uma destas hipóteses. Estabelece uma distinção entre o significado e os aspectos formais dos itens lexicais, e propõe que *a hipótese forte de aprendizagem explícita* é compatível com a aquisição do significado enquanto que *a hipótese fraca de aprendizagem implícita* é suficiente para garantir a aquisição dos aspectos estatísticos e as regularidades ortográficas e fonológicas; ou seja, enquanto para o significado quanto mais profundo for o seu processamento melhor será a sua aquisição, no que respeita os aspectos formais quanto mais um padrão se repete no *input*, quanto mais frequente ele for, mais rápida e melhor será a sua aquisição.

4.3.1. *Vocabulários fundamentais*

Na linha daqueles que defendem que o ensino do vocabulário deve ser controlado e ensinado explicitamente estão as listas de palavras. A primeira foi feita para o Inglês, tem as suas raízes em Ogden & Richards

1923, continuou a ser desenvolvida pelo primeiro nos anos seguintes, e insere-se no *Vocabulary Control Movement*. Consiste em 850 palavras: 600 nomes, 100 adjetivos, 50 antónimos destes formados pelo prefixo *un-* (*unable, unregular...*), mais 100 ‘palavras operacionais’ (*‘operation words’*) que incluem 18 verbos, preposições, pronomes e outras palavras gramaticais. É, por vezes, referido como *‘basic English’*, mas, na verdade, não se trata de uma lista de vocabulário e estruturas fundamentais da língua inglesa. De facto, o termo *‘BASIC’* é o acrónimo de *British American Scientific International Commercial* e este projecto pretende ser um sistema linguístico completo. O critério de selecção dos vocábulos não é a frequência mas a sua capacidade de cobertura, ou seja, as possibilidades que apresentam de, quando combinados com outros da lista, substituírem aqueles que dela não fazem parte ¹⁸⁵. Este projecto foi divulgado por todo o mundo, mas, na verdade, ele é um convite à formação de muitas expressões que não existem em Inglês. Ele inscreve-se, tal como o Esperanto, na procura de uma língua universal que resolva os problemas de comunicação entre os povos (Theo van Els *et alii* 1977/1984:215).

Ainda que, durante os anos 50 e 60, os *curricula* e os materiais de ensino fossem organizados em função de uma sequencialização dos conteúdos gramaticais, e os dicionários tivessem sido mantidos fora das aulas de língua (Rogers 1996:71), na linha de Harold Palmer, e à margem da teoria linguística, os estudos de léxico vão-se desenvolvendo com vista a apoiar o ensino. Inscreve-se neste contexto, para o Inglês, *A General Service List*, de 1953, antes referida. Mas é sobretudo em França e na Alemanha que eles adquirem, a partir deste período, uma certa proeminência (Stern 1983:131).

Insere-se neste quadro o *Français élémentaire*, preparado por Gougenheim, Michéa, Rivenc e Sauvageot, em 1956, e que, a partir de 1959, passou a chamar-se *Français Fondamental*. Esta é a primeira lista de palavras para o ensino do Francês como língua estrangeira compilada com base no Francês falado. Foram entrevistados 275 falantes nativos, de

¹⁸⁵ «Le **Basic English** créé à partir de l’anglais, est une langue artificielle au vocabulaire rigoureusement réduit mais susceptible, par le biais de la composition syntagmatique et de la paraphrase, de répondre – en principe – à tous les besoins sémantiques d’expression des locuteurs.» (Coste *et alii* 1976:35).

ambos os sexos, de diferentes idades, regiões e profissões. O *corpus* resultante compreendia 312.135 ocorrências. Inicialmente, o projecto privilegiou o critério frequência. Foram introduzidas na lista 1063 palavras, isto é, aquelas que obtiveram uma frequência superior a 20. Mas, ao analisarem a lista, verificaram que dela não faziam parte certas palavras muito frequentes. Decidiram então realizar um inquérito de disponibilidade e incorporar também na lista vocábulos assim obtidos. Uma comissão de linguistas decidiu ainda excluir alguns vocábulos (entre eles, ‘les mots à la mode’ e ‘les mots familiers et vulgaires’) e incluir 520 outros, muitos deles, segundo o próprio Rivenc (1979:22), de forma muito arbitrária ¹⁸⁶.

Os critérios de selecção usados para o Francês Fundamental foram aperfeiçoados e usados, durante os anos 60 e 70, em projectos semelhantes de outras línguas. Entre elas o Alemão, o Italiano, o Espanhol e o Português ¹⁸⁷. E apesar de eles poderem ser discutíveis, são com certeza mais fiáveis do que a introspecção, a intuição e a experiência. Não obstante, o *Lexicon for Threshold Level English* (1975) um apêndice ao trabalho pioneiro numa perspectiva comunicativa, realizado por uma comissão no âmbito do Conselho da Europa, é estes critérios que usa. A lista contém aproximadamente 1050 itens para uso produtivo e receptivo e mais 450 só para uso receptivo (Theo van Els *et alii* 1977/1984:181). Mas é certo que o repertório de estruturas aí incluídas (os ingleses chamam-lhes *language functions* e os franceses *actes de parole*) resultaram também de escolhas arbitrárias dos seus autores e não de inquéritos feitos aos falantes nativos. Os autores franceses reconhecem, no entanto, que «comme pour le *français fondamental*, des enquêtes de cette nature pourraient – devraient – être entreprises.» (Coste *et alii* 1976:36) ¹⁸⁸.

Mas a questão que se levanta com o aparecimento da abordagem comunicativa é como conciliar nos materiais comunicativos a frequência com a utilidade. A ordem das palavras numa lista nem sempre representa a melhor ordem pela qual elas devem ser ensinadas, sobretudo se tivermos em consideração as características e os fins específicos de cada aprendente.

¹⁸⁶ Para mais informação sobre listas de palavras e critérios de selecção, ver Carter & McCarthy 1988/ 1989 e Theo van Els *et alii* 1977/1984.

¹⁸⁷ Do *Português Fundamental*, porque ele constitui um dos materiais do estudo apresentado na parte II, falarei no capítulo 6 e 7.

¹⁸⁸ O *Nível Limiar* é o correspondente português deste projecto. Ver capítulo 7.

A presença de muitos estudantes estrangeiros em certas universidades, frequentando cursos muito diversos, nomeadamente na Holanda, em Inglaterra e nos Estados Unidos, e a necessidade de resolver o problema de estes não disporem de um vocabulário receptivo suficiente, tornou mais premente a distinção entre **vocabulário receptivo** e **vocabulário produtivo**. Este parece ser o motor de uma linha de investigação que propõe o ensino de um vocabulário 'âncora' (Kirsner & Lalor & Hirt 1993:226) para a compreensão, mas muito em especial para a produção, e que procura definir a dimensão mínima de um vocabulário receptivo para estudantes universitários não nativos.

Tem-se dito que, em línguas como o Francês, o Russo, o Inglês, ou o Neerlandês, as 5.000 mais frequentes palavras (ou 3.000 famílias) cobrem 90 a 95 % das ocorrências de um texto académico. Nation (1993:119) apresenta o seguinte quadro:

Percentage coverage of text for increasing vocabulary sizes	
Number of words	Coverage
1	7 %
10	25 %
100	50 %
1.000	75 %
2.000	81 %
5.000	89 %
43.831	99 %
86.741	100 %

Se estes números são válidos, um léxico com 5.000 palavras, se for suficientemente organizado, seria suficiente para atingir um nível adequado de compreensão. No entanto, segundo Arnaud e Sauvignon (1997:158) as restantes são, muitas vezes, as mais informativas. Além disso, assinalam, também não podemos esquecer a dificuldade que envolve a compreensão de unidades lexicais complexas. Esta questão põe-se de forma menos aguda em situações de interacção visto que, neste caso, é possível recorrer a estratégias de negociação. Hazenberg & Hulstijn 1996, que relatam uma investigação empírica para definir um vocabulário receptivo mínimo de Neerlandês para estudantes universitários não nativos, concluem que com menos de 10.000 lemas diferentes será muito difícil atingir um nível de compreensão satisfatório de textos académicos. Este número deve ser tido

em consideração no planeamento curricular de cursos para estrangeiros, já que estes incluem, frequentemente, nos níveis não básicos, para além da língua, disciplinas muito diversas, de que se destacam a literatura, a história e a geografia.

Nos anos 80, discutiu-se se haverá um **vocabulário nuclear**. Na verdade, os investigadores não se põem de acordo quanto à divisão entre palavras de alta frequência e de baixa frequência; concordam, no entanto, que essa divisão se situa entre as 1.500 de altíssima frequência e as 7.000 mais frequentes (Nation & Newton 1997:238). Carter, em 1987, acentuou que testes linguísticos não são suficientes para estabelecer esse tão discutido vocabulário nuclear, que são necessários testes psicolinguísticos, que é necessário ter em consideração a capacidade para as aprender e para as ensinar:

From a pedagogic point of view it is obvious that a purely linguistic determination of coreness in vocabulary, however systematic and detailed, will be insufficient for many purposes. One central problem is that of relationship between core words, their learnability and teachability. This involves such complex factors as the appropriate isolation of core meaning; decisions over which meanings to teach first, particularly in polysemous words; the relationship between memorization, lexical recall, and vocabulary items; the specification of usefulness in terms of learning purpose [...]; the fact that 'simple words' can be among the most problematic syntactically and semantically [...]; and the relationship between syntax and lexis in the process of simplification... (Carter 1987:188-189).

A investigação apresentada na Parte II desta obra insere-se nessa preocupação. Ela mostra que há, de facto, um vocabulário de altíssima frequência que deve ser trabalhado exhaustivamente para a produção e para a compreensão, sem esquecer as suas combinatórias mais frequentes ¹⁸⁹. Esse trabalho transformá-lo-á num léxico mental organizado e rico em conexões que constituirá uma teia em que se vem encaixar o vocabulário para a compreensão que deve ser muito mais vasto.

4.3.2. *Estratégias de inferência e de retenção*

¹⁸⁹ «Os dicionários de frequência, elaborados por Juilland *et alii*, mostraram que nas cinco línguas românicas (espanhol, português, francês, italiano e romeno) cerca de 80 % de qualquer texto é constituído pelas 500 palavras mais frequentes da língua, incluindo-se aí um conjunto de palavras de valor semântico muito geral e a totalidade das palavras gramaticais dessas línguas.» (Biderman 1996: vol III, 218).

Os últimos anos da década de 70 e os primeiros da de 80 ocuparam-se menos com a aquisição do vocabulário do que com o modo como manobrá-lo para que fosse mais facilmente aprendido. Surgem, neste período, e têm tido continuação até aos nossos dias, muitos trabalhos que se centram na busca de estratégias que facilitem a inferência a partir do contexto e a retenção de vocabulário novo. Um dos seus aspectos mais positivos desta linha de investigação foi ter contribuído para a tomada de consciência de que os aprendentes não têm dois vocabulários distintos – um vocabulário passivo e um vocabulário activo – mas que o léxico mental do falante de L2 é muito instável e que o conhecimento de cada item do vocabulário constitui um *continuum* que vai do reconhecimento à disponibilidade para a produção; podendo, no entanto, um mesmo item estar disponível para certos contextos e não para outros (ver, por exemplo, Bialystok & Sharwood Smith 1985).

Enquanto que a *hipótese forte de aprendizagem implícita*, de que a *input hypothesis* de Krashen é o melhor exemplo, enfatiza o papel do *input* simplificado e do contexto na compreensão e, conseqüentemente, na aquisição, e defende que não se justifica o ensino directo do vocabulário porque o estudante aprenderá aquele de que precisa através da leitura intensiva de textos ou através da exposição a *input* oral, Long 1981 argumenta que, de facto, *input* compreensível é necessário à aquisição, mas que aquilo que facilita a aquisição é a oportunidade de negociar soluções para os problemas de comunicação e as modificações operadas em interacção que essa negociação implica. Esta declaração ficou conhecida como *interaction hypothesis*.

Ellis, Tanaka e Yamazaki (1994:194) comentam que, muito provavelmente, aquilo que estes e outros investigadores tinham em mente, no início dos anos 80, quando falavam de aquisição, era a aquisição do sistema gramatical, mas parece razoável e adequado investigar a *interaction hypothesis* em relação à aquisição de vocabulário. Tendo presente que a aquisição de um item lexical consiste num processo complexo, fazem-no em relação ao significado, e concluem que, pelo menos no que respeita o significado de nomes concretos, a negociação em interacção promove a aquisição. No ano seguinte, Rod Ellis (1995) realiza um trabalho experimental com *input* modificado em interacção e com *input* pré-

modificado (ou seja, aquele em que as redundâncias são aumentadas e a complexidade sintáctica é diminuída) e conclui que o número de palavras aprendidas por minuto era maior no segundo caso. Rod Ellis faz notar que o facto de *input* pré-modificado ser mais eficiente para promover a aquisição do significado não quer dizer que isso seja válido para aspectos fonológicos, sintácticos ou outros de carácter semântico. Sendo escassos os trabalhos de investigação centrados em *input* oral, modificado ou pré-modificado, esta constatação é mais um argumento a favor do papel positivo que o ensino pode ter e uma acha na fogueira da discussão entre aqueles que defendem, até mesmo para principiantes, o uso de materiais autênticos ou o uso de materiais previamente modificados.

Investigação experimental como a realizada por Rod Ellis e colegas, ao evidenciar o papel do ensino justifica o trabalho daqueles que argumentam que devem ser fornecidas ao estudante algumas estratégias para processar o *input* que complementem as associadas à sua L1 ¹⁹⁰. Assim, ao contrário de uma abordagem tradicional em que o ensino do vocabulário era ocasional e deixado a cargo do estudante, a investigação realizada neste âmbito enfatiza o ensino de estratégias conducentes à aprendizagem do vocabulário e tem em consideração as necessidades e o estilo de aprendizagem de cada um ¹⁹¹ (Oxford & Scarcella 1994).

Alguns dos investigadores fazem um reparo prévio: é necessário distinguir **inferência** de **retenção**, **estratégias de inferência** de **estratégias de aprendizagem**. Ou seja, a estratégia de inferência usada pode consistir simplesmente em gerar um significado que sugira uma interpretação coerente da passagem em causa e em nada contribuir para a retenção da palavra (Lawson & Hogben 1996); por outro lado, um mesmo

¹⁹⁰ «The classification system that seemed best to capture the nature of **learner strategies** reported by students in these studies [‘think aloud’ interviews included listening to and reading academic content materials] was based on the distinction in cognitive psychology between metacognitive and cognitive strategies together with a third category for social/affective strategies. [...] Examples of strategies in each of these categories are: metacognitive strategies for planning, monitoring, and evaluating a learning task; cognitive strategies for elaboration, grouping, inferencing, and summarising the information to be understood and leaned; and social/affective strategies for questioning, co-operating, and self-talk to assist in the learning process.» (Chamot & O’Malley 1994:374).

¹⁹¹ «Learners from different culture groups sometimes have quite different opinions about the usefulness of various vocabulary learning strategies.» (Schmitt 1997:202).

contexto pode facilitar a compreensão mas não a retenção (Mondria & Wit-De Boer 1991).

A investigação sobre **estratégias de inferência** centra-se no papel dos cognatos, na informação que o contexto pode fornecer, e nos factores que afectam a inferência a partir do contexto (ver, entre outros, Haastrup 1985, Na & Nation 1985, Li 1988, Nagy 1997). Laufer 1997 faz notar que se acredita demais nas deixas que o texto oferece ao leitor para a inferência do significado, que o contexto pode criar mal entendidos, e que quanto menos palavras se sabe mais difícil é entrar no texto ¹⁹². Nick Ellis 1997 refere que o facto de uma palavra em L2 nos lembrar a palavra em L1, seja por razões ortográficas, fonológicas ou etimológicas, facilita a aprendizagem. Daí a importância dos **cognatos**. Meara (1993:284) afirma que, de facto, existe um grande número de cognatos entre as maiores línguas da Europa ¹⁹³ que o ensino não tem explorado por causa das classes multilingues. E Garcia 1991 conclui em trabalho experimental que a relação de cognaticidade nem sempre é transparente, que os estudantes, por vezes, não relacionam mesmo os óbvios e que, por isso, é preciso consciencializá-los. Também a **morfologia** é tida em consideração, na medida em que, nos processos de inferência, as regras que governam a formação de palavras na L1 restringem as hipóteses que os sujeitos põem acerca das regras lexicais da L2 e, conseqüentemente, interferem no processamento (Hancin-Bhatt & Nagy 1994).

Muitos dos que estudam a aquisição de vocabulário defendem que a maior parte é aprendido acidentalmente. No entanto, com base em estudos que atestam que os bons aprendentes são capazes de escolher e usar estratégias sofisticadas adequadas à aprendizagem do vocabulário, diferentes **estratégias de retenção** têm sido propostas para a sua aprendizagem. Entre elas destacam-se a técnica da palavra-chave e outras mnemónicas que consistem em criar associações semânticas, imagéticas ou fonéticas. No entanto, as conclusões vão geralmente no sentido de acentuar

¹⁹² «What I find hard to accept is taking for granted that guessing in L2 is indeed possible with most unknown words and that successful guessing depends mainly on learner's guessing strategies. This seems to be a naïve belief since a variety of factors will interfere with the guessing attempts of the reader.» (Laufer 1997:27-28).

¹⁹³ Ver nota 103:diferentes tipos de relação de cognaticidade (Meara 1993).

as suas limitações (Schmitt 1997, e Schmitt & McCarthy 1997:235)¹⁹⁴, de propor que o seu uso é mais eficiente quando combinado (Brown & Pery 1991), que elas ajudam mas não substituem os métodos mais naturais de aquisição accidental, e que devem ser usadas só no caso de palavras em que, por qualquer razão, haja uma manifesta resistência à aprendizagem espontânea (Prince 1996:481, Hulstijn 1997:220).

O papel da **leitura** na aquisição de vocabulário tem sido destacado e tem merecido muita atenção por parte de vários investigadores (ver, por exemplo, Huckin & Haynes & Coady (eds.) 1993). Nunan 1985 chama a atenção para o papel desempenhado pela **interculturalidade** e para a possibilidade de um mesmo texto ser afectado pelos esquemas mentais e pelo mundo experiencial do leitor. Associado à questão da leitura, tem merecido bastante atenção o papel do **dicionário**, e das definições que ele fornece, bem como o do **contexto**¹⁹⁵. Vários trabalhos têm procurado responder à questão se é melhor inferir/aprender a partir do contexto ou de definições. Nesi & Meara 1994 concluem que, muitas vezes, os utilizadores se fixam numa parte da definição sem, na verdade, perceberem a sua relação com a palavra que procuram, e que isso se deve ao modo como as definições estão redigidas e à falta de informação sobre restrições. No entanto, Luppescu & Day 1995 e Grabe & Stoller 1997 afirmam que o uso de dicionários bilingues de modo consistente e adequado parece ter um impacto positivo na aprendizagem do vocabulário e no desenvolvimento da leitura. Mas o que parece ser mais eficaz é o uso de duas estratégias combinadas: a apresentação da palavra em contexto seguida do seu

¹⁹⁴ «[...] strategies like the **Keyword Method** may well have their limitations. It is hard to imagine a learner retaining very large numbers of keyword images in their mind. Of course, one would assume that as words become connected in the mental lexicon in other ways, the keyword images would eventually be discarded. This raises the question of whether certain strategies are primarily short-term ones, and at which point they become less effective than, or even begin to compete with, unaided language processes.» (Schmitt & McCarthy 1997:235)

¹⁹⁵ «The context of a word can be described on a local as well as on a global level. The former is based on what linguists [...] have called the selectional restrictions and semantic subcategorization features of a word. [...] Words that are syntactically combined with a given word in a phrase form what will be referred to as its **lexical context**. [...] A word is not only embedded in the lexical context of other words, but also occurs within the description of some state of affairs, or some events. [...] the situational embedding of a word is called its **schematic context**.» (Fischer 1994:552-553).

isolamento e tratamento em um processo elaborativo ¹⁹⁶ de que pode fazer parte o recurso à tradução (Fischer 1994, Prince 1996).

À luz da investigação desenvolvida e da posição de Nick Ellis, o ensino directo de vocabulário é benéfico e necessário para muitos adultos. O trabalho na sala de aula que desenvolve a aplicação de estratégias de inferência e de retenção para o significado parece justificar-se e deve merecer a nossa atenção, mas a aprendizagem accidental ou indirecta é, sem dúvida, essencial e o principal motor para o desenvolvimento da linguagem (Oxford & Scarcella 1994:240). Daí que investigação sobre as características do *input* disponibilizado na sala de aula, como a levada a cabo por Rod Ellis e colegas, deva ser privilegiada não só para o significado mas também para aspectos fonológicos, sintácticos ou outros de carácter semântico.

4.4. *Ordem de aquisição, léxico e ensino de L2*

Trying to break a natural order is like trying to run before you can walk (Ellis & Laporte 1997:64)

Nunan (1991:253) recorda-nos que o objectivo final da investigação em aquisição de L2 consiste em explicar e predizer os estádios por que passa o aprendente e em identificar os processos graças aos quais adquire a língua alvo.

Um grupo de investigadores, estabelecido em 1974, e coordenado por Jurgen Meisel, está entre aqueles que mais se têm empenhado em identificar esses estádios e em aplicar ao ensino os resultados da investigação em aquisição. No âmbito de um projecto de investigação, conhecido pelo acrónimo *ZISA*, desenvolveram, durante 15 meses, estudos longitudinais e transversais sobre a aquisição do Alemão como L2 por crianças e adultos falantes de línguas românicas (Espanhol, Italiano e Português) ¹⁹⁷. A análise dos dados privilegiou sobretudo a **sintaxe** e a

¹⁹⁶ «When students learn by repetition, they go over the material to be learn by rote, with minimal linkages to existing schemata. In contrast, when students learn strategically by **elaboration**, they develop images of the new information, organise the information, relate what they know about the concepts to the new information, and in general, build strong and varied associations with existing memory.» (Chamot & O'Malley 1994:378)

¹⁹⁷ «**ZISA** (acronym for 'Zweitspracherwerb italienischer und spanischer Arbeiter' = second language acquisition of Italian and Spanish [immigrant workers]). [...] The empirical basis of this research involved 45 adult informants in a cross-sectional study, three-8-year old children in a one-year longitudinal study, and 12 adult learners in a two-year longitudinal

morfologia e os investigadores concluíram que o desenvolvimento destas passa por um conjunto ordenado de estádios. Esta sequência fixa é explicada por um certo número de estratégias de processamento dos enunciados (Meisel & Clahsen & Pienemann 1981). Os resultados quanto à aquisição da ordem das palavras do Alemão, que prevê uma sequência de quatro estádios fixos e predeterminados, é provavelmente a mais robustamente apoiada em evidência empírica. Esta sequência foi aplicada ao Inglês e é a base teórica do **Multidimensional Model**¹⁹⁸, que tem sido desenvolvido por Pienemann e Johnson, e pretende fornecer apoio para o ensino (Pienemann 1986, Pienemann 1989, Pienemann & Johnson & Brindley 1988).

Estes estádios de desenvolvimento podem ser acelerados pelo ensino mas, de acordo com *the teachability hypothesis*, o curso do desenvolvimento da L2 não pode ser alterado por factores externos ao aprendente. Esta hipótese prevê que **o ensino só pode favorecer a aquisição se a interlíngua se encontrar próxima do ponto em que a estrutura ensinada seria adquirida em situação de aprendizagem informal** (Pienemann 1989:60).

Ou seja, ter adquirido aquilo que caracteriza um estádio constitui um pré-requisito para o estádio seguinte. O grupo de Pienemann começou por dizê-lo no início da década de 80 e progressivamente tem vindo a afirmá-lo com mais convicção. Pienemann & Johnston & Brindley (1988:220) «is a concrete attempt to initiate the process of applying SLA findings to language pedagogy»; Pienemann (1989:63) afirma: «items in a syllabus need to be taught in the order in which they are learnable»; «if teaching is meant to promote acquisition – that is actual use of structures being taught – since it can be shown that ‘**storing up treasures in heaven**’ approach to learning,

study. These studies include a total of about 200 hours of recorded speech, which has been transcribed and analysed» (Pienemann & Johnston & Brindley 1988:220-221)

¹⁹⁸ O modelo tem duas dimensões. A primeira diz respeito aos estádios de desenvolvimento, não sujeitos a variação; a segunda focaliza-se na variação a que estão sujeitas outras áreas da língua, socio-psicologicamente relevantes mas menos relacionadas com sub-sistemas linguísticos.

far from promoting acquisition can actually produce disturbances in the acquisition process» (Pienemann 1989:72) ¹⁹⁹.

Esta proposta, ainda que alguns dos seus aspectos tivessem sido alvo de críticas ²⁰⁰, tem tido bastante aceitação, particularmente por parte de investigadores ligados ao ensino, e tem dinamizado a investigação em contexto formal de aquisição. Recentemente, Nick Ellis, cujos trabalhos temos vindo a seguir, afirmava:

Just as there is the issue of Learnability in L1, so there is that of Teachability in L2 – any empirical findings about natural developmental sequences should be respected in the design of instructional materials (Pienemann, 1985), and attempts to teach structures or transformations that build on still-to-be acquired procedures or representations are likely to fail. (Ellis & Laporte 1997:78)

Como disse antes, a investigação tem sido guiada pela linguística teórica. Daí que tivessem merecido lugar de destaque a morfologia e a sintaxe. Em 1985, Clahsen justificava o foco do modelo na morfossintaxe dizendo que se sabe muito mais sobre esta do que sobre a aquisição da semântica ou da pragmática, mas acrescenta: logo que se saiba se o desenvolvimento da semântica e da pragmática também seguem uma sequência de aquisição, será necessário rever os procedimentos e incluir esses resultados na teoria (Clahsen 1985:288). Esta prova de abertura parece, no entanto, não ter convencido Peter Skehan que, em *A Cognitive Approach to Language Learning*, aponta como um dos aspectos fracos do Modelo a falta de atenção prestada aos aspectos lexicais. Ele afirma que Pienemann 1984 parece ignorar que «normal communication relies upon lexicalized chunks of language which are necessary for the speaker to keep up with real time, and for the listener to avoid having to deal with consistently new and difficult-to-process language.» (Skehan 1998:82).

Mas, de facto, embora, como vimos, os estudos se tenham vindo a centrar cada vez mais no léxico, e que, uma vez mais, guiados pela teoria

¹⁹⁹ Lightbown (1985:183) e Gass (1988:213) são mais restritivos: «I take the position of Lightbown (1985) that studies of SLA make a major contribution to the field of pedagogy in the form of 'teacher education' and less of a contribution in terms of actual practice.»

²⁰⁰ Para crítica ao *Multidimensional Model*, em especial à sua componente socio-psicológica, ver Hudson 1995. Ver também a resposta de Pienemann & Johnston & Meisel 1995. Referência a outras críticas podem encontrar-se em Skehan 1998: 80-82.

linguística, a partir do início dos anos 90, os investigadores tenham começado a lembrar sistematicamente que «the grammar/vocabulary dichotomy is invalid» (Lewis 1993/1996:vi), que «the largest part of language learning is the learning of words and their properties» (Little 1994 – em Singleton 1999:20), que «**grammars are systems of rules that operate on a lexical database**» (Ellis 1996), a investigação em aquisição, por enquanto, ainda não se aventurou suficientemente por estes caminhos e os resultados ainda não são, até mesmo para línguas muito estudadas, tão abundantes (e coerentes, e compatíveis...) que permitam estabelecer uma sequência de estádios de desenvolvimento que, por sua vez, permita apoiar o ensino ²⁰¹. No entanto, e pelo menos, o estado actual dos conhecimentos sugere uma pedagogia que concilie uma abordagem comunicativa com o ensino centrado no léxico.

Um trabalho pioneiro, no sentido em que procura uma solução de compromisso («some sort of middle ground»), é *Lexical Phrases and Language Teaching* de James Nattinger e Jeanette DeCarrico, e teve a sua 1.^a edição em 1992. Os autores, de acordo com Widdowson (1989:131), consideram que abordagens que se concentram essencialmente nas regras de gramáticas podem conduzir a uma falta de preocupação com a adequação, enquanto que abordagens excessivamente comunicativas podem conduzir a dificuldades para compor e decompor frases (Nattinger & DeCarrico 1992/1997:xiv). Referem que a investigação em psicolinguística mostra que, a todos os níveis (fonológico, morfológico, sintáctico, semântico e entoacional), a aquisição se processa em dois estádios – «a stage of ‘item learning’ and a stage of ‘system learning’» – e propõem um programa de ensino centrado em ‘*lexical phrases*’:

This approach, we believe, avoids the shortcomings of relying too heavily either on theories of linguistic competence, or on theories of communicative competence. Though the focus is on appropriate language use, the analysis of regular rules of syntax is not neglected (Nattinger & DeCarrico 1992/1997:xv).

²⁰¹ Uma tentativa é Altman 1997. O autor propõe, num estudo longitudinal de aquisição de vocabulário de Hebraico como L2, em que ele é o sujeito da investigação, cinco estádios de desenvolvimento na aquisição dos verbos.

Podemos distinguir duas abordagens que advogam este compromisso: a **abordagem lexical** e a **abordagem cognitiva**. A primeira, desenvolvida na sequência da investigação sobre *corpora*²⁰², enfatiza o facto de as palavras terem entre si uma certa tolerância de compatibilidade e em haver certas combinatórias que são independentes de considerações de ordem gramatical e que, por isso, necessitam de uma abordagem lexical para o ensino (ver Sinclair & Renouf 1988, Willis 1990, Lewis 1993); a segunda, além de também ter em consideração a investigação sobre *corpora*, procura integrar a investigação sobre os aspectos psicolinguísticos e a dicotomia entre ‘*item learning*’ e ‘*system learning*’ (ver Cowie 1987, Ellis & Beaton 1993, Ellis 1994, 1995, 1996, 1998, 1999, Skehan 1998).

4.4.1. *A abordagem lexical*

No final dos anos 80, os investigadores começam a chamar sistematicamente a atenção para que o vocabulário deve ter uma posição central no ensino, guiando-o em todas as áreas, incluindo a gramática (ver Carter & McCarthy (eds.) 1988). No entanto, a questão que se põe é como planificar o ensino ou produzir materiais didácticos tendo em consideração, ao mesmo tempo, a gramática e o léxico. John Sinclair e Antoinette Renouf propõem, como solução, o recurso a *corpora*, e comentam:

If the analysis of the words and phrases has been done correctly, then all the relevant grammar, etc. should appear in a proper proportion. Verb tenses, for example, which are often the main organizing feature of a course, are combinations of some of the commonest words in a language (Sinclair & Renouf 1988:155).

Uma selecção correcta deve basear-se na frequência das combinatórias. Daqui decorre que a abordagem lexical, tal como a encaram estes autores, não encoraja a aquisição de um vocabulário muito abundante, pelo menos nos níveis iniciais. Propõe sim uma concentração no vocabulário de que o aprendente dispõe num determinado estágio, e a sua recombinação (Sinclair & Renouf 1988:155).

Lewis (1993/1996) defende também que «*language consists of grammaticalised lexis, not lexicalised grammar*», que «*much language consists of multi-word ‘chunks’*», e que, conseqüentemente, o ensino se

²⁰² Esta abordagem está directamente relacionada com o projecto COBUILD.

deve organizar em função das palavras (Lewis 1993/1996:vi). Não considera esta abordagem como uma revolução, mas sim como uma útil mudança de ênfase (Lewis 1993/1996:193). No entanto, tem uma posição diferente em relação à frequência. Argumentando que as palavras menos frequentes são geralmente aquelas que transportam mais sentido (Lewis 1993/1996:109) defende que «learnability and communicative power are at least as important in electing words for inclusion as frequency» (Lewis 1993/1996:110). Por isso, propõe que todos os níveis de iniciação devem fornecer aos estudantes um vocabulário vasto, «even if they are initially unable to grammaticalise it» (Lewis 1993/1996:106). Insiste em que os professores se devem concentrar mais no léxico do que em palavras isoladas, ou seja, as palavras de conteúdo ('content words') devem ser sempre ensinadas em contexto, com os verbos e adjetivos com que co-ocorrem, do mesmo modo que aos itens mais deslexicalizados, como são os verbos de alta frequência e as preposições, deve ser dado um tratamento lexical e não gramatical ²⁰³.

4.4.2. A abordagem cognitiva

A abordagem cognitiva, como vimos em 3.4, não só tem em consideração a investigação recente sobre *corpora* como, em muito maior escala do que a abordagem lexical, integra nas suas propostas, para explicar os processos que regem a aquisição de L2 e nas linhas mestras que devem guiar o ensino, os avanços feitos pela linguística cognitiva, pela sociolinguística, pela psicolinguística e pela neurolinguística ²⁰⁴.

Contrariamente à linguística generativa, destaca a importância da experiência ('*the power law of practice*') nos processos de reconhecimento e de produção lexical:

Fluent language users have had tens of thousands of hours on task. They have processed many millions of utterances involving tens of thousands of types presented as innumerable tokens. The evidence of

²⁰³ Em Lewis 1997 (*Implementing the Lexical Approach*), o autor apresenta sugestões para a construção de exercícios que ajudem os estudantes a focar a sua atenção nos vários tipos de combinatórias.

²⁰⁴ Um exemplo é Rod Ellis 1994. Ver nota 141.

language has ground on their perceptuo-motor and cognitive apparatus to result in complex language competencies (N Ellis 1998:641) ²⁰⁵.

Porque aprender uma língua consiste em analisar e aprender sequências de sons em palavras e de palavras em frases, e porque estas sequências formam uma base de dados a partir da qual o sujeito abstrai uma gramática, o falante pode mobilizar o conhecimento implícito (um sistema baseado em regras) ou o conhecimento explícito (um sistema baseado em elementos soltos) (Skehan 1998). Porque é a prática que desenvolve a competência linguística, **o ensino, ao proporcionar oportunidades de praticar a língua, terá um papel relevante se for adequado ao aprendente e ao seu estágio de desenvolvimento.** «Here, of course, lie the cruxes of applied linguistics» (N Ellis 1999: 30).

4.5. *Sumário*

Topics, notions, functions and grammar, therefore, all have a useful part to play in the vocabulary syllabus and it seems misguided to focus too much on one to the possible exclusion of the others (O'Dell 1997:272).

Cada movimento retoma sempre qualquer contribuição que, num passado mais ou menos remoto, tinha passado despercebida ou tinha sido abandonada. Por outro lado, é sempre uma reacção contra aquilo que vinha sendo feito no passado próximo ou no presente. É vantajoso, portanto, acreditar que nada é definitivo e cada movimento acrescenta sempre algum aspecto positivo àquele que o precede.

A abordagem comunicativa, ao preconizar o ensino do léxico em situação, constituiu, de certo modo, um passo em frente em relação às propostas anteriores. No entanto, porque se preocupa sobretudo com os aspectos funcionais, ela continua a não propor soluções quanto a uma possível sequencialização do material linguístico que tenha em conta os dados da aquisição. Assim sendo, ela **não contribui explicitamente para estabelecer uma rede de conexões entre os diferentes itens, e,**

²⁰⁵ A propósito de alguns princípios que têm guiado a investigação no âmbito da GG, nomeadamente a pobreza do estímulo, Ellis afirma: «Note that it is the assumptions of UG that are under attack, not the generative grammar descriptions of the relations between the linguistic units. These hold: To the extent to which they are systematic observations, they are as valid as the universals derived from typological classification research...» (Ellis 1998:633).

consequentemente, torna mais difícil o desenvolvimento de uma gramática. Nick Ellis (1994:3) afirma que foi exactamente o facto de os estudantes que aprenderam com métodos sem recurso à gramática terem ficado tão mal preparados que chamou a atenção para a importância do ensino e para a necessidade de repensar as opções metodológicas.

A abordagem lexical apoiada em dados de *corpora* não encoraja a aquisição de um vocabulário muito abundante, pelo menos nos níveis iniciais. Propõe sim uma concentração no vocabulário de que o aprendente dispõe num determinado estágio, e a sua recombinação. Embora se considere que deve existir um equilíbrio entre gramática e léxico, Sinclair e Renouf 1988 acreditam que uma selecção correcta de palavras e das suas combinatórias mais frequentes, feita com base em *corpora*, pode cobrir a gramática relevante; isto é particularmente válido para os verbos mais frequentes e mais deslexicalizados e para os auxiliares. Ou seja, uma abordagem preferencialmente lexical destes verbos parece mais adequada do que uma abordagem tendencialmente gramatical. E digo 'tendencialmente' porque, tal como advertem Sinclair e Renouf (1988:155), outras línguas podem ser diferentes do Inglês. Esta observação faz particularmente sentido no caso do português, língua com uma morfologia muito mais rica e a que, portanto, é preciso prestar muito mais atenção. Esta é também, como vimos, a proposta de Viberg 1993 para os verbos básicos.

Mas, de facto, aquilo que não podemos esquecer é que todos os contributos que as diferentes áreas do saber podem oferecer ao ensino de L2 só serão válidos se devidamente enquadrados por estudos de aquisição que possam conduzir ao estabelecimento de estádios de aquisição.

Parte II

Léxico aquisição e ensino de PE L2

Introdução

Como disse nas primeiras páginas deste trabalho, foram várias as motivações para empreender esta investigação centrada na aquisição do léxico. Mas, na escolha, também pesou uma grande “curiosidade”: com que palavras é que falam/escrevem os estrangeiros? De todas aquelas a que estão expostos no dia a dia e na sala de aula, quais são as que, de facto, usam? Ou seja, quais e como as aprendem? Todas as mais frequentes? Quais são mais fáceis de aprender? Será que os falantes nativos usam muito mais para dizer as mesmas coisas?

O caminho escolhido não foi o de comparar o vocabulário/léxico de indivíduos, como é costume, mas o de grupos. Esta opção justifica-se porque, até ao momento, nenhum teste individual de avaliação da dimensão do vocabulário parece satisfatório ²⁰⁶. Além disso, usar qualquer teste implicaria testá-lo e validá-lo previamente, já que, como é sabido, não existem testes desse tipo para o português. Por outro lado, essa metodologia de investigação permite obter informação quanto à dimensão do vocabulário mas não quanto à organização do léxico.

Excluída a hipótese de utilizar testes individuais de avaliação da dimensão do vocabulário, o procedimento que parecia mais adequado seria o estudo de um *corpus* constituído por um número de documentos orais ou escritos suficientemente representativo de um grupo com determinado perfil.

Em geral, espera-se de um trabalho de investigação que circunscreva uma questão como objecto de análise e adopte um quadro teórico

²⁰⁶ «It is not obvious that any particular form of test is either specifically or adequately suited for testing either R [receptive] or P [productive]. Various tests (checklist, multiple-choice, translation, illustration, etc.) have been used indiscriminately for measuring R and P. [...] Productively, the size of a vocabulary could ideally be evaluated through a spontaneous production test, but here again the tested field would be very restricted: only items produced by the subject would be counted. The computation would partially depend on extra-linguistic and/or psychological factors (verbosity, fluency of the subject, motivation toward the topic, ease in front of an audience, etc.) and, thus, be either too low or too high.» (Melka 1997:97-98)

perfeitamente definido. Espera-se ainda que os resultados da investigação contribuam para o avanço do conhecimento na área em que ele se inscreve. Esta dissertação não segue rigorosamente esse figurino. O que pretendo é, antes de mais, criar um enquadramento que me permita, no futuro, continuar a investigação sobre aquisição do léxico, identificando então questões mais finas e procurando tratá-las de modo mais exaustivo. Considero que desse enquadramento para estudos de léxico mais direccionados, especialmente se se trata dos primeiros estádios de aquisição e desenvolvimento de uma L2, faz parte uma avaliação global da dimensão e características do vocabulário disponível. Porque, apesar de todas as objecções que se possam levantar, desistir de avaliar vocabulário receptivo e/ou produtivo ²⁰⁷, até serem encontrados melhores métodos para medir o conhecimento das palavras ²⁰⁸, não me parece ser o caminho a seguir. Avaliar, separadamente para a escrita e para o oral, a dimensão do vocabulário, a partir de um *corpus*, combinando-a com a avaliação do conhecimento de cada uma das palavras que o constituem, ainda que no estado actual dos conhecimentos não se saiba muito bem como fazê-lo, parece-me ser, de momento, a melhor opção. Assim, como disse na Introdução, **o trabalho que aqui se apresenta tem como objectivo central avaliar que vocabulário usa e que conhecimento manifesta sobre ele um grupo de aprendentes de português língua não materna, depois de um mínimo de 120 horas de aprendizagem semi-formal, durante um período de mais ou menos 3 meses.**

Vários investigadores recomendam que nos mantenhamos atentos a toda a colaboração que diferentes abordagens do conhecimento e diferentes teorias põem à nossa disposição. Por outro lado, quanto mais estudos são levados a cabo, mais difícil é optar por uma teoria; por isso, a atitude

²⁰⁷ «The abundance of expressions used to describe the two notions (active vocabulary vs. passive, comprehension vs. production, understanding vs. speaking, recognitional vocabulary vs. actual or possible use) may suggest either that researchers do not agree on the reality of what they are working on, or that a distinction between R and P does not necessarily exist as such, but is justified for reasons of convenience. We are in the habit of setting off one term against the other. Studies have sometimes based themselves on this convenient dichotomy (R vs. P) as if it was self-evident, and further seem to assume, realistically or not, that one possesses two distinct vocabularies, one receptive and one productive.» (Melka 1997:84).

²⁰⁸ «[W]hat we need to do is look for better ways of measuring knowledge of words, and drop the idea of testing and measuring R and P until we have found them.» (Melka 1997:98)

recomendável é encontrar o ponto de vista mais adequado para a interpretação de um determinado problema em análise (cf. 0.4. *Enquadramento teórico*; Bialystok 1994:567). Em conformidade, depois de se ter indicado a natureza e objectivos do estudo, assumiu-se que o enquadramento teórico adequado devia ter em consideração várias áreas da Ciência Cognitiva enquanto ciência interdisciplinar e, na *Parte I*, reviu-se literatura das disciplinas consideradas relevantes. Essa revisão serviu para chamar a atenção para a importância que debates noutras disciplinas têm tido para os estudos de L2, nomeadamente, em dois pontos fulcrais nesta área e que aqui nos interessam particularmente: **o papel da L1 e o papel do ensino. Em que consiste o *transfer*? Como, quando, em que casos, em que circunstâncias o ensino pode influenciar a aquisição?**

Os aprendentes partem, para a L2, de estádios de aquisição diferentes. Esses estádios de aquisição estão relacionados com a relação entre a sua L1 e a L2 e com o modo como determinados parâmetros são fixados em cada uma delas, como defendem os generativistas. Corder propõe que o papel desempenhado pela L1 é bastante mais penetrante e subtil do que tradicionalmente se tem acreditado; segundo este autor, ***transfer*** é a transferência de conhecimento implícito da estrutura mental da L1 para a interlíngua, constituindo esta um conhecimento separado e com desenvolvimento independente da L1. *Transfer* é, assim, um mecanismo de facilitação que usa por empréstimo itens e traços da L1 como uma estratégia comunicativa e que, quando é bem sucedida, conduz à sua incorporação na interlíngua. Os tipologistas procuram relacionar os padrões de *transfer* com o binómio L1/L2. A pergunta que se põem e que desejam ver respondida pela investigação é: em que medida é que a noção de marcação pode prever os problemas de desenvolvimento e de aprendizagem que um determinado contexto de L1/L2 põe? Haverá paralelismo entre hierarquias de marcação e sequências de aquisição? E estabelecem como hipótese que elementos e estruturas interlinguisticamente não marcadas surgirão em fases iniciais de aquisição, independentemente das condições de marcação da L1 e da L2 (o que, de certo modo, corresponde à hipótese nula de Slobin). Põem ainda como hipótese que as condições de marcação de L1/L2 desempenharão um papel importante nas fases seguintes de

desenvolvimento: os padrões de *transfer* estarão dependentes da marcação em L1 e em L2 e das relações entre as duas línguas (Hyltenstam 1986).

Pienemann e os que com ele trabalham dizem-nos que o **ensino** pode acelerar a aquisição, mas não pode alterar os estádios de desenvolvimento. E, de acordo com a *teachability hypothesis*, o ensino só pode favorecer a aquisição se a interlíngua se encontrar próxima do ponto em que a estrutura ensinada seria adquirida em situação de aprendizagem informal (Pienemann 1989:60). Os cognitivistas partilham desta opinião (cf. Ellis & Laporte 1997). Paradis encontra noutras disciplinas explicações para esta hipótese. O conhecimento implícito/procedimental, resultante de processos adquiridos e automatizados aquando da aquisição da L1, e o conhecimento explícito/declarativo/metalinguístico, que o ensino disponibiliza, dependem de sistemas cerebrais diferentes e têm conteúdos diferentes, e, por isso, este último não pode ser transformado ou convertido graças a automatização. A prática que a sala de aula proporciona automatiza a capacidade para produzir sequências correctas. É a prática, e não o conhecimento metalinguístico, que, ao desenvolver a automatização, desenvolve por via indirecta a competência linguística (Paradis 1994). **Poderá então pôr-se como hipótese que, nos casos em que a L1 e a L2 não coincidam, para modificar estruturas conceptuais associadas à L1, serão necessárias muitas rotinas que conduzam à automatização de sequências correctas.** (Ou, como diriam os generativistas, nos casos em que a L1 e a L2 difiram no modo como um determinado parâmetro foi fixado).

Considerando como aplicáveis aos informantes do estudo (cujo contexto de aprendizagem é semi-formal) os processos de aquisição descritos na literatura sobre bilinguismo, considerando que (1) o léxico é central no processamento, na representação e na aquisição; que (2) qualquer falante tem uma representação rica do léxico o que significa que a correcta e completa introdução de uma palavra no léxico mental exige o seu estabelecimento numa série de sistemas de representação; que (3) o léxico mental de cada falante bilingue (ou multilingue) é diferente para a compreensão e para a produção (e diferente também para a escrita e para o oral); que (4) a aquisição de diferentes propriedades do léxico resulta de processos de aprendizagem diferentes; que (5) factores interlexicais, tais como a relação tipológica entre a L1 e a L2 e a dimensão e características do léxico mental do falante de L2, afectam a aquisição de certos aspectos de uma palavra; e que também (6) factores intralexicais e a interacção entre eles afectam a

aquisição de uma palavra ²⁰⁹, como é que os processos gerais enunciados acima se espelham na aquisição, na “aprendibilidade” (cf. **Capítulo 3; Laufer 1990 e 1997**), de cada um dos itens que constituem o léxico mental?

Paradis afirma que todos os aspectos da aquisição e uso da morfossintaxe e da fonologia são compatíveis com as características da memória procedimental, enquanto que pelo menos alguns aspectos do léxico caem no domínio da memória declarativa (cf. Capítulo 3; Paradis 1994). Nick Ellis di-lo de outro modo: no que respeita a aquisição do significado, trata-se de um processo de aprendizagem explícita; no que diz respeito aos aspectos formais e a regularidades, eles dependem de processos de aprendizagem implícita; diz ainda que a compreensão e produção de vocabulário é separável dos aspectos semânticos e conceptuais (cf. Capítulo 3; Ellis 1994b).

Ou seja, a disponibilidade vocabular poderá depender de características do sujeito, nomeadamente da sua memória, e da frequência de um determinado item no input; o ensino poderá acelerar a aquisição de certos aspectos formais chamando a atenção para eles; propriedades semânticas do léxico, e suas consequências na gramática, porque estão mais dependentes de relações interlexicais e de conhecimento implícito, reflectirão mais diferentes estádios de aquisição e resistirão certamente mais à intervenção do ensino. Diz-nos Paul Meara:

There is obviously a great deal of internal structure in the mental lexicon of most L1 speakers (cf. Aitchison 1987), but it is much less obvious that the same degree of organization exists in a typical L2 lexicon. My guess is that a 5,000 word lexicon consisting of 5,000 separate items, is unlikely to behave in the same way as a 5,000 word lexicon which is structured. (Meara 1993: 293-4)

A afirmação acima, de Paul Meara, sintetiza a questão da relação entre dimensão do vocabulário disponível e o grau de organização resultante do número e do tipo de conexões entre os itens que caracterizam o léxico mental de um falante de L2. Ou seja, é possível que dois falantes disponham de um léxico mental que se baseia num vocabulário composto por um número de palavras muito semelhante; mas que, no entanto, o grau e o modo de estruturação mental de cada um desses léxicos seja muito diferente, e que, conseqüentemente, esses dois falantes apresentem *performances* linguísticas muito distintas. Isto pode acontecer se compararmos dois falantes de L1, ou um falante de L1 com um aprendente de L2, mas também pode acontecer se compararmos entre si dois aprendentes de L2.

²⁰⁹ Cf. *hipótese dos efeitos específicos de cada língua*: factores relacionados com a acessibilidade das formas linguísticas em cada língua desempenham um papel em cada caso (ver Capítulo 3).

Isto conduz-nos, entre outras coisas, a operar com uma distinção fictícia entre **vocabulário** e **léxico**. Digo fictícia porque, como se pode inferir de tudo o que foi dito até aqui, o vocabulário não é uma entidade diferente do léxico mental, mas sim uma parte dele. Assim, o que me proponho é recolher e analisar dados que permitam, por um lado, avaliar a dimensão do vocabulário e, por outro, avaliar o grau de estruturação do léxico. Isto significa assumir como orientadoras da investigação as seguintes **hipóteses gerais**:

- (1) dois indivíduos (ou dois grupos) podem dispor de léxicos mentais semelhantes no que respeita o número de itens, mas
- (2) cada um desses léxicos mentais pode ter um grau de organização e, conseqüentemente, comportamentos muito diferentes entre si.

Mas, restringir a investigação à confirmação destas hipóteses seria como pôr semente em vaso e concluir que germina, como diria o próprio Meara. Tendo em consideração que o objectivo final da investigação em aquisição de L2 consiste em explicar e prever os estádios por que passa o aprendente e em identificar os processos graças aos quais adquire a língua alvo (cf. Capítulo 4; Nunan 1991:253), de modo a, em última análise, poder informar o ensino, parece desejável conduzir a investigação de modo a prever quais as componentes do léxico sobre as quais a intervenção do ensino se poderá fazer sentir mais directamente. Assim, o que me proponho é ir um pouco mais longe, analisando os dados de modo a avaliar o papel da relação tipológica entre L1 e L2 e o papel do ensino na aquisição das diferentes componentes de um item lexical. Para isso, procurarei algumas respostas para as questões enunciadas acima: em que consiste o *transfer*? Como, quando, em que casos, em que circunstâncias o ensino pode influenciar a aquisição? Acrescento, então, outras hipóteses que guiarão a análise dos dados:

- (1) a correcta introdução de um item lexical no léxico mental depende de um conjunto de factores: do grau de organização do léxico mental, da relação tipológica entre L1 e L2, da sua frequência, categoria gramatical e de outras características específicas;
- (2) factores intralexicais combinados potenciam a dificuldade de aquisição;
- (3) a aquisição e uso de certas componentes do léxico estão mais relacionadas com a relação tipológica entre L1 e L2 do que outras; ou seja, a relação tipológica entre duas línguas far-se-á

sentir de modo diferente na aquisição de diferentes componentes do léxico;

(4) as componentes do léxico que caem no domínio da memória declarativa, tais como disponibilidade vocabular e certas componentes formais dos itens lexicais, dependem mais de outras características do sujeito do que de sistemas interiorizados e relacionados com a relação tipológica entre a L1 e a L2 do que as propriedades semânticas do léxico;

(5) certas componentes do léxico são mais sensíveis à intervenção do ensino do que outras; assim, se a aquisição de uma componente de um item lexical puder depender mais da aprendizagem explícita, esse processo reflectirá em menor grau factores interlexicais (a relação tipológica) e intralexicais (hierarquias de marcação/características do português) e será mais sensível à intervenção do ensino.

A *Parte II* organiza-se em três capítulos. No capítulo 5, darei conta da metodologia seguida para a recolha e tratamento de dados; o capítulo 6 organiza-se em duas secções: na primeira, em *Vocabulário disponível*, avalia-se a dimensão do vocabulário de que dispõe cada grupo, na segunda, em *Léxico deficitário*, avalia-se o seu grau de organização, analisando e comentando os vários tipos de desvios e as estratégias usadas pelos aprendentes para compensar o vocabulário indisponível; finalmente, no capítulo 7, reflecte-se sobre as implicações que os resultados do estudo poderão ter no ensino de PE L2.

Capítulo 5

Recolha e tratamento de dados de PE L2

5.1. *Introdução*

Uma vez tomada a decisão de constituir um *corpus* para estudo, levanta-se uma série de questões a que é preciso responder: recolher um *corpus* escrito ou oral? Qual a dimensão adequada de uma amostra? Com que materiais de L1 confrontar os dados? Que ocorrências considerar como desviantes? Como anotar os documentos? Que etiquetas usar? Que programa informático usar?

No espaço deste capítulo, darei conta das respostas encontradas para estas questões. Referirei as condicionantes que me levaram a optar por um *corpus* escrito, e os critérios para a selecção, transcrição, organização e codificação dos documentos. Referirei também outros materiais e informantes envolvidos no estudo, nomeadamente, o grupo de controle de falantes de português L1, e ainda: o *Português Fundamental (PF)*, que foi usado, a par do *sub-corpus* de controle, para a avaliação do vocabulário disponível, e o *Projecto do Português Oral de Maputo (PPOM)* (Gonçalves 1997), cuja grelha de desvios serviu como ponto de partida para o estabelecimento da tipologia de desvios adoptada neste trabalho.

5.2. *Informantes e materiais*

Para a obtenção de materiais a partir dos quais fosse possível chegar à constituição de um ***corpus***, os informantes disponíveis eram os estudantes do Departamento de Língua e Cultura Portuguesa (DLCP) da Faculdade de Letras da Universidade de Lisboa (FLUL). Estes, na maior parte dos casos, só ficam entre nós um ou dois semestres lectivos. Muitos deles, sobretudo os oriundos de países europeus, têm, em Lisboa, um

número razoável de amigos, conhecidos e, nalguns casos, familiares, o que não os torna muito dependentes da escola, resumindo-se a sua presença aí ao tempo das aulas. Constituem excepção a este comportamento os estudantes chineses, os únicos que permanecem na escola para além dos tempos lectivos ²¹⁰.

Há uns seis ou sete anos, fiz uma experiência de recolha de discurso oral e tive oportunidade de constatar o que já sabia em teoria: como é consumidor de tempo e fonte de contrariedades uma recolha desta natureza. Isto para dizer que seria muito difícil optar por este tipo de material e mais ainda se se tratasse de material oral e longitudinal. Outros investigadores, trabalhando até em equipa, queixam-se das dificuldades e dos insucessos que este tipo de recolha acarreta e de como o número de sujeitos da amostra se vai reduzindo à medida que o tempo passa (veja-se, por exemplo, Singleton 1999:194 e The Trinity College Dublin Modern Language Research). É com certeza por isso que materiais com estas características têm quase sempre como informantes imigrantes que vivem na cidade em que a recolha é feita e que, em muitos casos, são pagos.

Por outro lado, tinha à minha disposição uma boa quantidade de materiais que vinha arquivando desde 1988. Trata-se dos exames escritos do Curso Básico, realizados no mesmo Departamento. Estes exames são constituídos por duas partes: o reconto, por escrito, de uma história que é apresentada oralmente (em Leiria 1991, usei uma selecção desta primeira parte dos exames disponíveis até esse data) e um segundo exercício que consiste numa composição que pode ser uma carta, um texto opinativo, narrativo ou até a redacção de uma receita de cozinha. Em cada enunciado da prova são sempre sugeridos ao examinando vários temas possíveis de entre os quais ele deve escolher um ²¹¹. Foi o material produzido como resposta a este segundo estímulo que veio a ser seleccionado para constituir o *corpus* que serviu de base a esta investigação.

Depois de uma observação do material em arquivo, verifiquei que, além de dispor de um pequeno número de documentos de falantes de um considerável número de línguas maternas, de outras dispunha de um

²¹⁰ Para mais detalhes sobre informantes e materiais também válidos para este trabalho, ver Leiria 1991:83-93.

²¹¹ Enquanto coordenadoras do Curso Básico, na altura em que a recolha deste material foi feita, a elaboração das provas de exame estava a meu cargo e de Helena Bárbara Marques Dias. A primeira parte do exame, a história para ser recontada, foi, em quase todos os casos, resultado da imaginação criadora da minha colega; a segunda parte, de que resultaram estes materiais, apoiava-se tematicamente na primeira e foi elaborada por mim. Agradeço à minha colega Helena Bárbara esta preciosa colaboração.

conjunto apreciável de provas de exame que me permitiria constituir uma amostra de, pelo menos, 50 documentos com as características desejadas. Ou seja, dispunha de mais de 50 textos de falantes de alemão, chinês, espanhol, francês e sueco. Eram, aproximadamente, 180 provas de chineses, 85 de alemães, 70 de suecos, 65 de espanhóis e igual número de franceses. Quanto ao número de palavras por composição, 23 de chineses, 6 de espanhóis e 2 de franceses tinham menos de 100 palavras; tinham mais de 400 palavras 2 composições de alemães e 2 de chineses, 3 de franceses, e 1 de espanhóis.

Poderia assim, a partir destes materiais, constituir e organizar um *corpus* de aproximadamente 50 documentos por L1, em que estariam representadas duas línguas românicas (espanhol e francês), duas línguas germânicas (alemão e sueco) e uma língua não indo-europeia (chinês) ²¹². A questão que se punha era se a amostra com estas dimensões seria suficientemente representativa para os fins em vista.

A **dimensão adequada de uma amostra** depende daquilo que se pretende investigar. Contrariamente a componentes da língua mais sistémicas, como a fonologia ou a morfologia, estudos de léxico exigem uma amostra relativamente ampla. Por isso, *The Bank of English*, que em 1997 tinha 300 milhões de palavras, ou o *Corpus de Referência do Português Contemporâneo* (CRPC), um pouco mais modesto, com aproximadamente 50 milhões ²¹³, são considerados como particularmente adequados para estudos desta natureza. Mas também têm sido recolhidos alguns *corpora* para investigação em L2. O *Longman Corpus of Learners' English (LCLE)*, por exemplo, totaliza aproximadamente 10 milhões de palavras de Inglês escrito. Inclui respostas a testes de exame, cartas, relatórios, diários e

²¹² Por razões que se prendem com o sistema de escrita e com as consequências que esse facto pode ter na representação mental de alguns aspectos do léxico, lamento imenso não ter podido incluir um grupo de falantes de árabe e de japonês, mas o número de documentos disponível era notoriamente insuficiente. Alguns dos documentos do *sub-corpus* 'árabes' de Leiria 1991 (ver vol. II *corpus*, 61-75) deixam antever o interesse que o estudo desses materiais poderia ter. Ver Ryan & Meara 1991, referido em nota a 2.2.2.3. *O léxico da L2*.

²¹³ Para informação sobre *corpora* em geral, ver, por exemplo, Kennedy 1998, e McEnery & Wilson 1996. Quanto ao Corpus de Referência do Português Contemporâneo (CRPC) «é um *corpus* que se pretende que venha a ser representativo do português de uso geral e corrente no séc. XX, sendo, actualmente, constituído por amostragens de língua falada e de língua escrita do português europeu, do português do Brasil, do português de países africanos de língua oficial portuguesa e do português de Macau.» (Nascimento & Gonçalves 1995:143). O CRPC integrou, para a língua falada, o *corpus* oral do PF.

composições de aprendentes de Inglês, de oito níveis de proficiência diferentes, de 160 línguas maternas (Kennedy 1998:42). Perante estes números, as menos de 70 000 palavras que os materiais disponíveis para o *corpus* que me propunha organizar totalizariam pareciam ser insuficientes. Mas quando comparado com o *LCLE*, se tivermos em conta as 160 línguas e os oito níveis diferentes, a desproporção atenua-se consideravelmente. Feitas as contas, por nível e por língua, o resultado é de 7812 palavras, menos do que cada um dos *sub-corpus* da amostra em estudo aqui. Parecia portanto possível constituir, a partir dos materiais disponíveis, uma amostra suficientemente representativa.

Os sujeitos que produziram os documentos seleccionados, todos estudantes do Curso Básico de Português para Estrangeiros do DLCP da Faculdade de Letras da Universidade de Lisboa, frequentaram nesta escola pelo menos um semestre lectivo (120 horas lectivas de língua, durante um período de mais ou menos 3 meses); 64 % são do sexo feminino e 36 % do sexo masculino, com idades compreendidas entre os 18 e os 57 anos, e são caracterizados por uma grande heterogeneidade no que respeita formação académica e profissional e interesses pessoais. Têm como língua materna o espanhol, o francês, o alemão, o sueco ou o chinês (mandarim ou cantonês). Na maior parte dos casos, conhecem pelo menos mais uma língua além da sua L1 e do português. Por se encontrarem a viver num espaço em que a língua em aprendizagem é a língua da comunidade, e por frequentarem aulas de língua, estão integrados num contexto de aprendizagem que costuma designar-se de **contexto de aprendizagem semi-formal**. Alguns deles, à data de chegada a estes cursos, já tinham tido contacto formal ou informal com o português. O que há de comum entre todos eles é o facto de, depois de um mínimo 120 horas lectivas, se terem apresentado ao exame do Curso Básico.

Dediquei muitos dias de trabalho a tentar estabelecer um conjunto de critérios que me permitisse, a partir de dados disponibilizados pelos informantes nas fichas de inscrição no Curso, construir uma grelha de equivalências que relacionasse tempo de contacto com a L2 (em aprendizagem formal exclusiva, em aprendizagem informal exclusiva, aprendizagem formal sem recurso a professor, etc...), conhecimento de outras línguas (maior ou menor proximidade do Português e o grau de

conhecimentos dessas línguas) e, assim, conseguir alguma homogeneidade quanto a **tempo de estudo** e **conhecimentos prévios** à data da chegada ao Curso. Concluí que estas variáveis são absolutamente incontroláveis. Já tinha concluído o mesmo aquando da realização de Leiria 1991 (ver vol. I:88-92). Uma das razões, entre muitas outras, que torna impossível o apuramento deste critério resulta da informação insuficiente que os estudantes são capazes de fornecer. Quando interrogados sobre o número de horas lectivas que frequentaram, dizem não se lembrar, dão respostas vagas, ou dizem não se recordar de qual a duração de cada tempo lectivo. Mas o que torna menos fiável este critério é a frequência com que a L1 foi usada na sala de aula. Outros investigadores têm-se posto a questão, até mesmo no que respeita a crianças. Por exemplo, *How much english vocabulary do swedish-speaking primary-school pupils know before starting to learn English at school?* é o título de um artigo de Palmberg (1985) em que se põe essa questão.

Por outro lado, e como disse também em Leiria 1991, seria impossível conseguir um conjunto de informantes que fosse em número suficiente, que só conhecesse a sua L1 e que nunca tivesse tido contacto com o português até à data da entrada no Curso Básico da FLUL. Além disso, e por isso mesmo, um estudo com essa população não me parece que fosse particularmente interessante porque não seria representativo do perfil do estudante de português. Para muitos estudantes, o português é a terceira ou mesmo a quarta língua que aprendem ²¹⁴.

De qualquer modo, não foram seleccionados documentos produzidos por informantes que, no momento da inscrição no curso, tivessem declarado saber bem ou relativamente bem qualquer língua românica como L2 (por exemplo, alemães que declararam saber bem espanhol, francês ou italiano; ou franceses que declararam saber bem espanhol ou italiano). Numa escala de 0 a 4, não foi seleccionado nenhum documento produzido por um informante que declarasse grau de conhecimento 3 ou 4 numa língua românica. Tendo em conta que os documentos foram produzidos, pelo menos, três meses depois da sua chegada a Portugal, assumi que o

²¹⁴ A este propósito, é interessante referir que começam a surgir vários congressos sobre aquisição de L3 e L4.

grau de activação de outras línguas românicas, em que, à chegada aqui, não eram muito proficientes, já seria bastante reduzido.

Assume-se, portanto, que **uma parte dos informantes, à data do início do curso na FLUL, poderia ter tido contacto com outra ou outras línguas românicas e tinha alguns conhecimentos de pelo menos uma outra língua, além da sua L1, e, em muitos casos, tinha conhecimentos de português adquiridos em situação formal ou informal de aprendizagem.** A maior parte dos **espanhóis**, à data do início do curso na FLUL, nunca tinha estudado português e sabia inglês; alguns sabiam também alemão ou francês. Muitos dos **franceses** estudaram entre 10 e 60 horas de português em França; muitos sabiam inglês e alguns um pouco de alemão, espanhol ou italiano. Muitos dos **alemães** estudaram entre 30 e 60 horas de português; quase todos sabiam inglês e muitos sabiam um pouco de francês. O perfil dos **suecos** é semelhante; além disso, à data do início do curso, vários já viviam em Lisboa há aproximadamente três meses. Todos os **chineses** declararam saber inglês; a maior parte já estudava português pelo menos há dois anos (em Macau ou na República Popular da China); e alguns já estudavam há um ou mesmo dois semestres, em Lisboa ou em Coimbra.

*Não sendo possível, como se viu antes, avaliar os conhecimentos de um informante pelo tempo de estudo, considerou-se como informação mais fiável para estabelecer o perfil linguístico do informante a outra parte do exame, ou seja, a prova de **compreensão oral**. Como disse antes, tratava-se do reconto de uma história. A história era lida três vezes por uma das coordenadoras do Curso. O exercício consistia em recontar, por escrito, a história com o máximo de pormenores possível*²¹⁵.

Com vista a estabelecer o perfil linguístico dos informantes para o presente trabalho, o desempenho de cada informante nessa prova foi avaliado numa escala de 0 a 4:

- 0** Parece não ter compreendido nada; deixou em branco o espaço na folha de prova reservado à resposta.
- 1** Compreendeu muito pouco; apenas escreveu algumas frases soltas com muitos mal entendidos.
- 2** Compreendeu apenas uma parte da informação; o texto produzido tem alguma coesão, mas revela que só uma parte da informação foi compreendida.

²¹⁵ Por se tratar de texto narrativo, esta primeira parte (das provas realizadas até 1989) foi seleccionada para constituir o *corpus* que serviu para o estudo da aquisição dos aspectos verbais expressos pelos Pretéritos Perfeito e Imperfeito (Leiria 1991). Para mais detalhes sobre esta prova, ver Leiria 1991.

- 3 Compreendeu uma boa parte; há algumas falhas que se manifestam em pequenos desvios em detalhes da história.
- 4 Compreendeu tudo; o texto produzido dá conta de toda a informação veiculada; nalguns casos, e já que o texto foi lido três vezes, pode até parecer um exercício de ditado.

Assim, o perfil linguístico dos informantes seleccionados, no que respeita a **compreensão do oral**, é este:

L1	Escala					Total de doc.
	0	1	2	3	4	
<i>Espanhol</i>	0	0	0	3	47	50
<i>Francês</i>	0	0	1	9	43	53
<i>Alemão</i>	0	0	10	22	21	53
<i>Sueco</i>	0	3	6	19	24	52
<i>Chinês</i>	3	17	19	8	2	49

Da leitura do quadro, ressalta que, independentemente de todas as variáveis não controláveis referidas acima, o nível de compreensão oral diminui à medida que a distância da L1 aumenta em relação ao português. Só a informantes chineses (três: o 23, o 45 e o 46) foi atribuído o grau 0 (zero) da escala, porque deixaram em branco o espaço na folha de prova reservado a esta tarefa que tinha como objectivo prioritário testar a compreensão oral; e só a dois informantes do mesmo grupo foi atribuído o grau 4 da mesma escala ²¹⁶. Como era esperável, são os espanhóis, que na quase totalidade chegaram ao Curso sem nunca antes terem estudado português, os que apresentam um mais alto nível de compreensão oral e são os chineses, que em muitos casos tinham já dois anos de estudos, os que revelam mais dificuldades.

5.3. Português Europeu Escrito por **ESTR**angeiros (**PESTRA**)

Sendo um dos objectivo do estudo avaliar de que vocabulário dispunham, para uso na escrita, aprendentes estrangeiros em contexto de aprendizagem semi-formal, era indispensável que os **temas** tratados fossem bastante diversificados. O Português Fundamental (PF) poderia fornecer-me orientação para a selecção, do ponto de vista temático, dos documentos. Fiz um levantamento dos temas abordados nas composições e verifiquei que cobriam *grosso modo* «os temas versados nas 1400 entrevistas do PF» (ver PF, tomo 1:26). Digo ‘grosso modo’ porque, de facto, certos temas, como é o caso de agricultura, fauna e flora nunca foram abordados, provavelmente por serem de pouco interesse para estudantes a viverem em espaço urbano e, por isso, considerados como não prioritários pelos professores e, em conformidade, pelos coordenadores que elaboraram os exames.

²¹⁶ O último campo do código (ver 5.3.3. *Codificação*) pelo qual é introduzido cada documento diz respeito à compreensão do oral.

Dada a natureza dos materiais, também eram variados os **gêneros discursivos** (ver 5.3.3. *Codificação*). Os estudantes redigiram textos opinativos ou narrativos, cartas pessoais ou um pouco mais formais, receitas de cozinha, ou até diálogos curtos, consoante as directrizes do exercício ou as suas preferências. No entanto, nem sempre foi fácil atribuir aos textos uma dada classificação, já que muitos textos eram mistos: cartas que exprimiam uma opinião, comentários ou cartas que pelo meio contavam uma história...²¹⁷. Os chineses, por exemplo, têm uma marcada preferência pela área temática trabalho e pelo género narrativo, enquanto que os suecos são bastante opinativos e preocupados com o tema alcoolismo. A observação destas e de outras preferências presentes na totalidade dos materiais poderia até constituir material interessante para estudos de análise do discurso e estudos antropológicos ²¹⁸. De qualquer modo, a quantidade de material disponível para selecção revelou-se suficiente para constituir um *corpus* variado e que observa um certo equilíbrio entre os diferentes *sub-corpora*, em particular no que respeita o número de documentos por tema (ver *ANEXO 1*).

Para poder confrontar o vocabulário usado por falantes de L2 com o usado por **falantes de português L1** nas mesmas circunstâncias, uma vez seleccionados os documentos que iriam constituir cada um dos *sub-corpora*, e tendo em consideração o peso de cada um dos temas na totalidade dos documentos, foram preparadas cópias da segunda parte dos enunciados das mesmas provas de exame que tinham sido apresentadas aos estudantes estrangeiros, e foi pedido a estudantes do 1.º ano do Curso de Língua e Cultura Portuguesa e do 4.º ano do Curso de Línguas e Literaturas Modernas da Faculdade de Letras da Universidade de Lisboa que redigissem

²¹⁷ Desta dificuldade dão testemunho também Nascimento *et alii* (1997:117), a propósito do projecto PAROLE: «Aquando do estabelecimento dos critérios a ter em consideração no desenho do *corpus*, foi amplamente debatida a necessidade de serem tidos em conta aspectos temáticos identificados por 'género' e 'tópico'. Constatou-se, no entanto, a impossibilidade de conseguir quer uma distinção objectiva entre estes parâmetros quer o consenso entre os parceiros quer, ainda, a obtenção de resultados....» [uma vez que] «uma classificação temática coerente decorre da análise linguística interna dos textos, o que não seria possível no tempo de execução deste projecto.»

²¹⁸ *Os documentos que constituem o corpus foram digitados por uma antropóloga (e mil vezes revistos por mim) que me fez esta observação. Agradeço aqui a Ana Mafalda Dourado o cuidado e o entusiasmo que pôs nesta tarefa e a ajuda que me deu na decifração de algumas ocorrências mais opacas.*

composições sobre os mesmos temas ²¹⁹. Os materiais assim obtidos constituem o **sub-corpus de controle**.

A amostra é constituída por aproximadamente 50 documentos (entre 100 e 400 palavras) por cada uma das L1, num total de 309 documentos: 52 documentos foram produzidos pelo grupo de controle de falantes de português L1; 50 por falantes que têm o espanhol como L1; 53 que têm o francês; 53 que têm o alemão; 52 que têm o sueco; e 49 que têm o chinês (cantonês ou mandarim) ²²⁰ (ver *ANEXO 1*). Cada um dos *sub-corpora* tem aproximadamente 11.300 palavras, ou seja, um total de 68.000, o que corresponde a um pouco menos de 10 % do total de palavras seleccionadas das 1400 entrevistas do *Português Fundamental* (1.400 x 500 = 700.000).

5.3.1. Transcrição

Como disse antes, os documentos foram produzidos em situação de exame. Por isso, o que disse em Leiria 1991 também se aplica neste caso: «os textos que compõem este *corpus* apresentam certas semelhanças com alguns manuscritos literários: são a versão que, num dado momento, o seu autor considerou ser a que mais se aproximava de uma hipotética versão final. Para se chegar a ela, e consoante o seu autor, mais ou menos hesitações foram ultrapassadas e logo a seguir substituídas por outras. No entanto, há marcas de que ainda se trata de um rascunho, ou seja, de um texto em construção de que os segmentos riscados dão testemunho, numa tentativa, até ao último minuto disponível, de melhorar o texto aproximando-o de um outro que, caso tivesse havido tempo, teria sucedido àquele que é observável.» (Leiria 1991, vol. II:ix).

Tal como aconteceu em Leiria 1991, a minha principal preocupação neste caso também foi transcrevê-los com o máximo cuidado de modo a que os dados fossem o mais fiáveis possível. No entanto, não encontrei, no âmbito da Linguística Aplicada, modelos de transcrição para materiais com

²¹⁹ Agradeço a Rita Marquilhas ter-se disponibilizado a aplicar numa das suas aulas estes testes e aos alunos do 4º ano de LLM que se dispuseram a realizá-los. Agradeço também aos meus alunos de Língua Portuguesa I do ano lectivo de 1996-1997.

²²⁰ Contrariamente ao que foi decidido no caso do PF (seleccionar um texto de 500 palavras de cada uma das entrevistas), decidi manter todo o texto; daí que, para obter o mesmo total de palavras, seja diferente o número de documentos de cada *sub-corpus*.

estas características. «Há, contudo, uma tradição editorial, nomeadamente entre os medievalistas, que tem em atenção e regista as variantes que ocorrem em diferentes cópias de um mesmo original hoje indisponível, ou que assinala, em caso de dificuldades de leitura, as formas conjecturadas pelo editor. Recentemente, estes métodos têm sido aplicados a textos literários modernos.» (Leiria 1991, vol II:viii). Assim, tomei como base uma pequena parte dos (muitos) símbolos e procedimentos adoptados pela Equipa Pessoa na transcrição de autógrafos literários. Os documentos manuscritos, de que o *ANEXO 2* é um exemplo, foram transcritos de modo a darem conta de particularidades de escrita e de leitura, segundo as seguintes convenções:

<p>< xxx > <i>segmentos riscados</i> < (.....) > <i>segmentos riscados ilegíveis</i> / xxx / <i>segmentos acrescentados</i> /* xxx / <i>leituras conjecturadas</i></p>
--

O *ANEXO 3* é a transcrição do *facsimile* reproduzido no *ANEXO 2*. A informação que uma **edição diplomática** como esta pode proporcionar, em particular a transcrição de segmentos riscados, revelou-se de muito interesse em Leiria 1991. Verifiquei, na altura, que muitos dos segmentos riscados, sobretudo em documentos de alemães e suecos, diziam exactamente respeito ao problema em estudo – o aspecto verbal codificado pelos pretéritos perfeito e imperfeito – e davam conta das muitas hesitações que os falantes nativos destas línguas experimentam quando se devem decidir por um ou outro. No entanto, para não sobrecarregar a transcrição, não dou conta das possíveis variantes de leitura. Em muitos casos, nem sempre foi fácil decidir entre dois ou, por vezes, três grafemas possíveis num mesmo contexto. Isto aplica-se particularmente à distinção entre os grafemas **r/v** ou **m/n** e a certas vogais em determinados contextos de escrita. Apesar de eventuais erros de leitura, a opção por este tipo de edição revelou-se particularmente produtiva no caso presente porque, como se poderá ver na análise, desvios formais na escrita das palavras permitiram-nos chegar a resultados interessantes, entre outras coisas, sobre a aquisição do sistema fonológico.

O rigor exigido por uma edição diplomática só não foi observado no caso da **pontuação**. Uma vez apartados da informação fornecida pela organização do texto na página, a não intervenção do transcritor a este nível teria tornado muitos documentos, depois de transcritos, completamente incompreensíveis. Isto aplica-se particularmente aos documentos do *sub-corpus* chinês. Neste caso, não só foram introduzidos pontos finais, como nos outros, mas também foram eliminados outros ²²¹.

5.3.2. Organização

O ANEXO 1 também dá conta da ordem por que foram organizados os diferentes *sub-corpora* uns em relação aos outros. Essa ordem foi estabelecida tendo em conta a relação tipológica e aquilo que me pareceu ser a distância linguística entre o português e as outras línguas. Embora o conceito de distância linguística seja difícil de aplicar de uma forma rigorosa, a ordem *línguas românicas* → *línguas germânicas* → *chinês* não me parece passível de controvérsia. Do mesmo modo que me parece incontestável a ordem *espanhol* → *francês*. No entanto, não posso dizer, sobretudo antes de chegar a alguns resultados da análise, que não tivesse tido algumas dúvidas quanto à posição relativa do alemão e do sueco. Mas, na verdade, os resultados da análise parecem aconselhar e reflectir a ordem adoptada: primeiro o alemão e depois o sueco.

O ANEXO 1 também dá conta dos diferentes temas presentes e do modo como foram organizados os documentos dentro de cada *sub-corpus*. Eles foram organizados em três grandes áreas temáticas (o **indivíduo**, a **sociedade** e o **meio ambiente**). Dá conta ainda da distribuição dos documentos por cada uma das sub-áreas temáticas. Apesar das preferências de certos grupos por determinados temas, a selecção foi feita de modo que o número de documentos dentro de cada sub-área fosse sensivelmente o mesmo. Quando comparada com «os temas versados nas mil e quatrocentos entrevistas do Português Fundamental» (*PF Métodos e Documentos*, tomo 1:26), a informação fornecida no quadro mostra que, de facto, esses temas são bastante coincidentes.

²²¹ Os chineses usam pontos finais que parecem ter a função de separar um sintagma do seguinte. Cagliari (1993/1999:172-173), a propósito do alfabeto grego, comenta: «Os antigos costumavam escrever as palavras sem separação, emendando uma na outra. Quando queriam evitar ambiguidades ou destacar palavras, usavam, em geral, um ponto separando-as. É curioso encontrar hoje crianças em processo de alfabetização que fazem isto por iniciativa própria.»

5.3.3. Codificação

Cada documento encontra-se identificado no *corpus* **PESTRA** por um código que contempla:

1. <i>L1 do informante</i>	Espanhol, Francês, Alemão, Sueco, Chinês
2. <i>número do documento</i>	1- 53
3. <i>área temática e tema da composição</i>	K (a – j) ; X (1 – p) ; Z (q – v)
4. <i>género discursivo</i>	a : opinativo; b : narrativo; c : carta; d : outro
5. <i>compreensão do oral</i>	0- 4 ²²²

Tome-se como exemplo o código pelo qual foi introduzido o documento reproduzido no *ANEXO 2* e transcrito no *ANEXO 3*:

A 35 x1 b 2

Ele significa que o documento faz parte do *sub-corpus* ALEMÃO, é o número 35, da área temática X (a sociedade) e do tema L (hábitos sociais e comportamentos), que é um texto narrativo (b) e que à prova de compreensão oral do informante foi atribuído o nível 2.

5.4. Para a construção da base de dados

Recolher, seleccionar, organizar e codificar um certo número de materiais de modo a obter um *corpus* adequado à investigação é só um primeiro passo no longo e árduo caminho para a construção de uma base de dados que permita aceder, num curto espaço de tempo, à informação que ele pode disponibilizar. No caso da investigação em L2, em geral, o que interessa é obter informação que nos permita descrever e caracterizar um ou vários aspectos de um certo estágio de interlíngua. Neste estudo, em particular, o que me interessa é construir uma base de dados que forneça informação quanto à dimensão do vocabulário e quanto a desvios no léxico. Por isso, era necessário definir **critérios de lematização** e estabelecer um **sistema de marcação de desvios**.

²²² Este código não se aplica ao grupo de controle.

5.4.1. Para a avaliação da dimensão do vocabulário

Para além do grupo de controle de falantes de português L1, pareceu-me pertinente recorrer também a um outro instrumento com que pudesse confrontar o vocabulário usado pelos informantes estrangeiros. Disponível e adequado para esse fim era o **Português Fundamental** (PF). O PF faz parte de uma linha de investigação, desenvolvida entre meados dos anos 50 e o início dos anos 80 (ver, capítulo 4, *Léxico e ensino de L2*), e pode definir-se como «o vocabulário indispensável a uma efectiva capacidade de comunicação em situações da vida corrente». Porque «o ensino de uma língua impõe, como é evidente, uma selecção do extenso vocabulário que a constitui, de modo a estabelecer uma progressão na aprendizagem», «competiu desde sempre aos autores de manuais e aos professores proceder a essa selecção, tarefa que, na falta de um método científico em que pudessem apoiar-se, realizavam de forma empírica e baseados na intuição» (PF, tomo 1:15). Um dos seus objectivos consiste exactamente em contrariar o apoio na intuição, fornecendo «os dados que permitam estabelecer conteúdos lexicais adequados ao ensino do Português, língua estrangeira, desde o nível mais elementar da aprendizagem.» (PF, tomo 1:15) ²²³.

A lista de vocabulário designada por *Português Fundamental* é constituída por 2217 palavras. Este número corresponde aproximadamente a 1/5 do total de 6 000 vocábulos diferentes que o inquérito português permitiu recolher. Um *Corpus de Frequência* e um *Corpus de Disponibilidade* constituem o suporte material que serviu de base à sua selecção. O vocabulário do PF é constituído por vocábulos obtidos

- 1) exclusivamente do Inquérito de Frequência (741);
- 2) exclusivamente do Inquérito de Disponibilidade (646);
- 3) simultaneamente dos dois inquéritos (407);
- 4) e por vocábulos acrescentados por uma comissão de filólogos e linguistas (423) ²²⁴.

²²³ O outro objectivo, lamentavelmente nunca atingido, consistia em «apresentar os resultados de análises gramaticais sobre aspectos morfo-sintácticos do discurso oral, de interesse mais imediato para o ensino do Português.» (PF, tomo 1:15)

²²⁴ Para mais informações sobre o vocabulário seleccionado, ver Maria Luísa Segura da Cruz (1984), *PF Métodos e Documentos*, tomo 2: parte quarta.

Os vocábulos que constituem a lista foram seleccionados porque atingiram o limiar de frequência 40 ou porque, embora não o tendo atingido, foram incluídas por recomendação da comissão ou por pertencerem a séries (ex: *abril, outono*).

Projectos como o *PF* revelam um avanço significativo no ensino das línguas estrangeiras, na medida em que estão particularmente preocupados com as «necessidades de utilização da linguagem no dia-a-dia» (Rivenc 1987:8). No entanto, eles foram desenvolvidos numa época em que a preocupação estava muito mais centrada no ensino do que na aprendizagem. Actualmente, a investigação conducente a um melhor ensino procura apoiar-se também em investigação desenvolvida sobre a aprendizagem. Este trabalho inscreve-se exactamente nesta perspectiva.

No entanto, uma questão metodológica se levanta. Ela diz respeito ao confronto de dados obtidos a partir de um **corpus escrito** com dados obtidos a partir de um **corpus oral** como é o *PF*. McCarthy e Carter (1997:23-25) comparam, com base em 330 000 palavras do material escrito com outras tantas do oral do *Cambridge International Corpus*, as 50 palavras mais frequentes das duas amostras. Como seria de esperar, as palavras funcionais dominam o topo de ambas as listas, mas a lista de palavras do oral inclui um certo número de palavras lexicais (*know, well, got, think, right*) que, observadas mais de perto, se verifica serem marcadores interpessoais (*you know, I think...*) ou *organizadores do discurso* (*well, right*). Comparemo-las agora com a “Lista de vocábulos por frequências decrescentes” do *PF* (*PF* tomo 1:689-751), tendo embora em consideração que esta última é uma lista lematizada. Se juntarmos à lista de 13 “vocábulos gramaticais homógrafos, de alta frequência não analisados” mais umas 40 palavras, verificaremos que há muito mais palavras lexicais do que na lista inglesa correspondente. É claro que uma das razões para isso prende-se certamente com o facto de, na verdade, algumas delas fazerem também parte de marcadores interpessoais (*ser – não é? , saber – sabe(s)...*) ou organizadores do discurso (*depois*). Mas a principal razão para que isto aconteça parece-me resultar do tipo de materiais do *PF* (ver *Entrevistas*, tomo 1:79-309). Eles não são aquilo a que McCarthy e Carter (1997:29) chamam “*language-in-action*”, ou seja, situações em que a língua é directamente gerada por acções que os falantes

estão a realizar, mas sim, em muitos casos, a narração dessas mesmas acções ou a expressão de opiniões sobre elas. Isso é patente no número de linhas que ocupa, em muitos casos, cada uma das falas dos entrevistados. E nesse sentido aproximam-se dos textos escritos que constituem este *corpus* de L2. Se por um lado são mais ricos em palavras funcionais, por outro também o são em palavras lexicais. É por esta razão que me parece possível confrontar dados obtidos a partir deste *corpus* escrito com dados obtidos a partir um *corpus* oral com as características do *PF*.

5.4.2. *Para a construção da tipologia de desvios*

Outra questão particularmente controversa diz respeito à pertinência ou não de a investigação em aquisição de L2 optar por descrever a interlíngua em si mesma ou optar por compará-la com uma variedade nativa ²²⁵. Mas, se, como é aqui o caso, estamos interessados naquilo que separa a interlíngua da variedade-alvo, ou seja, entre aquilo que está representado na mente do aprendente e aquilo que está representado na mente do falante nativo, se a essa representação só temos acesso indirecto através do discurso produzido, e tendo em conta que as diferenças se manifestam nos desvios que o discurso produzido apresenta em relação aos falantes nativos, então uma **análise de erros** parece inevitável. O que não implica, necessariamente, excluir a possibilidade de descrever a interlíngua em si mesma. Ou como diz Carl James (1998:7): «We do have to have a detailed and coherent description of the learners' repertoires, but we cannot stop there...». Assim, antes de proceder à marcação dos desvios, impunha-se uma resposta a algumas questões: **Quais os critérios para considerar uma forma ou uma estrutura como desviante? Que sistema**

²²⁵ «There is a **constant tension** between, on the one hand, the long-term descriptive and explanatory priorities of people engaged in IL studies and dedicated to SLR and, on the other hand, the shorter-term pedagogic priorities of foreign language educators who do EA [error analysis]. One is tempted at times to conclude that SLR is not a part of applied linguistics, and is not interested in language teaching, but is a branch of pure linguistics, interested in the properties of language rather than the problems of learners, in language learnability rather than the processes of teaching. There does not have to be tension if we agree that the SLR and EA enterprises are different and have different goals. After all, as Cook puts it: 'Error analysis was [and still is! CJ] a methodology for dealing with data, rather than a theory of acquisition' (1993:22). Let those who want theory, and those who want them, ways of 'dealing with data'. In one sense, however, SLA is inescapably comparative. [...] The SLA researcher who studies IL developmentally and longitudinally, like the historical linguistic, will be forced to make comparisons of these successive stages.» (James 1998:7).

de marcação, económico e satisfatório, usar para identificação da natureza tão diversa dos desvios? Como lidar com o problema da não existência de fronteiras entre o léxico e a sintaxe?

Lennon (1991:182) define **erro** como «a linguistic form [...] which, in the same context [...] would in all likelihood not be produced by the learner's native speaker counterparts.» Carl James (1998:64) chama a atenção para a forma engenhosa como nesta definição, ao ser usada a expressão '*linguistic form*', se evita a distinção entre intenção semântica e intenção formal, e como, ao comparar os aprendentes com os '*native speaker counterparts*', se introduzem as variáveis nível de educação, idade, classe socio-económica, sexo e até mesmo profissão. É claro que a abrangência desta definição, se por um lado relativiza o peso de certos desvios, por outro também nos faz pensar que, na sua identificação, é preciso ir muito para além do mero critério de boa ou má formação de um enunciado ou de uma frase. James considera que o desvio deve ser avaliado segundo quatro categorias: gramaticalidade, aceitabilidade, correcção, e estranheza e impropriedade. **Gramaticalidade**, ou seja, boa formação, anda de mãos dadas com aceitabilidade; por outras palavras, a gramaticalidade é um pré-requisito para a **aceitabilidade**; no entanto, enquanto que é o conhecedor de uma língua quem decide se um enunciado é gramatical, é aquele que a usa quem decide se esse enunciado é aceitável. E, para decidir da sua aceitabilidade, não tem em consideração regras mas contextos. Por vezes, uma frase ou um enunciado pode ser inaceitável por apresentar um certo grau de estranheza, que é uma forma de impropriedade, e não de agramaticalidade. E essa estranheza resulta de trair as expectativas de **idiomaticidade** do ouvinte ou do leitor. Do mesmo modo que uma frase ou um enunciado pode ser espontaneamente aceite, mas, quando filtrada por conhecimento metalinguístico (ou explícito) pode ser rejeitada por não responder ao critério de **correcção** (James 1998:66-76).

Tratando-se de materiais, nalguns casos, tão desviantes a todos os níveis, a questão que se punha em seguida – encontrar um sistema de identificação e marcação de desvios eficiente e económico que desse resposta a todos estes critérios – não se afigurava como tarefa fácil. Graeme Kennedy diz a propósito da etiquetagem gramatical usada por diferentes projectos na análise automática de *corpora*:

In the final analysis, the value or 'success' of the various approaches to word-class tagging may not be crucially dependent on theoretical differences. It is largely an empirical matter within a theoretical framework as to whether a tagging system is more or less successful in correctly tagging a sentence, a text or indeed a whole corpus. (Kennedy 1998:223-224)

Creio que o mesmo se pode dizer neste caso. Estabeleci, então, um **sistema de marcação de desvios** relativamente impressionista, mas muito económico. Havendo sempre a possibilidade de recuperar informação a partir do contexto, o conjunto que se segue de nove etiquetas coladas a um item lexical ou a uma sequência permitiu, na fase da análise, uma reinterpretação ou uma afinação do seu significado. Estou absolutamente convicta de que, se tivesse estabelecido, à partida, uma marcação menos empírica e mais apoiada em hipóteses linguísticas, ela ter-me-ia impedido o confronto entre certas ocorrências que se revelou decisivo para a sua interpretação.

MARCAÇÃO DE DESVIOS	
X1	<i>o falante nativo teria usado outro item lexical</i>
X2	<i>a variedade-alvo não reconhece ou este contexto não aceita este item lexical ou esta sequência</i>
X3	<i>este item não coincide com a forma normalizada</i>
X4	<i>este item apresenta um desvio relativo a qualquer tipo de acordo</i>
X5	<i>o falante nativo teria usado outro tempo verbal</i>
X6	<i>este item <u>não</u> foi usado; o falante nativo tê-lo-ia usado</i>
X7	<i>este item foi usado; o falante nativo <u>não</u> o teria usado</i>
X8	<i>o falante nativo teria colocado este item noutra local da frase</i>
X9	<i>não foi possível pôr nenhuma hipótese quanto à intenção do autor</i>

5.4.3. Anotação do corpus e marcação dos desvios

Tratando-se de um projecto que, a partir de um *corpus* de L2, pretende relacionar dimensão do vocabulário disponível para a produção escrita com profundidade do conhecimento lexical, a construção de uma base de dados que permitisse atingir os objectivos da investigação exigia uma intervenção a vários níveis: anotação, lematização, e identificação e etiquetagem dos vários tipos de erros.

Mas, chegados a este ponto, não é difícil prever a dificuldade que representa definir o que se entende por palavra. Como é sabido, o conceito de palavra pode ser abordado de vários pontos de vista – ortográfico, fonético, semântico e gramatical – apresentando

*qualquer um deles vantagens e inconvenientes para a sua definição*²²⁶. *Outros sistemas de escrita diferentes do nosso mostram-nos que o critério ortográfico não é o mais científico. O critério da pausa potencial, usado na abordagem fonética, não parece melhor. A abordagem semântica seria boa, se não houvesse unidades simples de sentido codificadas por mais de uma palavra (palavra-chave, por exemplo...). A abordagem gramatical e o critério da mobilidade dariam resposta adequada, se não tivéssemos, por exemplo, preposições e artigos... Mas o quadro complica-se ainda mais quando, como é aqui o caso, estamos a operar com léxicos mentais bilingues. Esse facto torna ainda mais evidente não só a dificuldade de definir o que se entende por palavra mas também a de distinguir entre léxico, morfologia e sintaxe. Admitamos então que este problema não tem solução à vista e aceitemos o conceito de palavra como um ‘primitivo’, como uma unidade linguística sobre a qual há “um acordo intuitivo” (Scalise 1994:59)*²²⁷. *Esse acordo intuitivo parece, de facto, coincidir com a palavra gráfica. É verdade que a palavra não é um produto da literacia, mas também é certo que para nós, sujeitos alfabetizados, o ponto de vista ortográfico – conjunto de caracteres entre dois espaços em branco – parece ser o que corresponde melhor à representação mental que temos da palavra palavra.*

Apesar de se tratar de material oral, esta foi também a opção dos investigadores que trataram os materiais do PF: «A escolha do critério de segmentação do texto em palavras gráficas, que é geralmente utilizado em estatística lexical, impõe-se por necessidade de normalização, já que garante o tratamento uniforme e homogêneo do corpus em análise. Foram ainda tomadas como unidades lexicais alguns grupos de palavras que, pela estabilidade da relação significante-significado, pelo valor monossémico do grupo e frequência de emprego, já estão consagrados, na maior parte dos casos, pela grafia. São exemplos disto palavras compostas por justaposição como FIM-DE-SEMANA, MEIO-DIA, MEIA-NOITE, que passarão a ser referidas neste texto apenas por ‘palavras compostas’». Posteriormente, nos casos de homografia, foram feitas, excepto no caso das palavras gramaticais com frequência superior a 500, «as análises necessárias à distinção entre classes gramaticais diferentes ou, nalguns casos, entre comportamentos diferentes dentro da mesma classe»; também posteriormente, «foi feita uma pesquisa sistemática de locuções de tipo gramatical – locuções prepositivas, conjuncionais e pronominais e, ainda, de locuções adverbiais» (PF, volume I, tomo 1: 21, 23 e 29). Os mesmos critérios foram adoptados neste trabalho. Optou-se, no entanto, por anotar no corpus tanto as palavras compostas como as homógrafas e as locuções.

A anotação, que permitisse a posterior **lematização** das formas, tinha, portanto, em vista, fundamentalmente, **(1)** juntar sob uma mesma entrada as diferentes formas de um mesmo verbo ou de ocorrências de outras classes que devido a desvios formais (ortográficos ou outros) não

²²⁶ Ver Capítulo 1, nota 31.

²²⁷ Ver também, entre outros, Aronoff 1976 e Villalva 1994.

fossem reconhecidas pelo programa como semelhantes; **(2)** impedir a contagem em separado de formas que constituíssem, por exemplo, um tempo verbal composto ²²⁸ ou uma locução; e **(3)** desambiguar formas homógrafas. Para a identificação das homógrafas, e para não sobrecarregar a já de si morosa e aborrecida tarefa de anotação manual do *corpus*, estabeleceu-se como critério geral que só se anotariam as menos frequentes; por exemplo, aos artigos *o* e *a* não foi acrescentada nenhuma etiqueta, mas sim à preposição e aos pronomes homógrafos.

Para o modo como foi anotado o *corpus*, sirva como exemplo o *ANEXO 4*. Como se pode ver, (confrontando o *ANEXO 3* com o *ANEXO 4*) a anotação foi feita sobre uma cópia da qual foram apagados todos os segmentos riscados. Para que não fossem contadas como itens diferentes, todas as formas foram separadas, por um espaço, dos sinais de pontuação que as seguiam. Como também se pode ver, todas as etiquetas são constituídas por grafemas maiúsculos (na maior parte dos casos por, pelo menos, dois) de modo a serem interpretadas inequivocamente como não fazendo parte do texto original; as diversas etiquetas foram coladas às formas por um “_”; o mesmo procedimento foi usado para prevenir o afastamento de quaisquer outras formas que constituíssem uma sequência (tempos compostos, auxiliares aspectuais mais preposição, locuções...). As formas verbais conjugadas foram precedidas do seu infinitivo, imediatamente seguido de uma etiqueta que identifica o tempo verbal em que está conjugado (independentemente de ser ou não adequado ao contexto); às formas verbais homógrafas na primeira e terceira pessoa foram acrescentadas as etiquetas “11” e “33” (ver, por ex., “ela DIZER_VPP_disse33”, no *ANEXO 4*).

Como disse acima, foi estabelecido um sistema de nove etiquetas para a **marcação dos desvios**. No caso de **X6**, que marca as omissões, como se pode ver no mesmo anexo, o item omitido foi introduzido em maiúsculas de modo a ser facilmente identificado como não pertencendo ao original. Quanto a **X4**, no caso dos verbos, sempre que uma forma verbal não se apresentava conjugada na pessoa verbal exigida pelo contexto, além desta etiqueta, que marca um erro de acordo, foi acrescentada uma outra referente à pessoa exigida pelo contexto. Veja-se:

²²⁸ Ver Gonçalves 1992.

[Nós] IR_VPP_**Foram44**_X4 de carro e TER_VMP_**tinham44**_tido_X4 a mesma problema_X4 com_X1 todas as outras pessoas com a mesma idea_X3

No entanto, como era previsível, muitas dúvidas foram surgindo no que respeita a identificação de certos desvios. Veja-se só um exemplo, no mesmo ANEXO 4:

E a minha amiga não PERCEBER_VPP_percebiu_X4_X5 que_X1 eu ESTAR_A_VCI_estivesse11_a_X5 rir. Mas **porque**_X1 , isto SER_é muito_AV bom !

Como interpretar este último erro? Como um erro de ortografia (X3), como um erro de sintaxe (X1), ou como um erro de pontuação? Optei por classificá-lo como **X1** porque a experiência de ensino me diz que os aprendentes costumam ter muitas dificuldades nestas interrogações não frásicas e que, por isso, não se tratava, com certeza, de um mero “esquecimento” de um diacrítico. Mas nem sempre a opção foi tão fácil e tão apoiada. No entanto, não se trata aqui de análise mas de marcação de desvios, e, como referi antes, o programa de busca e concordâncias permite, em fases posteriores da investigação, aproximar formas iguais ainda que marcadas com etiquetas diferentes. E se, de facto, se tratava de um erro significativo, porque frequente, ele pôde ser reavaliado ²²⁹.

5.4.4. O programa

A exploração de um *corpus* é uma tarefa muito repetitiva e que consome muito tempo; itens e estruturas semelhantes têm de ser marcados cada vez que ocorrem nos dados. Por outro lado, de cada vez que uma hipótese é modificada, os mesmos dados têm de ser todos novamente analisados. Daí a necessidade de usar ferramentas computacionais. Para isso, impunha-se dar resposta a duas questões prévias e interrelacionadas:

²²⁹ Destaco aqui a vantagem de todo o *corpus* ser anotado pelo mesmo investigador e a importância do hábito de lidar com produções desviantes. Na verdade, todo o *corpus* deveria ser anotado, pelo menos, por dois ou mesmo três investigadores independentes e comparadas as diferentes versões *a posteriori*. Mas, como se pode compreender, neste caso, não houve condições para isso.

Que programas de tratamento computarizado de dados estão disponíveis para este tipo de *corpora*? Que programa escolher de modo a conciliar o tempo disponível para a anotação do *corpus* com a obtenção de uma base de dados capaz de vir a fornecer informação satisfatória?

Era preciso, então, escolher um programa informático que respondesse às restrições impostas por um *corpus* de interlíngua, neste caso específico, particularmente desviante em relação à variedade-alvo. Se outros desvios não houvesse ou não fossem tomados em consideração, bastariam todas as omissões e as inúmeras formas com grafias divergentes, consequentemente irreconhecíveis para qualquer programa informático de tratamento de variedades nativas, para dificultar a sua aplicação. Isto seria suficiente para que tivessem de ser excluídos os muito difundidos programas que elaboram listas de palavras e fornecem o número de ocorrências de cada uma delas. Por outro lado, e ainda que se pretendesse exclusivamente obter frequências de palavras, este tipo de programas não distingue formas homógrafas, comete alguns erros de lematização ²³⁰ e não reconhece formas constituídas por mais de uma palavra gráfica.

Apesar de haver um certo número de programas de tratamento computarizado de dados bastante difundido entre os investigadores da língua portuguesa ²³¹, e de ter tido conhecimento da existência de programas desenvolvidos para o tratamento de L2 ²³², depois de ponderar as

²³⁰ Sobre o estado da lematização automática, ver, por exemplo, Kennedy 1998.

²³¹ Nomeadamente o *VARBRUL*, muito usado pelos investigadores brasileiros, e que se destina ao tratamento estatístico de dados linguísticos variáveis, analisados na perspectiva da teoria da variação linguística laboviana; ou o *Oxford Concordance Program (OCP)*, que faz codificação, levantamento e quantificação de dados e que foi usado pelo projecto *Panorama do Português Oral de Maputo (PPOM)* para tratamento de dados do português de Moçambique, ou o *CHILDES (Child Language Data Exchange System)*, desenvolvido por Macwhinney & Snow (1985), que é constituído por três ferramentas : uma base de dados (*CHILDES*), um sistema de transcrição (*CHAT*), e programas de análise (*CLAN*). Para informação detalhada sobre esta ferramenta, ver Batoréu 2000:587-602 e Faria & Batoréu 1994.

²³² Nomeadamente, *COALA* (acrónimo de *Computer-Aided Linguistic Analysis*) desenvolvido no *Language Acquisition Research Centre* da Universidade de Sydney, apresentado em Pienemann 1992; e *COMOLA* (acrónimo de *Computer Model for Language Acquisition*), desenvolvido no *Department of Language and Speech* da Universidade de Nijmegen, apresentado em Jagtman & Bongaerts 1994.

vantagens e desvantagens da sua utilização ²³³, optei por encomendar a um informático um programa simples de contagem de frequências e busca de palavras a que fosse possível ir acrescentando módulos à medida das minhas necessidades de análise.

O programa é constituído por um conjunto de módulos que actuam dentro do *Microsoft-EXCEL*. Um módulo – **programa de frequências** – prepara listas de palavras e conta a frequência de diferentes formas num dado ficheiro. Outros dois – **programas de busca** – procuram uma dada forma lexical sem contextos; ou procuram uma dada sequência de grafemas independentemente de constituírem ou não uma palavra. Finalmente, um outro módulo – **programa de busca e concordâncias** – procura uma dada palavra ou uma sequência de grafemas e fornece todos os seus contextos (três palavras à esquerda e três à direita) ²³⁴ e indica o ficheiro em que cada um deles ocorre. O facto de os programas correrem sobre *EXCEL* permite introduzir colunas com informação suplementar e ordenar os dados por qualquer uma das colunas.

As características do programa permitiram uma grande liberdade no que respeita a anotação do *corpus*, que se traduziu na possibilidade de usar tantas etiquetas quantas se revelaram necessárias. Isto não significa, no entanto, que não tivesse sido feito um grande esforço de contenção, na medida em que quanto maior fosse o seu número mais probabilidades de ocorrerem erros de digitação existiriam ²³⁵.

5.5. *Dados para a análise*

Uma vez anotados todos os documentos de um *sub-corpus*, foi cada um desses documentos introduzido em *EXCEL* e tratado pelo programa de frequências tendo-se obtido uma lista de ocorrências de cada um deles. O *ANEXO 5* mostra a lista resultante do documento **A35**. A partir desta foi possível também fazer uma contagem dos desvios em cada um dos

²³³ Informação sobre programas de tratamento computarizado de dados pode encontrar-se em McEnery & Wilson 1996 e Kennedy 1998.

²³⁴ A opção por um contexto tão limitado revelou-se vantajosa e desvantajosa. A principal vantagem consiste em ocupar só uma linha no écran. A principal desvantagem consiste em, muitas vezes, o contexto se revelar insuficiente e obrigar a consultar o texto. De qualquer modo, a primeira sobrepôs-se à segunda.

²³⁵ Sobre codificação e anotação ver, entre outros, McEnery & Wilson 1996.

documentos e, em seguida, em cada um dos subcorpora. É claro que, nesta fase, não se trata de uma análise, mas sim de uma avaliação em bruto de cada um dos documentos ²³⁶.

O conjunto de documentos que constitui cada um dos *sub-corpora* foi transformado num único ficheiro, foi tratado pelo mesmo programa, e obteve-se uma lista da totalidade das ocorrências nesse *sub-corpus*. O ANEXO 6 é um extracto da lista obtida do *sub-corpus* ALEMÃO. Como se pode ver, essa lista apresenta todas as diferentes formas de uma mesma palavra e os diferentes desvios que cada uma delas apresenta ²³⁷.

Tendo em conta que um dos objectivos da investigação consiste em avaliar o vocabulário disponível, independentemente dos desvios que o caracterizam, para atingir esse objectivo foi necessário proceder, a partir destas listas, a uma **lematização** manual, necessariamente consumidora de muito tempo e, certamente, não isenta de falhas. O ANEXO 7 é um extracto da lista lematizada do *sub-corpus* ALEMÃO, com a frequência de cada forma e, na terceira coluna, com a informação sobre se faz (1) ou não (2) parte do *PF*. Para a elaboração desta lista foram ignorados todos os nomes próprios (*Alemanha, China, Lisboa, Costa da Caparica, etc.*), estrangeirismos (*fast-food, hippies, jeep, kodak, etc.*) e desvios. Os desvios formais foram contados como se se tratasse de formas não divergentes. Assim, por exemplo (ver ANEXOS 2 e 3) *idea, engressado, la, agente*, foram contados nas entradas *ideia, engraçado, lá e gente*.

5.5.1. *Vocabulário disponível*

A avaliação feita sobre um *corpus* como este, apesar do número de ocorrências ser relativamente limitado, é paralela à desenvolvida para L1, nomeadamente para o estabelecimento de vocabulários fundamentais. **Assume-se portanto que pelo menos o vocabulário usado nestes**

²³⁶ Algumas das características de cada documento foram tratadas mas não foram exploradas na análise. Essa informação será útil para, em trabalhos futuros, tentar relacionar ocorrências correctas e desviantes no mesmo documento.

²³⁷ Número de entradas diferentes por L1: português: 3621; espanhol: 3564; francês: 3747; alemão: 3388; sueco: 3477; chinês: 3544. Note-se que estes resultados, na verdade muito semelhantes entre si, não são particularmente significativos do ponto de vista da riqueza lexical porque estas listas incluem todos os desvios, nomes próprios, estrangeirismos e sinais de pontuação. São listas em bruto que resultaram de aproximadamente 24 horas (por cada *sub-corpus*) de actuação do programa de frequências sobre os documentos anotados.

documentos está disponível para a produção escrita e que, se vários aprendentes conhecem uma palavra, é provável que pelo menos os outros do mesmo grupo linguístico, embora não tivessem tido oportunidade para a usar, também a conheçam.

Sendo um dos objectivos do trabalho confrontar o vocabulário usado no *corpus* de L2 com a lista do *Português Fundamental* (PF), para a selecção das entradas, seguiram-se os critérios desta ²³⁸. Assim, por exemplo (ver *ANEXOS 6 e 7*) os determinantes *essa, esse e esta, este* constituem entradas diferentes; do mesmo modo, sempre que uma locução atingiu o nível limiar de frequência no *PF* (ver *por exemplo* nos mesmos *ANEXOS 6 e 7*) foi incluída na lista de locuções (ver *ANEXO 10*). Desta tarefa de lematização resultaram seis listas, uma por cada L1, que permitiram avaliar a dimensão do vocabulário usado por cada um dos grupos e confrontá-lo com a lista do *PF*.

O *ANEXO 8* apresenta um extracto do **vocabulário disponível pertencente ao PF** (lista ordenada por ordem decrescente de número de ocorrências no PESTRA). Como referi antes, os *sub-corpora* não tinham, à partida, o mesmo número de palavras. Isso deve-se ao facto de ter optado, ao contrário do que aconteceu por exemplo com o *PF*, por não seleccionar o mesmo número de palavras de cada um dos textos. Se o tivesse feito, teria perdido muito material, porque os textos tinham extensões muito diferentes (entre 100 e 400 palavras). Tentei, por isso, seleccionar os textos de modo que os vários *sub-corpora* tivessem, entre si, um total de palavras aproximado. Mas, ainda que tivesse seleccionado à partida o mesmo número de palavras, os resultados finais seriam diferentes. E isso porque, por um lado, foram contados os dois elementos que constituíam uma forma contraída (por exemplo, **da** foi contado em **de** e em **a**), e, por outro, foram eliminados os segmentos riscados, os nomes próprios, estrangeirismos ²³⁹ e números. Essa é também a razão por que os comentários que se seguem no capítulo 6 têm preferencialmente em consideração percentagens e não totais.

²³⁸ Como é sabido, os critérios para decidir o que deve constituir uma entrada divergem bastante. Além disso, mesmo tendo-os bem estabelecidos à partida, é preciso, muitas vezes, tomar decisões caso a caso.

²³⁹ Por estrangeirismos entenda-se palavras de outras línguas usadas conscientemente e não como empréstimos para resolver uma indisponibilidade vocabular.

Paul Rivenc, co-autor do *Francês Fundamental* e do *Espanhol Fundamental* e conselheiro científico do PF, diz que René Michéa, a partir da sua experiência como professor de alemão a alunos franceses do ensino liceal, distingue entre **palavras atemáticas**, «aquelas que, com maior ou menor regularidade, podemos estar certos de encontrar em qualquer texto relativamente extenso (com mais de um milhar de palavras) independentemente do seu tema dominante», e **palavras temáticas**, aquelas que estão «ligadas a um determinado tema, ou a um conjunto de temas bem definidos, exprimindo um domínio particular da experiência humana, do pensamento ou da acção». Segundo Michéa, as primeiras servem mais para nos exprimirmos a propósito das coisas do que para exprimir as próprias coisas; as segundas designam os seres, os objectos ou as acções e são por isso, na sua maioria, palavras concretas. A partir desta distinção, Michéa e Gougenheim evoluem para a distinção entre **vocabulário frequente** e **vocabulário disponível**, com «características próprias, mas complementares um do outro e igualmente necessários no funcionamento da comunicação verbal espontânea.» O vocabulário frequente é fornecido pela «análise dos textos recolhidos em situação real de comunicação»; o vocabulário disponível é constituído por «palavras de uma frequência fraca e pouco estável» mas «usuais e úteis», e, ainda que não sejam usadas frequentemente, «estão à nossa disposição, podendo falharmos momentaneamente ou mesmo desaparecer por efeito de uma doença» (Gougenheim *et alii* 1964:137-152 citados por Rivenc 1987:8-9).

O falante de uma L2 não sofre necessariamente de uma qualquer 'doença' linguística, mas o seu léxico mental, no que respeita ao armazenamento, acesso e controle, pode ser comparado com o de um falante de L1 que, por efeito de desenvolvimento, ou seja, porque está a aprender palavras novas, apresenta uma certa instabilidade. Além de poder ser muito grande o número de situações em que o acesso a um determinado item lexical é difícil, o seu léxico mental é completamente deficitário em relação a certas palavras da L2. Nomeadamente nos primeiros estádios de aquisição, ele só dispõe de um número limitado de vocábulos a que pode recorrer para actividades receptivas e/ou produtivas. Acontece que todos os vocábulos usados na amostra estavam disponíveis no momento da produção dos textos independentemente da frequência com que foram

usados. Ou seja, o falante de L2 (e também o de L1...) tem disponível um certo número de palavras que se posicionam num *continuum* que vai das mais atemáticas para as mais temáticas, dependendo o acesso a estas últimas de uma maior ou menor activação, em função de interesses ou hábitos particulares. Assim, **usarei a designação de vocabulário disponível para me referir ao conjunto de vocábulos que foram usados na amostra, vocabulário disponível atemático ou frequente** para me referir aqueles vocábulos que foram usados com frequência relativa alta independentemente do tema, vocabulário *disponível temático* para me referir a vocábulos de frequência relativa baixa, **e usarei o termo vocabulário indisponível para me referir aos vocábulos de que os aprendentes mostraram não dispor no momento em que escreveram o texto tendo usado, por isso, uma qualquer estratégia compensatória.**

5.5.2. Tipologia de desvios

O vocabulário disponível revelou-se por vezes insuficiente, o que obrigou os aprendentes a usarem estratégias que pudessem compensar os vocábulos em falta. Por outro lado, muitos dos vocábulos disponíveis apresentam em muitos casos desvios de naturezas muito diversas que, como vimos, foram assinalados na amostra, e que evidenciam um léxico deficitário.

Dito de outro modo: num dado momento,

1. o aprendente, para uma dada representação conceptual, tem disponível um item no seu léxico mental da L2; no entanto, ele apresenta um ou mais desvios formais, ou seja, desvios que afectam exclusivamente o item em questão, mas não, ou só indirectamente, o enunciado ou a frase. Pode ser um erro de ortografia (accidente por <acidente>) ou de acentuação (esta por <está>), uma forma aproximada (livradade por <liberdade>), um desvio na flexão (papels por <papéis>); sento por <sinto>), ou na atribuição de género (a problema por <o problema>).

2. o aprendente não tem disponível no seu léxico mental da L2 nenhum item lexical que corresponda à representação conceptual; então, porque é activado e o falante não consegue controlá-lo, ou porque está convencido de que faz parte do léxico da L2, usa de empréstimo um item de uma outra língua, aplicando-lhe ou não estratégias que o tornem mais conforme com aquilo que considera ser típico da L2; ou seja, consciente ou inconscientemente, usa um item de outra língua ou faz um item novo. Outra estratégia possível, que parece ser a mais frequente sempre que o falante tem consciência de que o item não está disponível, consiste em usar palavras

disponíveis da L2, recorrendo a uma estratégia de reconceptualização. Esta estratégia, quando bem usada, pode ser de difícil identificação.

3. e, como de um léxico mental não fazem parte só itens lexicais simples, uma outra forma de déficit lexical consiste em o aprendente não dispor de certas combinatórias frequentes e mais ou menos cristalizadas; para colmatar essa deficiência, ele recorre a processos aditivos e usa palavras da L2 disponíveis para criar combinatórias aproximadas.

4. o aprendente dispõe de uma representação de um dado item cujas propriedades apresentam um desvio qualquer no que respeita o seu significado referencial, extensões metafóricas, adequação pragmática ou valores afectivos. Estes desvios, que resultam de relações paradigmáticas (sinónimos, antónimos, hipónimos...) e relações sintagmáticas das palavras, fazem-se sentir, com certeza, no seu comportamento numa frase ou enunciado, sobretudo nos usos menos prototípicos e estão relacionados, em boa parte, com as diferentes escolhas que cada língua faz para expressar sentidos idiomáticos.

Como referi antes, a marcação de desvios por que optei pressupunha um tratamento posterior. Assim, para estabelecer uma **tipologia dos desvios**, numa primeira abordagem dos dados, usei a grelha de análise do Panorama do Português Oral de Maputo (PPOM) ²⁴⁰.

Como é sabido, em Moçambique, «o Português é tipicamente uma L2, falada sobretudo em meio urbano, por locutores que têm línguas do grupo bantu como línguas maternas (L1)». O principal objectivo do PPOM consiste na «recolha de um corpus de língua oral (equivalente a cerca de 130 horas de gravações), produzido por 100 falantes da cidade de Maputo. Com este corpus, transcrito e processado em programa computarizado, pretende-se criar um banco de dados que possa ser disponibilizado para a investigação sobre o processo de dialectação do Português oral de Maputo, a realizar na segunda fase desta pesquisa (ver volume II, A Construção de um Banco de 'Erros')» (Gonçalves 1997:54) ²⁴¹. A partir de uma amostragem do material recolhido que «equivale a 12 horas de discurso gravado, e a 20 % do total dos informantes» (Moreno & Tuzine 1997:75-76), é feito um levantamento e uma **tipologia geral de 'erros' do Português Oral de Maputo**. Esta tipologia inclui sete grandes áreas: léxico, léxico-sintaxe, sintaxe, morfossintaxe, fonética e diversos (ver errata às páginas 69 e 70 do volume II).

Embora se tratasse de discurso oral e a abordagem fosse diferente, esta tipologia de erros revelou-se de grande utilidade para o meu trabalho. Ela funcionou, numa primeira fase, como ponto de partida para a

²⁴⁰ Agradeço a Perpétua Gonçalves todos os e-mails em que, gentil e prontamente, respondeu a todas as minhas dúvidas sobre esta tipologia.

²⁴¹ Até ao aparecimento deste projecto, os estudos mais sistemáticos «tiveram como base empírica o registo formal de locutores de uma faixa social específica, de adultos instruídos.» Por isso, muitos fenómenos considerados desviantes em relação ao PE estavam quase ausentes destes estudos ou tinham neles uma frequência muito baixa. Para responder a esse enviesamento, foram integrados na recolha falantes com um nível de escolaridade igual ou superior à 4ª classe, entre os 16 e os 55 anos, de cinco dos cinquenta bairros de Maputo. Desses informantes, 92 % têm como L1 uma língua bantu, 62 % aprendeu português em situação formal, e 94 % destes nunca frequentou a universidade. Para 36 % dos informantes o português é a única língua de comunicação em casa, para 34 % é exclusivamente uma língua bantu, enquanto que os restante 30 % usam ambas no mesmo contexto comunicativo. As profissões são variadas: quadros médios e superiores, empregados de escritório, trabalhadores manuais, e trabalhadores independentes, em proporções equilibradas (Gonçalves 1997:47-73).

classificação dos erros do *PESTRA*. Numa fase posterior ela foi reformulada em função da abordagem lexical seguida neste trabalho ²⁴². Digamos, portanto, que ela funcionou como uma fase intermédia entre a marcação dos desvios e a tipologia que aqui ensaio.

Assim, por exemplo, o *PPOM* considera como erros de léxico: «neologismos de forma, neologismos semânticos, empréstimo, conversão, e expressões idiomáticas». Inclui na área do léxico-sintaxe: «selecção categorial, selecção semântica, passiva, expressões quantitativas, expressões locativas, expressões temporais, expressões recíprocas, pronome pessoal reflexo, e género dos nomes». Decorrente daquilo que neste trabalho se entende por saber uma palavra, na **tipologia de erros adoptada para o *PESTRA*** foram considerados como erros de léxico, a par das “expressões idiomáticas”, as sequências desviantes (trata-se sobretudo de expressões temporais e locativas) mais ou menos cristalizadas. O género dos nomes foi igualmente considerado como um erro de léxico, assim como os casos de concordância nominal (no *PPOM* incluídos em morfossintaxe) identificados como decorrentes de uma atribuição desviante do género do nome.

Visto que saber uma palavra pressupõe conhecer as suas diferentes formas, foram também classificados como erros de léxico os desvios na flexão de nomes, adjectivos e verbos. Foram também incluídos, exigidos pela modalidade escrita, outros três tipos de erro: ortografia, acentuação, contracção e formas aproximadas. Neste último caso, trata-se de formas que não coincidem com formas do português, que parecem não resultar nem de uma deficiente aplicação de regras ortográficas da nossa variante, nem de *transfer* directo de outra língua, mas sim de vocábulos de que o aprendente tem ainda uma representação difusa. Este tipo de erros está relacionado com a aquisição do sistema fonético-fonológico do português europeu e substitui a entrada relativa a erros de “fonética” (ver *PPOM*).

²⁴² «[A] pesar da validade que possa ter a sistematização estabelecida, há que reconhecer igualmente os seus limites, ou pelo menos a possibilidade de os dados do *corpus*-amostragem serem agrupados de outra forma. [...] Esta é uma fase exploratória da investigação, em que se pretende essencialmente proporcionar um quadro indicativo dos principais tipos de desvios do POM relativamente à norma europeia. Fica, por conseguinte, a recomendação no sentido de que esta classificação e organização dos ‘erros’ seja revista, em função dos objectivos específicos de estudo que venham a ser estabelecidos.» (Gonçalves 1997:44-45).

Embora em *Léxico deficitário* nem todos os desvios sejam tratados com o mesmo grau de detalhe, porque não foram classificados como erros de léxico, ou porque, apesar de terem sido, serão analisados mais superficialmente, a todos os desvios identificados no *corpus* foi atribuída uma classificação, ainda que sujeita a reformulações posteriores. Assim, estabeleceu-se uma **tipologia de desvios** que orientou a análise dos dados e que é apresentada na página seguinte.

*Tudo isto será desenvolvido no capítulo 6. Depois de tratar, na sua primeira parte, o **Vocabulário disponível**, apresentarei, guiada por esta tipologia, a análise e o comentário dos dados na segunda parte do capítulo, em **Léxico deficitário**.*

LÉXICO DEFICITÁRIO

DESVIOS FORMAIS

ortografia
acentuação
contração

formas aproximadas

flexão de nomes e adjectivos
flexão verbal

atribuição de género
concordância em género

VOCABULÁRIO INDISPONÍVEL

empréstimos
neologismos a partir de empréstimos
neologismos formais

desvios em combinatórias

DESVIOS NA SELECÇÃO DO LÉXICO

nomes
adjectivos
verbos
preposições
outras categorias
recategorização

DESVIOS SINTÁCTICOS, MORFOSSINTÁCTICOS E OUTROS

organização sintáctica
ordem das palavras
colocação do pronome pessoal
determinação dos nomes (det+ art)
morfologia flexional:
concordância verbal
concordância nominal em género
concordância nominal em número
flexão casual

tempo verbal
modo verbal
aspecto verbal
infinitivo

outros

Capítulo 6

Vocabulário disponível e léxico deficitário

6.1. Introdução

Apesar de neste trabalho se considerar que saber uma palavra é muito mais do que a uma imagem gráfica (ou sonora) associar um significado, o que se avalia em 6.2. *Vocabulário disponível* é exclusivamente de quais e quantas representações gráficas a que é atribuído um dado significado dispõe cada um dos sub-grupos, independentemente de elas coincidirem total ou parcialmente com a forma que um falante nativo usaria no mesmo contexto. Assim, e para efeitos de contagem, todas as formas divergentes foram normalizadas, no que respeita os seus aspectos formais, e contadas como se se tratasse, na variedade culta, da forma correspondente àquela que pareceu ser a intenção do autor. O ANEXO 8 apresenta o vocabulário disponível pertencente ao PF **com maior número de ocorrências**. O ANEXO 9 apresenta o vocabulário disponível não pertencente ao PF que, em cada um dos *sub-corpora*, atingiu maior número de ocorrências. A análise e comentário dos dados relativos a vocabulário disponível será feita com base nestes resultados.

A análise apresentada em 6.3. *Léxico deficitário* foi feita a partir de uma base de dados que contém todos os desvios classificados segundo a tipologia apresentada no final do capítulo anterior.

6.2. Vocabulário disponível

6.2.1. *Vocábulos*

Como se pode ver na TABELA 1, **apenas 1454 entradas PF ocorreram mais de 0 vezes no PESTRA**. Cada um dos grupos usou 6 ou 7 dessas entradas mais de 250 vezes. Dessas 1454, **42 entradas somaram mais de 200 ocorrências** na totalidade do *corpus*, o que perfaz um total de

mais de metade (29 371) do total de 51 785 ocorrências PF em todo o corpus. Este resultado dá-nos uma primeira ideia da frequência com que ocorrem certos vocábulos, sobretudo palavras gramaticais, e do peso que têm em qualquer texto (confrontar com *TABELA 3*).

Tabela 1

ENTRADAS E OCORRÊNCIAS NO CORPUS DE VOCÁBULOS PF							
Entradas com mais de:	português	espanhol	francês	alemão	sueco	chinês	total
250 ocorrências	7	7	7	7	7	6	32
200	7	7	7	8	10	7	42
150	10	10	11	13	13	11	60
100	14	19	14	16	17	14	83
50	23	29	29	34	34	34	150
20	72	80	87	87	91	89	324
15	96	98	111	100	118	113	388
10	141	133	156	149	148	162	517
5	306	254	264	281	254	277	772
2	481	455	468	477	474	473	1044
1	641	604	652	605	617	594	1203
0	960	931	971	873	865	883	1454

Paul Nation (1993a:118) diz que um aprendente que saiba apenas a palavra *the*, a mais frequente em Inglês, poderá reconhecer 7 % dos itens de qualquer texto escrito. No nosso caso, os correspondentes *o*, *a*, *os*, *as* perfazem 9,8 % dos itens do *PESTRA*. Rivenc (1987, tomo 2:5) afirma que cinquenta vocábulos do português cobrem mais ou menos 50 % do *corpus* do *PF*. No *PESTRA*, eles perfazem um pouco mais, ou seja, 60 % do total de ocorrências. Mas para fazermos uma leitura global dos resultados, observemos a *TABELA 2*. Ela apresenta os **resultados quanto ao número total de entradas e ocorrências, PF e não PF**, em cada um dos *sub-corpora*.

O inquérito português que esteve na base do estabelecimento da lista do *PF*, como vimos antes, permitiu recolher um total de 6.000 **vocábulos diferentes**. Em relação ao número de vocábulos de que dispõe a língua portuguesa, este número é aparentemente insignificante ²⁴³. Apesar deste resultado e daquilo que a bibliografia já fazia prever, não deixa de ser

²⁴³ A 6ª edição do *Dicionário da Língua Portuguesa* da Porto Editora, um dos mais divulgados entre utilizadores comuns, ou a 1ª edição do *Dicionário Universal da Língua Portuguesa* da Texto Editora, têm entre 90 000 e 100 000 entradas. Este segundo tem exactamente 93 034 vocábulos (ver sobrecapa do volume).

surpreendente que os informantes portugueses que constituíam o **grupo de controle** tivessem conseguido escrever sobre um conjunto relativamente

Tabela 2

Entradas e ocorrências PF e não PF no PESTRA						
	portug.	espanhol	francês	alemão	sueco	chinês
TOTAL DE ENTRADAS	1895	1465	1514	1239	1177	1188
entradas PF	960	931	971	873	865	883
entradas não PF	935	534	543	366	312	305
% entradas PF	50,70 %	63,50 %	64,10 %	70,50 %	73,50 %	74,30 %
% entradas não PF	49,30 %	36,50 %	35,90 %	29,50 %	26,50 %	25,70 %
TOTAL DE OCORRÊNCIAS	11318	10946	11260	11102	11169	10635
ocorrências PF	9871	10160	10424	10480	10606	10115
ocorrências não PF	1447	786	836	622	563	520
% de ocorrências PF	87,20 %	92,70 %	92,60 %	94,40 %	95 %	95,20 %
% de ocorrências não PF	12,80 %	7,30 %	7,40 %	5,60 %	5 %	4,80 %

variado de temas usando menos de 2.000 vocábulos diferentes, dos quais aproximadamente metade pertencem ao *PF* (50,70 %). Este era, aliás, o principal objectivo da inclusão de um *sub-corpus* de controle de falantes nativos: avaliar qual a dimensão e a natureza do vocabulário usado, para a produção de textos escritos sobre os mesmos temas, por falantes nativos e compará-la com a de não nativos ²⁴⁴.

Se tivermos agora em consideração os dados relativos aos **estrangeiros**, e se lermos o quadro da esquerda para a direita, veremos que, à medida que a distância linguística vai aumentando, não só o número de vocábulos diferentes vai diminuindo, mas também a percentagem desses mesmos vocábulos que não pertencem ao *PF*. Ou seja, enquanto que os portugueses usaram 49,30 % de vocábulos diferentes não *PF*, os chineses só usaram 25,70 %. As duas últimas linhas do mesmo quadro – **% de ocorrências PF e não PF** – dão-nos uma outra informação: a tendência para o decréscimo de ocorrências não *PF* acentua-se (dos 12,80 % dos portugueses para os 4,90 % dos chineses). Isto significa que um mesmo

²⁴⁴ Voinomaa 1993:10 faz a seguinte observação: «native speakers do not always employ more words than learners.»

vocábulo não *PF* é usado mais vezes, sendo assim muito mais rentabilizado, à medida que a distância linguística vai aumentando.

Observemos agora mais de perto os resultados de cada um dos grupos e as diferenças entre eles, no que respeita a **distribuição das ocorrências PF** (TABELA 3). Também no caso dos “**vocábulos gramaticais homógrafos de alta frequência não analisados no PF**” a mesma tendência de decréscimo em função da distância linguística se mantém. Dois dos grupos (suecos e chineses) ficam abaixo da média (22,30 %). Além de possíveis diferentes opções ao nível da estrutura sintáctica seleccionada, uma explicação para estes resultados prende-se com as omissões deste tipo de vocábulos (Ver código X6: *este item não foi usado; o falante nativo tê-lo-ia usado*).

Tabela 3

Distribuição das ocorrências PF no corpus	port.	espanhol	francês	alemão	sueco	chinês	Total
voc. gram. homógrafos frequentes não analisados no PF	2775	2588	2596	2541	2440	2144	12309
100 mais frequentes vocábulos	4301	5007	4999	5216	5488	5036	25746
restantes vocábulos PF	2795	2565	2829	2723	2678	2935	13730
% voc. gram. homógrafos frequentes não analisados no PF	24,50	23,60	23,10	22,90	21,80	20,20	22,30
% dos 100 mais frequentes vocábulos	38,00	45,70	44,40	47,00	49,10	47,40	46,70
% dos restantes vocábulos PF	24,70	23,40	25,10	24,50	24,00	27,60	24,90
%	87,20	92,70	92,60	94,40	95	95,20	93,90

De facto, foram marcadas no *corpus* 49 omissões no caso dos espanhóis, 106 no dos franceses, 177 no dos alemães, 265 no dos suecos, e 372 no dos chineses. Da observação de cada uma destas listas, e como seria fácil de prever, ressalta que, quase na totalidade, o que é omitido são palavras gramaticais, sobretudo artigos e preposições ²⁴⁵.

Examinemos agora, para além das palavras gramaticais, os **100 mais frequentes vocábulos** no *PESTRA* (Ver TABELA 3). Eles representam,

²⁴⁵ As omissões serão analisadas em 6.3.3. *Desvios na selecção do léxico*.

no *sub-corpus* português, 38 % do total de ocorrências, e 46,70 % do total das ocorrências no conjunto dos outros *sub-corpora*. Além dos portugueses, só os franceses, com 44,40 %, ficam abaixo desta média. Os suecos (49,10 %), seguidos dos chineses (47,40 %), são os que mais se apoiam nestes vocábulos mais frequentes para a produção textual. Os **restantes vocábulos PF** representam 24,70 % do total das ocorrências no *sub-corpus* português, e 24,90 % no conjunto dos outros *sub-corpora*. Acima desta média ficam os franceses e os chineses (27,60 %). Este grupo, que, como vimos, usa menos vocábulos gramaticais e menos vocábulos não PF, parece concentrar mais a sua atenção em vocábulos plenos bastante frequentes.

Em conclusão, aproximadamente metade dos 2217 vocábulos que constituem o PF correspondem a mais de 90 % do total de ocorrências do corpus. Contrariamente àquilo que talvez fosse de esperar, os falantes de línguas românicas, e em particular os de espanhol, não usam na escrita um vocabulário muito mais alargado que os outros grupos. Este resultado surpreender-nos-ia, com certeza, se não tivéssemos um grupo de controle de falantes de português L1.

A literatura sobre aprendizagem de **línguas próximas** refere muitas vezes que os aprendentes, particularmente nos primeiros estádios de aquisição da L2, mostram pouca confiança no uso de **cognatos**, como estratégia para a expansão do seu vocabulário. À luz destes resultados, e tendo em consideração os do grupo de controle, este comentário pode ter uma leitura diferente. Pode, pelo menos, levar-nos a reflectir sobre a que tipo de cognatos se referem: cognatos atemáticos de alta frequência que “servem mais para nos exprimirmos a propósito das coisas”, ou cognatos temáticos pouco frequentes que “designam seres, objectos ou acções»? É bem possível que se trate destes últimos e que o facto de os terem encontrado poucas vezes no *input* explique essa falta de confiança. Ou será que, de facto, eles não fazem muita falta quando nos queremos exprimir “a propósito das coisas”, como era aqui o caso ?

A validade desta hipótese pode ser reforçada pela observação da origem dos vocábulos PF usados no *PESTRA*. Como referi antes, os 2217 vocábulos que constituem o PF foram seleccionados a partir de um inquérito de frequência e de um inquérito de disponibilidade. A *TABELA 4* apresenta resultados relativos exclusivamente às entradas do PF que no

inquérito de frequência atingiram o limiar de frequência mínima de 40. Ou seja, aquelas entradas que não dependeram nem do inquérito de disponibilidade nem da intervenção da Comissão para serem seleccionados. Como se pode verificar, eles representam, em todos os casos, mais de 70 % do total de entradas do *PF*.

Tabela 4

FREQUÊNCIA (PF)	port.	espanhol	francês	alemão	sueco	chinês
entradas só do inq. de frequência	730	700	727	671	660	673
ocorrências só do inq. de frequência	9421	9707	9870	9988	10104	9638
% do total de entradas PF	71	70,50	70,60	70,60	71,90	71,70
% do total de ocorrências PF	95,40	95,50	94,70	95,30	95,30	95,30
% do total de ocorrências	83,20	88,70	87,70	90	90,50	90,60

Em contrapartida, dos 646 vocábulos seleccionados exclusivamente a partir do **inquérito de disponibilidade**, os números são muito mais modestos, como se pode ver na *TABELA 5*.

Tabela 5

DISPONIBILIDADE (PF)	port.	espanhol	francês	alemão	sueco	chinês
entradas do inq. de disponibilidade	147	155	153	138	136	138
ocorrências do inq. de disponibilidade	270	303	354	337	307	318
% do total de entradas PF	14,20	15,60	14,90	14,80	14,80	14,70
% do total de ocorrências PF	2,70	3,00	3,40	3,20	2,90	3,10
% do total de ocorrências	2,40	2,80	3,10	3,00	2,80	3,00\

Apenas 371 desses vocábulos foram usados, perfazendo um total de 1891 ocorrências no conjunto de todos os *sub-corpora*. Desses 371 vocábulos diferentes, 132 só ocorreram uma vez. Esses 371 vocábulos representam apenas à volta de 14 % do total de entradas do *PF* e aproximadamente 3 % das ocorrências na amostra de cada um dos grupos.

Estes resultados, e o facto de 532 vocábulos dos 2217 do *PF*, na sua maioria substantivos e adjectivos, tendo uma boa parte deles origem na disponibilidade, não registarem nenhuma ocorrência no total da amostra, confirmam a afirmação de Paul Rivenc: o vocabulário obtido a partir do inquérito de disponibilidade é constituído por “palavras de uma frequência fraca e pouco estável” mas “usuais e úteis”. Mas uma questão que parece pertinente levantar é se esses vocábulos incluídos na lista do *PF* serão mais úteis do que outros que dela não fazem parte, mas que foram usados na amostra. A essa questão responde o *ANEXO 9* que, como disse antes, apresenta o vocabulário disponível não pertencente ao *PF* que, em cada um dos *sub-corpora*, atingiu maior número de ocorrências.

Quando comparado com outros, realizados anteriormente, «a maior extensão do *VPF* pretende responder a uma das críticas que têm sido feitas aos vocabulários fundamentais que é a de se apresentarem selecções vocabulares demasiado restritas. Pretendeu-se assim corresponder de forma mais eficaz às exigências das correntes actuais da didáctica das línguas estrangeiras que põem em causa uma excessiva restrição no estabelecimento dos conteúdos linguísticos, mesmo para o nível elementar da aprendizagem, e preconizam uma maior diversificação desses conteúdos de forma a melhor corresponder às exigências, naturalmente diversificadas também, dos diferentes públicos a que se destinam» (Segura da Cruz 1987:472). O que aqui se afirma é incontestável. O que se pode discutir é se, em vez de uma lista final de 2217 vocábulos, não teria sido preferível optar por uma lista que privilegiasse o vocabulário tendencialmente atemático, resultante em boa parte do inquérito de frequência, que não conteria mais de 1000 vocábulos, e de uma série de listas, por centros de interesses, em que os vocábulos estariam ordenados por ordem decrescente de frequência. É verdade que todo esse material está disponível para os professores nos volumes “Métodos e Documentos”, mas, perante a lista final, ele pode ficar perdido entre o excesso e a escassez. Excesso e escassez que se prende também com uma outra questão de base: a distinção entre dimensão de um vocabulário suficiente para a recepção e dimensão de um vocabulário suficiente para a produção. Retomarei este assunto no Capítulo 7, *Léxico e ensino de PE L2*.

Também é verdade que, para além da dimensão das amostras e para além de estarmos a comparar uma amostra escrita com uma outra que privilegia o oral, pode haver outras razões para que certos vocábulos *PF* não ocorram no *PESTRA* ou tenham nele um peso muito diferente. Uma delas é, certamente, a comparação de uma amostra de L2 com uma de L1. Olhemos mais de perto os **100 mais frequentes vocábulos no *PESTRA*, no *sub-corpus de controle* e no *PF***. Da lista dos 100 mais frequentes no *PF*, mas não no *PESTRA*, fazem parte, entre outros, os vocábulos *claro*, *pá* (bordão), *ah* e o verbo *olhar* (*olhe*, *olha*) que são, ou eram na altura da recolha (ver *pá* e *portanto*), típicos do oral. Nenhum destes, como seria de esperar, pertence aos 100 primeiros do grupo de controle. Do mesmo modo que o primeiro lugar ocupado na lista pelo verbo *ser* se deve, com certeza, aos marcadores discursivos *não é* e *não é verdade* (podendo dizer-se o mesmo do lugar mais cimeiro ocupado pelo vocábulo *não*).

Quando observados por ordem decrescente de ocorrências e se comparamos o seu número de ordem com o mesmo no *PF*, verificamos que no caso de certos vocábulos essa ordem é muito diferente. Nomes como *peessoa(s)*, *vida*, *amigo*, *cidade*, *dinheiro*, *problema*, *carro*, *lugar*, *rua*, *mundo*, *trabalho* e *português*, adjectivos como *diferente*, *importante*, e *grande* e verbos como *esperar*, *falar*, *pensar* e *comer*, têm comparativamente muito mais ocorrências no *PESTRA*, mas em muitos casos também no *sub-corpus* português, do que no *PF*. Isso pode dever-se aos temas, aos interesses do grupo etário, predominantemente jovem, ou à situação de estrangeiro pronto a viver um mundo diferente e a compará-lo com o seu. As cartas dirigidas a amigos explicam certamente a presença entre as 100 primeiras de *tu*, *te* e *teu*, bem como a ausência de *senhor*. O lugar cimeiro que ocupam os possessivos *meu*, *minha*, *tua*, *sua* e *seu* na amostra dos estrangeiros aparece como sintoma de domínio insuficiente da expressão da posse em PE, que recorre menos aos possessivos do que outras línguas.

Interessante, porque revela com certeza **características da interlíngua**, é a comparação entre os dados do *PESTRA* e os das duas amostras de L1. Se observarmos os vocábulos que são comuns ao *PF* e ao grupo de controle, mas que não fazem parte do *PESTRA*, verificamos que uma boa parte deles estão relacionados com a **expressão das relações espaciais**. Fazem parte dos 100 mais frequentes no *PESTRA* o verbo *ir*, mas

não *vir*, os deícticos *aqui*, *este* e *esta*, que exprimem proximidade em relação ao locutor, mas não *esse* e *essa*. Responsável pela entrada de *isso* (com 152 ocorrências) são as locuções *além disso* e *por isso* ²⁴⁶. A TABELA 6 apresenta os resultados dos deícticos no *PESTRA* e no grupo de controle.

Tabela 6

	port.	espanhol	francês	alemão	sueco	chinês	total
aqui	3	9	14	24	35	8	90
aí	11	2	0	0	0	0	2
ali	0	8	2	1	4	6	21
cá	4	3	3	4	5	8	23
lá	13	7	5	7	15	18	52
este	36	22	45	36	42	34	179
esta	44	13	30	25	22	28	118
isto	14	21	7	1	22	8	59
esse	14	12	12	7	6	4	41
essa	22	5	14	8	3	1	31
isso	11	14	21	10	9	11	65
aquele	10	6	2	2	4	15	29
aquela	10	9	8	2	5	5	29
aquilo	3	2	0	0	0	0	2
	195	133	163	127	172	146	741

Como se pode verificar, vistos no seu conjunto, e se não tivermos em consideração os desvios, os resultados até nem se afastam muito dos resultados de L1. Como é sabido, algumas línguas românicas, entre elas o português e o espanhol, mantiveram os três lugares do Latim. Mas esse carácter mais marcado do sistema parece, neste estágio de aquisição, não constituir já uma grande dificuldade, pelo menos no que respeita os demonstrativos. No entanto, todos os grupos apresentam desvios no uso dos marcadores espaciais *aqui* (em muitos dos casos presentes no *corpus*, o falante nativo preferiria *cá*) e *ali* (em todos os casos desviantes – 12 em 21 – usado em vez de *lá*; o grupo espanhol é responsável por apenas um destes

²⁴⁶ Na Tabela 6, não são consideradas as ocorrências de *isso* nestas locuções.

desvios)²⁴⁷. Isto pode levar-nos a pôr como hipótese que o sistema dos demonstrativos fica estabelecido na interlíngua de acordo com a variedade-alvo antes do destes marcadores espaciais²⁴⁸.

Uma análise mais detalhada destes dados, bem como dos verbos *ir-vir*, *ser-estar-ficar*, *andar*, *deixar*, *passar*, *pôr*, *tornar* e *chegar* e das preposições *a-de-em-para-por-sobre*, todos eles vocábulos de alta frequência, seria desejável e levar-nos-ia certamente a conclusões muito interessantes sobre a aquisição da expressão do espaço (e da sua extensão ao tempo). No entanto, ela constituiria, só por si, uma dessas investigações mais finas de que falei antes e a que este trabalho pode conduzir, mas que não desenvolverei aqui. Batoréo 2000 é um estudo sobre a expressão do espaço em crianças falantes de PE, enquanto língua materna, que reúne vasta informação linguística e cognitiva sobre este tema, de grande utilidade também para a compreensão do mesmo fenómeno por aprendentes de L2. O que me parece de destacar, desde já, é o facto de uma simples contagem de ocorrências de vocábulos nos poder conduzir a possíveis diferenças num aspecto tão fundamental como, por exemplo, a conceptualização e a expressão linguística do espaço²⁴⁹.

6.2.2. *Combinatórias*

Tempo e espaço são, de facto, categorias básicas. Essa a razão pela qual os vocábulo que as exprimem surgem entre os mais frequentes nas

²⁴⁷ Sobre a distinção entre *aqui - ali* e *cá - lá*, Batoréo 2000:439 cita Silva-Joaquim 1988: «Trata-se da manifestação de uma diferente granularidade, que para *cá/lá* é vasta, extensa e para *aqui/ali* se apresenta fina, estreita quase a tender para o ponto. As manifestações de diferentes granularidades e as possibilidades discursivas que daí advêm são frequentes na construção textual, mas não tanto nos sistemas linguísticos.»

²⁴⁸ «Where L1 has an **unmarked** and L2 a corresponding **marked category**, the unmarked category will often be transferred from L1 to L2, and this unmarked category will in many cases remain as a feature of the learner's interlanguage for a long period of time. [...] Where both L1 and L2 have a marked category, the unmarked category can still turn up in the learner's interlanguage. It will, however, fairly quickly be abandoned in favor of the marked category.» (Hyltenstam 1986:68)

²⁴⁹ «[A] expressão básica do Espaço no Português Europeu é assegurada por um número muito reduzido de itens. Tomando em consideração o vocabulário do *Português Fundamental* – isto é, o que tem a frequência superior a quarenta ocorrências – trata-se de uma centena (cento e quarenta) marcadores espaciais nucleares, o que constitui cerca de cinco por cento do vocabulário global do Português Fundamental. [...] Repare-se que a constatação de que um número muito reduzido de itens linguísticos é responsável por uma parte substancial da expressão linguística de um idioma – e, neste caso específico, da expressão espacial entendida no sentido lato da Teoria Localista – é um argumento que vem ao encontro dos pressupostos talmianos da existência de um esqueleto espacial que estrutura todo o funcionamento da linguagem.» (Batoréo 2000:448-449).

línguas. Mas estes vocábulos básicos não são suficientes para fazerem referência a uma multiplicidade de situações espaciais e temporais; daí que muitos deles devam também a sua alta frequência ao facto de se combinarem entre si ou com outros em **sequências cristalizadas** de conteúdo semântico mais ou menos transparente. É o caso dos substantivos de alta frequência *vez, tempo, dia, ano, hora*, e de outros ligeiramente menos frequentes, tais como *mês, semana, manhã, tarde, noite, fim*, que se combinam com verbos (como *haver* e *fazer*), preposições (em especial, com *a, em* e *de*), e com advérbios (*já, agora, depois, antes, sempre, aqui, quase, ainda, hoje, então...*) para formarem um complexo sistema de expressão temporal.

O ANEXO 10 dá conta de muitas das locuções que, tendo atingido a frequência mínima de 40 ocorrências no PF, foram usadas no PESTRA e também pelo grupo de controle. Das 95 locuções que no PF atingiram a frequência de 40, os portugueses usaram mais ou menos metade (49). No corpus estrangeiro foram identificadas 62 diferentes, tendo cada um dos grupos usado aproximadamente o mesmo número, ou seja, entre 34, no caso dos franceses, e 40, no caso dos espanhóis. Entre as mais usadas estão *por isso, por exemplo, mais ou menos, além disso, pelo menos, depois de, antes de, por causa de, e muitas vezes*.

Mas, para além destas, outras do mesmo e de outros tipos, relativamente frequentes, foram usadas. O ANEXO 11 é um extracto do levantamento sistemático de diferentes tipos de sequências feito a partir do sub-corpus ALEMÃO. Essas sequências, ou combinatórias, são, para além de vocábulos compostos (*bode-expiatório, parque de estacionamento...*) grupos completamente cristalizados (*apesar de, cerca de, ou melhor, se calhar...*), cabeças de frase (*acho que, penso que, quer dizer*) e certas coocorrências privilegiadas (*animais domésticos, crise económica, bairro antigo*). Também fazem parte da lista certas sequências que constituem fórmulas frequentes no oral (*É ótimo!, Até à próxima!*) e fórmulas que iniciam ou fecham cartas (*Querida amiga, Muito obrigado pela tua carta, Um abraço*).

O modo habitual e adequado de estabelecer o que é uma combinatória frequente consiste em seleccionar uma palavra nó e, sobre um

corpus de grande extensão, observar todas as suas concordâncias ²⁵⁰. Aqui, dada a pequenez relativa de cada *sub-corpus*, o procedimento foi diferente. Três das amostras foram percorridas e foram seleccionadas todas as sequências que apresentavam um certo grau de fixidez e possível recorrência. Quero com isto dizer que o critério pode ter variado ligeiramente caso a caso e pode ser que algumas delas tenham sido indevidamente seleccionadas. Feito este reparo, posso dizer que foram levantadas aproximadamente 1100 do *sub-corpus* português, 950 do espanhol e 800 do alemão. Este decréscimo em função da distância linguística, aliado ao facto de nas amostras estrangeiras, em particular, e como se pode ver, no caso dos alemães, a mesma sequência aparecer muito mais vezes repetida (ver *por causa de, por exemplo, por isso, acho que, espero que, gosto muito de, penso que*) parece confirmar a hipótese segundo a qual «an important index of nativelike competence is that the learner uses idioms fluently» (Ellis 1997:130). Ou seja, à medida que a interlíngua se vai aproximando da variedade-alvo as combinatórias frequentes vão sendo usadas em maior número e cada vez mais automaticamente. Do mesmo modo que vão desaparecendo da interlíngua eventuais usos de sequências não analisadas (ver C45 *uma vez esperei o eléctrico há quase uma hora por quase uma hora*).

6.2.3. Conclusão

Como é sabido, muitos vocábulos, embora tendo uma categoria sintáctica de base, podem, quando em determinados contextos, desempenhar funções de outra (nomes que podem ser adjectivos ou advérbios, adjectivos que podem ser nomes...). Daí que, resultados mais rigorosos quanto a ocorrências de cada uma das categorias exijam uma anotação de todas as formas do *corpus*, tendo em consideração a sua categoria num determinado contexto, o que não foi feito. No entanto, e apesar de os resultados apresentados a seguir deverem ser encarados como provisórios e aproximados, a observação da amostra centrada em algumas das **classes de palavras** também é reveladora das características da

²⁵⁰ «How do we determine whether a phrase is really a lexical unit or simply a phrase made up of separate words? At the end of the continuum we might have free collocation and at the other end an unbreakable set.» (Singleton 1999:199).

interlíngua. A TABELA 7 apresenta o total de entradas apuradas de nomes, adjectivos, verbos, advérbios e locuções adverbiais, e conectores frásicos presentes no PESTRA e no grupo de controle.

Tabela 7

CLASSES DE PALAVRAS	port.	espanhol	francês	alemão	sueco	chinês
nomes	804	670	697	588	535	536
% de nomes PF	48	60	61	66,2	70	71,6
adjectivos	307	201	201	175	168	135
% de adjectivos PF	39,1	48,8	49,8	55,4	60,1	65,2
nomes/adject	43	51	65	28	41	46
% de nom/adj PF	37,2	47,1	50,8	53,6	48,8	54,3
verbos	437	293	300	235	231	255
% de verbos PF	49,4	68,6	65,3	77	79,2	75
adv. e locuções adv.	166	128	90	98	88	77
% de adv. e locuções adv. PF	57,2	73,4	70	83,7	86,4	80,5
conectores frásicos	53	33	29	23	22	25
% de conectores frásicos PF	43,4	66,7	82,8	95,7	91	92

Como se pode ver, a diferença entre o número de entradas *PF* no grupo de controle e no PESTRA vai aumentando progressivamente dos nomes para os verbos. A amplitude, entre o grupo de controle e o grupo com resultados mais afastados deste, é de 23,6 % no caso dos nomes (em relação aos chineses), 26 % no caso dos adjectivos (também em relação aos chineses), e atinge os 30 % no caso dos verbos (em relação aos suecos). **Estes resultados apontam para uma maior dificuldade no alargamento da categoria verbo**, o que não surpreende dada a centralidade e complexidade desta categoria. **Mas mais revelador ainda é a grande diferença quanto a advérbios e, muito em especial, quanto a locuções adverbiais e conectores frásicos usados pelo grupo de controle e pelos aprendentes estrangeiros. Estes parecem enfrentar grandes dificuldades, particularmente na aquisição de locuções conjuncionais.** De facto, estas restringem-se a três ou quatro casos altamente frequentes (*para que, mesmo que*), enquanto que os portugueses usam uma enorme

variedade deste tipo de sequências cristalizadas (*dado que, se bem que, uma vez que...*). Se a estes resultados juntarmos alguns indicadores relativos à expressão anafórica e à co-referência,²⁵¹ teremos uma panorâmica relativamente esclarecedora do vocabulário disponível e do seu uso.

Podemos concluir que a frequência dos vocábulos no input é um factor determinante na aquisição. Por outro lado, e os resultados são bastante consistentes neste ponto: à medida que a distância linguística vai aumentando, os aprendentes vão-se restringindo mais ao vocabulário mais frequente. Acontece, no entanto, que, por vezes, os franceses parecem dispor de recursos ligeiramente superiores aos espanhóis. Isto pode dever-se ao facto de haver entre os franceses um certo número de professores de línguas. Alguns eram professores de francês de filhos de imigrantes portugueses e, por isso, decidiram aprender português. É provável que competências associadas ao ensino de uma língua, visíveis sobretudo quando se trata de produção que pode ser monitorizada, desenvolvam estratégias que conduzam a uma aprendizagem mais eficaz. Também os chineses, em certos aspectos, evidenciam um desempenho menos afastado dos outros grupos do que os suecos. No entanto, por um lado, não nos podemos esquecer de que o seu tempo de aprendizagem é muito mais longo e, por outro, que o que temos estado a avaliar é, em certa medida, a capacidade para associar a um dado significativo *um* significado, o que envolve bastante conhecimento explícito adquirido graças à memória. De como é que este vocabulário disponível se comporta, e de quais as suas características enquanto componente do léxico, trataremos em seguida.

6.3. *Léxico deficitário*

O que ressalta da avaliação do vocabulário disponível é que, quando comparado com o do grupo de controle, o número de vocábulos de que os estrangeiros dispõem parece ser quase suficiente para a produção escrita de textos de temáticas pouco específicas como estes. No entanto, esse vocabulário disponível apresenta um certo número de desvios de naturezas muito diversas. Ou seja, além de nada se saber quanto a extensões de significado de que dispõe o falante para cada um desses itens, também nada foi avaliado no que respeita todos os outros aspectos relacionados com o conhecimento de uma palavra.

²⁵¹ O grupo de controle usa 38 formas de clítico, os espanhóis 25, os franceses 10, os alemães 15, os suecos 15 e os chineses apenas 1.

A **TABELA 8** dá conta do **total absoluto de desvios identificados no PESTRA**. É, portanto, em relação a estes totais que as percentagens, para cada tipo de desvio analisado e comentado em seguida, foram calculadas.

Tabela 8

Total absoluto de desvios no <i>PESTRA</i>					
espanhol	francês	alemão	sueco	chinês	total
1325	1422	1770	2061	2068	8646

Como se pode verificar, foram anotados e classificados um total de 8.646 desvios no **PESTRA**. O número de desvios, em cada um dos *sub-corpora*, vai de 1325 a 2068 o que não representa uma amplitude tão grande como se poderia esperar. Mas vejamos como é que eles se distribuem pelas diferentes tipos de desvios estabelecidos na tipologia (ver, no final do Capítulo 5, a *tipologia de desvios*).

6.3.1. *Desvios formais em vocabulário disponível*

Como referi antes, segundo Laufer (1997:141), saber uma palavra implica, entre outras coisas, saber **(1)** a sua forma oral e escrita e **(2)** a sua estrutura de base, as derivações mais comuns e a sua flexão. A **TABELA 8** inclui todos estes desvios referidos em **(1)** e **(2)**, ou seja, aqueles que considerámos como desvios formais em vocábulos disponíveis.

A observação rápida desta tabela alerta-nos para diferentes comportamentos em função da natureza dos desvios. Fica imediatamente claro que mais de metade dos 1325 desvios advertidos no *sub-corpus* espanhol são desvios formais e muitos deles resultam certamente daquilo a que os contrastivistas chamariam *interferência negativa*. Esse tipo de transferência vai diminuindo à medida que a distância linguística vai aumentando, ou seja, à medida que o léxico das duas línguas tem características mais distintas e se encontra armazenado mais em separado. Mas, como é evidente, as consequências dessa distinção entre os dois léxicos far-se-ão sentir nos aspectos não formais, nomeadamente na representação semântica e nas suas consequências a nível morfossintático.

Tabela 9

DESVIOS FORMAIS EM VOCÁBULOS DISPONÍVEIS					
	espanhol	francês	alemão	sueco	chinês
ortografia	130	92	84	93	36
acentuação	372	308	248	247	80
contração	16	9	2	7	1
%	39,1	28,8	18,9	16,8	5,7
formas aproximadas	122	111	127	92	205
%	9,2	7,8	7,2	4,5	9,9
flexão de nomes e adjectivos	9	10	7	1	4
flexão verbal	50	35	31	54	42
%	4,5	3,2	2,2	2,7	2,2
género	8	25	54	78	26
concordância em género	6	0	44	52	26
%	1,1	1,8	5,6	6,3	2,6
total	713	590	597	624	420
%	53,9	41,6	33,9	30,3	20,4

6.3.1.1. Ortografia e acentuação

Foram classificados como erros de ortografia e acentuação aqueles que claramente resultam de uma não conformidade com as regras vigentes para o português europeu ²⁵². Como se pode ver na TABELA 9, o número de erros diminui em função da distância linguística, atingindo uma soma baixíssima no caso dos chineses. No caso dos grupos cuja L1 também usa uma escrita alfabética, muitos deles têm origem em *transfer* das regras de ortografia de outra língua ou resultam de deficiente domínio de regras do português. São exemplo de **erros de ortografia**: *accidente, diferente, coppos, controllar, sympatico, característica, antigua, ficemos, chegei, Suécia, Franca, Olanda, viagem, inpotência, emgarafamento, esemplo, certesa, sítiu, prufissão, homems, e por tanto*. No entanto, aproximadamente 50 dos 400 erros de ortografia concentra-se em formas dos verbos *chegar, conhecer, ficar, poder, e ter*, de que são exemplo *chegei, conheco, ficei, pudia,*

²⁵² A ortografia oficial do PE rege-se pela Convenção Ortográfica Luso-Brasileira (Lisboa, Imprensa Nacional, 1945, 56 pgs.), o chamado Acordo de 1945, vigente em Portugal mas não no Brasil. Acrescenta-se a este Acordo o artigo único do decreto-lei n.º 32/73 que elimina «os acentos circunflexos e os acentos graves com que se assinalam as sílabas subtónicas dos vocábulos derivados com sufixo *mente* e com os sufixos iniciados por *z*.» Sobre a ortografia do português ver Castro, Duarte & Leiria 1987.

e *teem*. Quase outros tantos em formas de *acidente*, *diferente*, *português*, *quando*, e *viagem*. Também foram detectados alguns erros resultantes da não **contracção** de preposição e artigo ou determinante (ex: *em um*, *em aquele*) ou da não análise de uma forma já contraída (ex: *no esse momento*).

Quanto a **erros de acentuação**, uma boa parte consiste na falta do acento gráfico; de outros são exemplo: *atrásada*, *agradável*, *inflência*, *você*, *razão*, *atenção*, *professor*, *patrimônio* e *rapidamente*. Mas, como pode ver no ANEXO 12, mais de metade dos 1255 erros de acentuação (639) concentra-se também num reduzido número de formas muito frequentes (formas dos verbos *ser*, *estar*, e *ter*; e, entre outros, em *também*, *próximo*, *mas*, *ninguém*, e *já*)²⁵³. Da observação atenta das formas divergentes não ressalta nem o conhecimento explícito de umas regras de acentuação e não de outras nem qualquer estratégia sistemática geral ou particular. Os aprendentes parecem acentuar graficamente uma forma porque têm dela uma imagem que inclui um acento gráfico, independentemente da sílaba tónica e, por conseguinte, de onde ele deve ser colocado.

As palavras também estão representadas na memória a longo prazo pela sua imagem gráfica, quando se trata de sujeitos escolarizados. Por isso mesmo, quando temos dúvidas acerca da ortografia de uma palavra da nossa L1, geralmente basta-nos escrevê-la, ou seja, confrontá-la com a imagem que temos dela, para resolvermos a nossa incerteza. Aqui há tempos, podiam ver-se nas ruas de Lisboa uns cartazes em que se lia, depois de uns segundos de alguma estranheza: “*dabliu dabliu dabliu*” e “*imeile*”. A estranheza e a dificuldade de leitura que experimentávamos perante estas formas é semelhante à que sentimos quando vemos numa língua estrangeira uma palavra que corresponde a uma da nossa L1 mas com uma ortografia diferente. Hummel 1986 fala de níveis de processamento e diz que a codificação semântica é processada a um nível mais profundo, enquanto que aspectos estruturais e perceptuais são

²⁵³ O suplemento de educação do jornal *El País*, de 06.11.00, publica um artigo, subordinado ao título «Los alumnos de ESO son incapaces de escribir 25 palabras sin cometer faltas». O artigo refere um estudo realizado com 1.400 jovens de 13 anos em que se conclui que «dos de cada tres fallos son de acentuación». Da comparação de três estudos (realizados em 1950, 1989 e 1999) conclui-se que «hay 14 palabras que aparecen siempre entre las 25 que más faltas originan. Esas palabras son: *más, también, día, porque, mí, está, después, a, están, así, allí, había, há y cómo.*»

processados a um nível mais superficial. O processamento destas formas, porque não automatizado, exigia de nós um esforço suplementar.

Esta superficialidade no processamento e a representação gráfica forte que temos da palavra na L1, ou em outra próxima da L2, podem explicar o muito maior número de erros de ortografia do **grupo espanhol**. Uns parecem resultar da estratégia: em português é diferente (*esperimentar, orarios*); outros: em português é igual (*en geral, digote, reuniase, verte*). Há também um tipo de erros exclusivos deste grupo e que são semelhantes aos cometidos pelos portugueses: *falavasse* por *falava-se* e *sair-mos* por *sairmos*. Isto deve-se provavelmente também ao facto de não usarem conhecimento explícito, mas sim, como acontece com os falantes nativos, conhecimento implícito e fortemente automatizado para a produção destas formas ²⁵⁴.

Em contrapartida, o **grupo chinês** apresenta um número muito baixo de erros de ortografia. E, em quase todos os casos, resultam de características específicas da ortografia do português (*fiho* por *filho*, *tihamos* por *tinhamos*, *conlhecido* por *conhecido*; *ficuei* por *fiquei* e *luguar* por *lugar*). Como vimos no Capítulo 2, se, por um lado, o facto de a sua língua usar um sistema de escrita não alfabético constitui uma desvantagem no processamento lexical, por outro, o sistema de escrita logográfico permite-lhes desenvolver extraordinariamente a memória visual. Assim, é provável que também o processamento das formas da L2 se faça a um nível mais profundo, tal como acontece para a L1. O mesmo se pode dizer em relação à acentuação. O hábito de na escrita da sua língua prestarem atenção a muito pequenas diferenças dos caracteres que são significativas, aliado a um menor grau de automatismo e ao uso de mais conhecimento explícito, explica com certeza o incomparável menor número de erros de acentuação. São bom exemplo do que acabo de dizer as formas *é*, *nós*, e *está*, muitas vezes confundidas pelos outros grupos com as formas não acentuadas *e*, *nos*, e *esta*, mas raramente pelos chineses.

²⁵⁴ Como é fácil de prever, há neste grupo um número de erros de difícil classificação, em relação aos quais ficam algumas dúvidas. Destaca-se o caso frequentíssimo de *mais* por *mas* e vice-versa, que não é certamente um desvio semântico, e que foram incluídos em neologismos formais.

Prestei particular atenção a estes dois grupos por eles se encontrarem nos extremos do *continuum* no que respeita a distância linguística. Os 39,1 % de desvios dos espanhóis e os 5,7 % dos chineses, bem como os valores intermédios dos outros grupos (28,8 dos franceses; 18,9 dos alemães; e 16,8 dos suecos), são a expressão clara do peso que, também para escrita, têm as características da L1 e a sua distância e relação com a L2. Também a nível da ortografia, quanto mais afastada é a L1 menos ela interfere na L2. Além disso, e de modo mais evidente do que nos outros casos, os chineses parecem transferir estratégias desenvolvidas para a escrita da sua L1, e, por isso, parecem estar mais atentos aos detalhes e usar bastante memória visual.

6.3.1.2. *Formas aproximadas (ou o que nos diz a escrita sobre a aquisição da fonologia)*

Além das formas divergentes que resultam da não aplicação das regras ortográficas, há outras formas que não coincidem com as do PE. Elas parecem não resultar nem da não aplicação de regras ortográficas da nossa variante, nem de transfer directo de itens lexicais de outra língua, mas sim de vocábulos de que o aprendente tem ainda uma representação deficiente. Essas formas deficientes, a que chamei formas aproximadas, reflectem as estruturas fonético-fonológicas da interlíngua.

Num artigo de 1986, Meara e Ingle afirmam que a investigação em L2 se tem concentrado nos aspectos semânticos e nas combinatórias e tem prestado muito pouca atenção aos aspectos formais, considerando-os, embora de forma não explicitamente assumida, como simples e não problemáticos. No entanto, referem: “There are, for example, a number of collections of spelling mistakes made by learners which clearly point to learners experiencing severe difficulty with written forms, even in languages whose orthography is more rational than we are accustomed to with English.” (Meara & Ingle 1986:160). Perante este estado de coisas, estudam material escrito de falantes de inglês a aprenderem francês, de modo a analisarem como aprendem a representar os aspectos fonológicos de palavras recentemente adquiridas. Estabelecem como hipótese que “certas partes das palavras” (“certain parts of words”) serão mais fáceis de aprender que outras, e concluem que, de facto, o início das palavras é relativamente resistente ao erro, enquanto que as consoantes seguintes, incluindo as finais, são menos. Concluem ainda que existe uma certa instabilidade, que palavras correctas podem apresentar incorrecções em testes realizados posteriormente (Meara & Ingle 1986:169). Mas, considerar sons consonânticos situados em “certas partes das palavras” sem ter em conta quais as suas características, parece-me insuficiente. Contudo, e este trabalho é de 1986, não encontrei na bibliografia de L2 trabalhos que retomassem

esta questão. Porque existe evidência de que a aquisição da representação das estruturas fonológicas, de facto, não é simples, é preciso encontrar abordagens mais consistentes para estudar a sua aquisição.

A dissertação de doutoramento de Maria João Freitas (Freitas 1997), Aquisição da estrutura silábica do Português Europeu, realizada no quadro teórico da Gramática de Princípios e de Parâmetros, observa «o comportamento verbal de crianças monolíngues no percurso de aquisição do Português Europeu, no sentido de descrever o processo de aquisição da estrutura silábica neste sistema linguístico» e tem como um dos objectivos parciais «definir a ordem de aquisição dos constituintes silábicos no Português Europeu» (Freitas 1997:9).

A autora, por «nem todas as crianças demonstrarem o mesmo comportamento verbal no mesmo momento do seu desenvolvimento», adopta uma escala de desenvolvimento de base linguística e não de base etária. E comenta: «a partir de escalas de base linguística que definam ordens de aquisição da estrutura silábica das línguas particulares, é possível chegar a uma única escala de desenvolvimento silábico, a integrar num quadro mais geral de descrição do desenvolvimento fonológico» (Freitas 1997:151). Esta opção por uma escala de base linguística dá-nos a oportunidade de tomar este trabalho como modelo e de confrontar o resultado da análise dos dados de Freitas 1997 com os dados do *PESTRA*. No entanto, há que ter em consideração que, neste caso, o material em observação é escrito e não oral o que introduzirá inevitavelmente factores de risco e incertezas quanto a eventuais conclusões. De qualquer modo, algumas tendências observadas apontam para a existência de uma ordem de aquisição com muitos pontos de contacto com a proposta por Freitas 1997 para o português europeu L1.

Tratamento quantitativo das várias **estruturas silábicas**, realizado sobre *corpora* de PE, permitiu concluir que 90 % das ocorrências registadas correspondem às estruturas CV²⁵⁵ (52,39 %), CVC (12,39 %), V (17,32 %), e CVG (7,54 %). Os restantes 10 % são maioritariamente preenchidos pelas estruturas VC (3,44 %), CCV (2,94 %), e VG (1,66 %) (d' Andrade e Viana 1993, citado por Freitas 1997:112).

Se adoptarmos, como fez Freitas 1997, um modelo de representação da sílaba de *ataque – rima*, reconheceremos que o PE tem dois tipos de

²⁵⁵ C – consoante; V – vogal; G – semi-vogal.

ataque silábico: **simples** (como em **CV**, **CVC**, **CVG**), e **ramificado** ou complexo (como em **CCV**); o ataque simples pode também ser **vazio** (como em **ØV**, **ØVG**). Todas as consoantes fonológicas do PE podem estar associadas a ataque simples (oclusivas, nasais, fricativas e líquidas). Quanto aos ataques ramificados, os grupos consonânticos mais frequentes são: oclusiva + vibrante /r/, oclusiva + lateral /l/, fricativa + vibrante /r/, e fricativa + lateral /l/ (respectivamente, como em [pr]ato, [bl]usa, li[vr]o, e [fl]or)²⁵⁶. O constituinte silábico **rima** ou está apenas associado a um **núcleo** (não ramificado, como em **CV**; ou ramificado, como em **CVG**) ou ramifica em **núcleo e coda** (como em **CVC** e **CVGC**). O núcleo não ramificado pode estar associado a qualquer vogal, enquanto que o núcleo ramificado tem sempre uma vogal associada à primeira posição e uma semi-vogal associada à segunda (como em **faixa**, **pauta**, **cão**, ou **mãe**). Na posição de coda apenas /l, r, s/ são fonologicamente possíveis (como em **testa**, **sorte**, **salto**; **fausto**, **mães**, **tostões**).

Seguindo trabalhos de Fikkert (1994) para o Neerlandês, Freitas estipula **estádios e sub-estádios de aquisição do ataque não ramificado**: 1.º - ataques obrigatórios associados a oclusivas; 2.º - ataques vazios; 3.º - outros tipos de ataques simples: 3.ºa - nasais em ataques simples; 3.ºb - outros modos de articulação em ataques simples (fricativas, líquidas e semivogais, “em ordem variável entre si, consoante o sujeito considerado”) (Freitas 1997:152). E **estádios de aquisição do ataque ramificado**: 0. - pelo uso de estratégias de selecção, os grupos consonânticos não surgem como alvos possíveis; 1.º - o grupo consonântico é reduzido ao primeiro elemento, uma oclusiva ou uma fricativa; 2.º - o grupo consonântico é reduzido ao segundo elemento, uma líquida, uma consoante nasal ou uma semivogal; 3.º - os dois elementos do grupo consonântico são produzidos.

Há, no entanto, certas estruturas tratadas na literatura como sendo problemáticas em PE e que apontam para a questão das **fronteiras**

²⁵⁶ Freitas (1997:108-9) refere a existência de **outros grupos consonânticos** menos frequentes, como em [pn]eu, [ps]icologia, a[bs]urdo, o[bt]er, a[ft]a, a[mn]ésia. Segundo Mateus 1994 (citada em Freitas 1997:108-9), a especificidade destas estruturas é validada pela inserção de uma vogal epentética pelas crianças e pelos falantes de Português do Brasil e pelos erros de ortografia dos falantes de PE (*peneu*, *afeta*). No PE coloquial são gerados vários grupos consonânticos fonéticos, como em [sp]arar ‘separar’, [[p]aço ‘espaço’, [dvd]edor ‘devedor’, e outros.

silábicas. Freitas (1997:284) refere três: estruturas fonológicas com **ataque vazio em início de palavra**, como ‘estar’ ou ‘escrever’, palavras que são produzidas pelos falantes com apagamento da vogal átona inicial ²⁵⁷; estruturas com **semivogais** no nível fonético, muito em especial as relativas aos ditongos crescentes ²⁵⁸; e «segmentos ou sequências de **segmentos nasais** cuja produção implica a definição de fronteiras silábicas à difusão da nasalidade».

Quanto à **nasalidade**, Freitas 1997 faz a revisão das várias análises possíveis da nasalidade nas línguas, considerando que ela «não parece ser uma questão decorrente da avaliação da estrutura silábica» e que «há a possibilidade de o auto-segmento nasal ser um traço de palavra» (Freitas 1997:190 e 345). Porque há sistemas linguísticos sem segmentos [-consonântico] nasais, Freitas 1997 põe a hipótese de «a nasalidade associada a vogais e a semivogais não se encontrar disponível no sistema das crianças no início da produção». Contudo, «a produtividade do processo de nasalização dos segmentos [-consonânticos] no Português Europeu deverá fazer com que a emergência e a estabilização do funcionamento da nasalidade não ocorra tarde na aquisição» (Freitas 1997:346).

Vejamos agora o que é que **os dados do PESTRA**, que, recorde, são **dados da escrita**, nos podem dizer acerca da aquisição das estruturas fonológicas do PE. Como seria de esperar, nem sempre foi fácil classificar um desvio como aproximação ao sistema ou como erro de ortografia. É claro que, também por isso, os dados disponíveis não nos permitem ir tão longe como permitiriam se se tratasse de um *corpus* longitudinal de oralidade. É possível que, uma vez realizada esta primeira fase da investigação, os vários documentos que constituem cada um dos *sub-corpora* possam vir a ser sequencializados, em função do número e da natureza destes e de outros desvios, possibilitando então que se formulem hipóteses mais consistentes

²⁵⁷ A predominância da forma [ˈtar] no português coloquial «argumenta a favor da independência de [j] relativamente à(s) consoante(s) que o seguem.» As crianças inicialmente passam por um longo período de [ˈtar], depois começam a seleccionar também como alvo lexical *estar*. (Freitas 1997:290-291) O mesmo sucede com [ˈpɛrɔ] / [ˈjɛrɔ] (Freitas, comunicação pessoal).

²⁵⁸ «A análise fonológica defendida para o Português Europeu é a de que não existem ditongos crescentes no nível fonológico» (Freitas 1997:330). A autora analisa os dados das crianças procurando confirmar ou infirmar esta teoria, mas, segundo a autora, eles não são conclusivos (Freitas 1997:332).

quanto a estádios de aquisição, nomeadamente no que respeita a hierarquia de aquisição das informações fonológica e fonética.

Apesar da qualidade dos dados que trata, Freitas (1997:369) põe como hipótese sujeita a confirmação em investigação futura que «primeiro, as crianças trabalham no sentido da estabilização das estruturas fonológicas» e «mais tarde, as crianças lidam com informação relativa à estrutura fonética do sistema-alvo». Particularmente no caso da L2, tendo em conta as características do PE, a relação entre os níveis fonológico e fonético, sobretudo no que respeita as vogais átonas, é bastante complexa e, conseqüentemente, muito mais difícil de observar em material escrito em que outros factores interferem. É essa a principal razão por que centrarei a análise que se segue nos **segmentos consonânticos** e na **nasalidade**.

A **TABELA 10** dá conta da totalidade dos desvios considerados como sendo de natureza fonético-fonológica e distingue os desvios relativos a posições associadas a vogais e semivogais dos desvios relativos a consoantes e à nasalidade. O que salta imediatamente à vista, ao contrário do que acontece com a ortografia, é a semelhança entre os resultados dos dois grupos mais afastados no *continuum* de distância linguística (9,2 % e 9,9 %) e o decréscimo nos três grupos intermédios.

Tabela 10

FORMAS APROXIMADAS	espanhol	francês	alemão	sueco	chinês
desvios no núcleo	62	88	63	63	115
desvios no ataque, na coda e nasalidade	60	33	64	29	90
total	122	111	127	92	205
%	9,2	7,8	7,2	4,5	9,9

Mas vejamos na **TABELA 11** como é que esses resultados se distribuem e qual a sua origem.

Tabela 11

CONSOANTE ASSOCIADA A:	espanhol	francês	alemão	sueco	chinês	totais
ataque não ramificado						
oclusivas	8	4	5	2	25	44
nasais	1	3	8	0	6	18

fricativas	32	3	13	9	5	62
líquidas	5	7	5	5	11	33
ataque ramificado	0	1	1	1	12	15
coda	4	4	2	3	15	28
total	50	22	34	20	74	200

Tendo em consideração que cada *sub-corpus* tem mais de 11 000 ocorrências, é provável que os tipos de desvios que atingem somas muito baixas não tenham grande significado. Eles podem ter origem em leituras incorrectas ao nível da transcrição ou em lapsos do autor do texto. Mas as que atingem somas mais elevadas são, com certeza, a face visível de uma dificuldade na aquisição, resultante certamente de características da L1 e ou de especificidades do PE, que a memória visual não consegue esconder. Prestarei, por isso, mais uma vez, particular atenção aos espanhóis e chineses.

Os desvios dos **espanhóis** são bastante transparentes e resultam de diferenças fundamentais entre os sistemas fonológicos das duas línguas, nomeadamente no que respeita o modo de articulação de **oclusivas** e o traço de vozeamento em **fricativas**. No primeiro caso, as formas desviantes têm origem numa diferença marcante entre a variedade centro-meridional do PE, que constitui o padrão, e o Castelhana: a não oposição fonológica nesta língua entre /b/ e /v/ ²⁵⁹. As formas desviantes são quase na totalidade aquilo a que se costuma chamar casos de ultra-correcção e têm na origem esta não distinção: *amabilidade*, *convinar*, *convoios*, *liveralizar*, e *povres*, por <amabilidade>, <combinar>, <comboios>, <liberalizar> e <pobres>. Quando o alvo era a fricativa /v/, surge-nos *cabalos* e *habia*, em vez de <cavalos> e <havia>. Dos 37 desvios em fricativas, 30 resultam de o Castelhana só dispor do fonema surdo /s/ enquanto que em PE existe o seu par sonoro /z/. São exemplos: *acasso*, *andaluces*, *corajossa*, e *messa* por <acaso>, <andaluzes>, <corajosa> e <mesa>; são-no também: *abrazo*, *esperanza*, *esquezer*, *fozem*, *lembranzas*, *zapateiro*, por <abraço>, <esperança>, <esquecer>, <fossem>, <lembranças> e <sapateiro>. Alguns casos referidos acima (*habia*, *andaluces*, *esperanza*, *zapateiro*), dado que a

²⁵⁹ O *corpus* estudado em Barbosa 1999, de que falarei daqui a pouco, apresenta vários casos de «permuta de por <v>, e vice-versa» o que «permite identificar redatores das cartas comerciais» como sendo «naturais do norte de Portugal.» (Barbosa 1999:185).

diferença entre as duas línguas só se situa no fonema/grafema em causa, poderiam até ser considerados como empréstimos. Menos frequentes são os desvios nas ocorrências do par [ʒ] – [ʃ]: *arranchasse* por <arranjasse>, e *dejarte* e *embajada* por <deixar-te> e <embaixada>.

As 8 ocorrências de alvos desviantes em fricativas no sub-corpus dos **alemães**, quando observados em contexto, porque os documentos em que surgem não apresentam indícios de *transfer* de outras naturezas, parecem não estar relacionados com eventuais conhecimentos de Castelhana; além disso, a sua dispersão e repartição são altas: *fotogravia* por <fotografia>; *libro* e *probas* por <livre> e <provas>; *felizidade* por <felicidade>; *arranchar* por <arranjar>; e *cassado*, *dissem*, *confusão*, *decisão*, e *nacismo* por <casado>, <dizem>, <confusão>, <decisão>, e <nazismo>. O mesmo se pode dizer dos dados **suecos**. Mas mais uma vez é o par /s/ - /z/ a maior fonte de desvios (*presos* por <preços>; *conducir*, *dicer*, *decisão*, *precizam*, e *quisseres* por <conduzir>, <dizer>, <decisão>, <precisam> e <quiseres>). A duplamente desviante forma *sourricço* por <chouriço>, que o aprendente sueco que a produziu muito provavelmente nunca tinha visto escrita, alerta-nos para o papel que o contacto com a imagem gráfica tem na aquisição das estruturas fonológicas.

No que respeita as sílabas iniciais com **ataque vazio**, é pouco explícita a relação que a grafia tem com a fonologia e o peso que podem ter as formas escritas respectivas noutras línguas. Mas é entre os alemães que se encontra o maior número de formas aproximadas; as ocorrências são: *campar*, *pareceu* e *presentar*, em vez de <acampar>, <apareceu>, <apresentar>, por um lado, e *speciais* e *spero* em vez de <especiais> e <espero>, por outro. Mas também os espanhóis fazem *levador* por <elevador> e *stão* por <estão>, os franceses *presentadas* por <apresentadas> e *special* por <especial>, os suecos *stúpida* por <estúpida> e os chineses *dois nos* por <dois anos>. Mas, mais uma vez, nomeadamente no caso dos franceses, se coloca a questão de poder tratar-se de neologismos a partir de empréstimos.

Embora muito menos do que seria de esperar, é no caso dos **chineses** que o peso da distância linguística e as características da L1 se fazem sentir. Destaco, por isso, alguns aspectos do sistema fonológico do

Chinês ²⁶⁰, particularmente aqueles que podem ter consequências mais imediatas na interlíngua. Antes de mais, é uma língua tonal e não acentual; ou seja, os tons, quatro em *Putonghuà* e nove em Cantonês, têm valor fonológico. Porque há outras diferenças entre o Mandarim e o Cantonês, e porque a maior parte dos informantes são falantes de Cantonês, é o sistema fonológico desta última que aqui privilegio e passo a descrever, muito resumidamente, seguindo, de muito perto, Nóia 1991:

1. O Cantonês apresenta um sistema de três consoantes **oclusivas**, /p t k/, mas, porque não recorre ao traço de vozeamento neste subsistema, não tem /b d g/. Operando com o traço de africacão que afecta o /t/, resultando /tʃ/, eleva para quatro o seu número, /p t k tʃ/. Aplicando o traço de aspiração obtém um sistema de oito elementos: /p t k tʃ/ vs. /ph th kh tʃh/.

2. O Português tem três **fricativas**, /f s ʃ/, a que aplica o traço de vozeamento, daí resultando /v z ʒ/; o cantonês tem apenas duas, /f s/, e também neste subsistema não opera com o traço de vozeamento.

3. O Cantonês, tal como o Português, dispõe de um sistema de três consoantes **nasais**: /m n ɲ /; no entanto, o ponto de articulação de /ɲ / é diferente: palatal em Português e velar em Cantonês.

4. O Cantonês tem apenas uma consoante **líquida**, /l/, enquanto que o Português tem quatro, /l ʎ r R/.

²⁶⁰ **Chinês** é, como se sabe, o termo genérico para referir um grupo de sete grandes variedades que pertencem à família sino-siamesa. As diferenças entre elas, a nível fonológico, lexical e gramatical, são consideráveis. Entre essas variedades destaca-se um sub-grupo: o *Mandarim* (designação ocidental para *guanhuà*) composto por quatro grandes variedades. O termo *guanhuà* usa-se para referir a língua materna da maior parte da população e a língua franca usada por falantes de outras variedades ininteligíveis entre si. *Putonghua* (*putong* literalmente significa 'comum', e tem implícita a ideia de não sofisticado) é o nome que se dá actualmente a esta variedade, que adopta como ponto de partida os dialectos do Mandarim do norte, o sistema fonológico da variedade de Pequim como base do padrão fonético e a literatura moderna como modelo gramatical (Chen 1993). Às variedades que não pertencem ao grupo do Mandarim, dá-se a designação de variedades do sul. Entre elas, destaca-se o *Cantonês*, falado por exemplo em Hong-Kong e Macau.

5. Em Cantonês nenhuma sequência consonântica pode ocorrer dentro da mesma sílaba, ou seja, o **ataque** silábico é sempre simples.

6. Somente seis, das vinte consoantes existentes em Cantonês, podem ocorrer em ataque de sílaba e em **coda**, /p t k m n ɲ/, estando as restantes confinadas ao ataque. Ou seja, dos três segmentos que em Português podem estar associados à posição de coda, o /r/ não existe em Cantonês e o /s/ e o /l/ não podem ocupar essa posição.

Observemos agora os dados. Começemos pelo **ataque não ramificado**. As **oclusivas** estão muitas vezes associadas a esta posição. De /p/ vs. /b/ só temos *patata* por <batata> e *bouco* por <pouco>. Das oclusivas, é a oposição /t/ vs. /d/ que parece causar mais desvios: *acretidar*, *convitou*, *natariam*, e *quanto* por <acreditar>, <convidou>, <nadariam> e <quando>; *acretidar*, *estudande*, *enquando*, *quando*, e *agretável* por <acreditar>, <estudante>, <enquanto>, <quanto>, e <agradável>. E ainda, *ajua-me* por <ajuda-me>, *optimisna* por <optimista>, *Porqual* por <Portugal>, e *recitava* por <recebia>. Mas também /k/ vs. /g/ é problemático: *Porqual*, *guente*, *figuei* e *segundária* por <Portugal>, <quente>, <fiquei> e <secundário>; outros casos são: *polítio* por <político>, *pirioso* por <perigoso>, e *marioso* por <marisco> ²⁶¹. Apesar de as **fricativas** sonoras também não existirem em cantonês, elas parecem constituir um problema menor. Registam-se só as formas *paucinhas*, *confusão* e *múçica* por <pauzinhos>, <confusão> e <música>.

De desvios em **líquidas** são estes os casos: *chulau*, *Falo*, e *cheio* por <chorou>, <Faro> e <cheiro>; *terefonou* por <telefonou>; *carro* e *burrulho* por <caro> e <barulho>; no caso de /ʎ/ a situação complica-se um pouco. Recorde-se que *fiho* por <filho>, e *conlhecido* por <conhecido> foram classificados como erros de ortografia. Mas temos também *mulheres*, *melihor* e *moiho* por <mulheres>, <melhor> e <molho>, formas que me fazem

²⁶¹ Neste caso, é provável que além da realização em análise, se acrescente o problema da coda na sílaba anterior. Uma análise alternativa é considerar a forma como uma derivação de *mar*, já que é de lá que o marisco vem.

ter dúvidas quanto à opção de considerar aquelas como desvios ortográficos, embora aqui a estratégia ortográfica seja outra. De facto, contrariamente ao que se seria de esperar deste grupo de aprendentes, a interferência da qualidade da palatal alvo no núcleo da sílaba anterior, provocando uma forma ortográfica com <i>, parece revelar alguma acuidade acústica.

Um **ataque ramificado** ou uma **coda** podem sofrer deslocação (*precisa* por <precisa>; *proque*, *preguntar* e *tronou* por <porque>, <perguntar> e <tornou>), apagamento (*sofiam*, *outa*, e *forestes* por <sofriam>, <outra> e <florestas>; *convesaríamos* e *desenvolvimento* por <conversaríamos> e <desenvolvimento> e *quaique* por <quaisquer>, *meno*, *depoi* e *dua* por <menos>, <depois> e <duas>) ou pode ser introduzida uma vogal epentética (*Boraga* por <Braga>; *aperecebemos*, *dificuldade* e *expecialimento* por <apercebemos>, <dificuldade> e <especialmente>). Em relação à coda, como mostram os exemplos, verifica-se a queda da consoante quer em posição interna quer em final de palavra e a inserção de uma vogal epentética que transforma a estrutura silábica: a consoante que estaria em coda da sílaba anterior passa a ataque da nova sílaba. Casos como *proque*, *preguntar* e *tronou* consistem em metátese de consoantes o que, claramente, releva da dificuldade em dominar a distinção entre sílaba com ataque ramificado e sílaba com coda ²⁶². Este *sub-corpus* fornece-nos até algumas formas que só muito dificilmente o contexto ajuda a descodificar. Os desvios afectam mais do que uma sílaba e resultam geralmente de dois ou de vários problemas acumulados: traço de vozeamento, ataque ramificado, e coda, a que se juntam a nasalidade e o sistema vocálico. É o caso de *pargema* por <programa>, *carlo* por <claro>, *peridos* por <prédios>, *branco* por <barco>, *estelas* por <estreitas> e *encomonia* por <economia>.

Também no caso da **nasalidade** (TABELA 12), é neste grupo de sujeitos que os resultados são mais consistentes. Nos outros grupos há vários casos em que hesitei em classificá-los como desvios desta natureza ou como erros de ortografia, nomeadamente no que respeita ao apagamento, já que, muitas vezes, e como acontecia na acentuação, o erro

²⁶² De notar que, em falantes não cultos do PE, também ocorrem este tipo de fenómenos (*precisa* por <precisa> *preguntar* por <perguntar> ou *prefeito* por <perfeito>).

consiste na falta do diacrítico ou no uso de outro (ESP e FRAN: *manha*, *amanhá* por <*manhã*> e <*amanhã*>; ALEM: *alemá*, *manha* e *informações* por <*alemã*>, <*manhã*> e <*informações*>; quatro ocorrências de *maé* por <*mãe*> do mesmo informante; SUEC: *macas* e *amanha* por <*maças*> e <*amanhã*>). Das 9 ocorrências de associação do grupo alemão, 5 são do mesmo informante e na mesma palavra (*Alemanhã* por <*Alemanha*>).

Tabela 12

NASALIDADE	espanhol	francês	alemão	sueco	chinês	totais
apagamento	4	5	12	5	6	32
associação	4	3	9	2	8	26
deslocação	0	0	0	1	1	2

De qualquer modo, a questão da nasalidade põe-se para todos, embora de modo mais nítido para os chineses e alemães, e parece mostrar que a nasalidade associada a vogais está a ser interpretada como traço de palavra. Os casos de associação da nasalidade a uma sílaba são quase sempre consequência de a nasalidade estar presente noutra sílaba da palavra, mesmo que não contígua, ou de estar associada a uma outra palavra que, por razões semânticas ou formais, tem, no léxico mental do falante, conexões com a palavra alvo. Veja-se: ESP: *començar*, *minhã* e *mão* por <*começar*>, <*minha*> e <*mau*>; FRANC: *agenciã*, *Lisbõa* e *pãe* por <*agência*>, <*Lisboa*> e <*pai*>; ALEM: *monstrar* e *Alemanhã* por <*mostrar*> e <*Alemanha*>; CHIN: *normanlmente*, *encomonia*, *bonintas*, *constumam* e *Macão* por <*normalmente*>, <*economia*>, <*bonitas*> e <*Macau*> ²⁶³. Como se pode ver, de facto, a nasalidade é atribuída a uma sílaba quando ela já está presente noutra, ou quando a palavra está mentalmente associada a uma outra que tem esse traço na L1 ou na L2 (vide *mãe* – *pãe*, *Lisbonne* – *Lisbõa*).

Laufer 1997 convidava-nos a avaliar de que modo é que determinados componentes concorriam para potenciar os problemas. De facto, não só no caso dos chineses mas em todos, traço de vozeamento,

²⁶³ É interessante referir que, embora não existam ocorrências de marcação ortográfica de nasalidade associada a *muito*, há várias formas rasuradas, em particular nos alemães e chineses, em que isso se verifica (<*muinto*>; <*muintas*>). A correcção sistemática deve-se certamente ao peso do frequente contacto com a forma gráfica.

ataque ramificado, coda, nasalidade, /r/ intervocálico, e o sistema vocálico (em particular [α] e [e]), quando combinados, criam problemas à estabilização do sistema fonético-fonológico. Por isso, tal como acontece em relação à ortografia, há um certo número de palavras frequentes que concentram uma boa parte dos desvios e que surgem em pelo menos dois dos *sub-corpora*. Dou como exemplos: *camino*, *compania*, e *carro* por <caminho>, <companhia> e <caro> (nos franceses e alemães); *arenjar* e *arranchar* por <arranjar> (nos espanhóis e alemães); *difrente*, ou *defrente*, e *intressante* por <diferente> e <interessante>, em todos ²⁶⁴. E o caso surpreendente da palavra <dinheiro> que, das 139 ocorrências no *corpus*, tem mais de 10 % de formas desviantes (*dinero*, *dinhero*, *dinherio*). No entanto, todos parecem ter problemas com <dinheiro>, excepto os chineses...

Porque se trata de L2, e porque se trata de dados da escrita, não é possível levar muito longe o **confronto com o PE L1**. No entanto, algumas semelhanças existem. Freitas 1997 conclui que a última classe a emergir e a estabilizar são as líquidas (Freitas 1997:139), que a última estrutura a estabilizar é o ataque ramificado (Freitas 1997:167), que o último constituinte a emergir e a ser adquirido é a coda, e que o processo de associação da nasalidade à cadeia segmental põe dificuldades à aquisição. Quanto ao momento da emersão, é claro que não se aplica neste caso de L2, mas, no que respeita a estabilização, como vimos, são estes os principais problemas que subsistem nos dados em apreço, e isso é particularmente claro no caso dos dados dos chineses, a L1 mais afastada. **Para além da distinção vozeada vs. não vozeada, é a estabilização da nasalidade e de sílabas que não seguem a estrutura canónica CV, ou seja, as que têm ataque ramificado e ou coda, que parecem oferecer mais resistência à aquisição. Isto é particularmente verdade quando numa mesma palavra se combinam duas ou mais destas componentes.**

Sendo os materiais escritos a única fonte de informação de se que dispõe para conjecturar sobre momentos passados da língua, a linguística histórica tem-lhes prestado bastante atenção. Entre nós, fizeram-no Clarinda de Azevedo Maia 1986 para o Galego-Português,

²⁶⁴ De qualquer modo, é de assinalar que, em certos casos (ver *difrente* e *intressante*), a forma produzida corresponde à disponibilizada pelo *input*.

e, mais recentemente, Rita Marquilhas 1996 para o PE do século XVII e Afrânio Barbosa 1999 para o Português no Brasil em finais do século XVIII.

Rita Marquilhas 1996 estudou um *corpus* constituído por “provas judiciais arquivadas pelos promotores da Inquisição”, escritas durante o século XVII por portugueses que se iniciavam na “faculdade das letras”. A autora designa-os de “mãos inábeis”²⁶⁵ e conclui que as características ortográficas dos documentos por eles produzidos «permitem confirmar o que se tem apurado nos estudos dedicados à psicogénese da escrita: que os falantes em fase de aquisição de uma escrita alfabética representam a fonologia como uma sucessão de unidades silábicas cuja estrutura interna se lhes apresenta frequentemente insondável, desde que haja uma fuga à estrutura canónica Consoante-Vogal» (Marquilhas 1996:376). Apresenta vários exemplos de sílabas de ataque ramificado e de rima ramificada com /r/, e comenta: «A antologia podia alongar-se porque a ortografia irregular de formas com cadeias de consoantes que incluam /r/ constituem a característica mais recorrente das mãos inábeis seiscentistas. O traço não é histórico, nem específico da escrita do português. Entre inábeis franceses de 10 anos dos séculos XIX e XX observou-se igual comportamento.» (Marquilhas 1996:294). E, embora haja entre as grafias dos inábeis seiscentistas formas em que o segmento deslocado é a letra <l>, esta deslocação «é muitíssimo mais rara do que a de <r>, pelo que parece ser a lateral, na representação do inábil, um segmento mais estável do que a vibrante» (Marquilhas 1996:306).

Afrânio Barbosa 1999, que defende que só sabendo como era o Português no Brasil se pode saber o que viria a ser o Português do Brasil, estuda um *corpus* constituído por documentação oficial e cartas de comércio escritas no Brasil por portugueses em finais do século XVIII²⁶⁶. O autor, seguindo Marquilhas 1996, considera estes comerciantes como um grupo intermédio, “nem *inábeis*, nem *hábeis*”, mas “*pouco hábeis*”, enquanto que os produtores de fontes oficiais seriam “*mãos hábeis*”. Com

²⁶⁵ Segundo Marquilhas (1996:279) esta é a melhor tradução que encontrou para o termo «*scripteurs maladroits*» de Claire Blanche-Benveniste 1993.

²⁶⁶ O *corpus* reunido e estudado por Afrânio Barbosa é constituído por um total de 117 documentos: 93 cartas de comércio escritas no Brasil por portugueses e 14 documentos oficiais expedidos do Rio de Janeiro para Lisboa. (Barbosa 1999:15).

base nos dados recolhidos, Barbosa conclui que «as flutuações gráfico-fonéticas [estão presentes, no século XVIII] tanto nas mãos hábeis dos profissionais da escrita, produtores das fontes oficiais [como nas] dos escritores menos hábeis das cartas de comércio». «A diferença entre eles reside na proporção de incidências de cada tipo de atestação gráfica dos processos fonéticos » (Barbosa 1999:165). Duas das ocorrências registadas, tanto em cartas comerciais como em documentos oficiais, coincidem com as do *PESTRA*: *perguntar* por <perguntar> ²⁶⁷ e *percizar* por <precisar>. A forma *perfiro* por <prefiro>, que ocorre em carta comercial, ocorre também no nosso *corpus*. Ela é, dos 63 desvios formais advertidos no material produzido pelas “mãos hábeis” dos nossos estudantes universitários que constituem o grupo de controle, o único caso relativo à estabilização do sistema consonântico do PE.

Segundo Ellis (1997:124-5) a repetição de sequências fonológicas na memória a curto prazo (MCP) permite a sua consolidação na memória fonológica a longo prazo (MLP). Quanto mais um estímulo se repete na MCP, maior é a MLP e mais fácil é repeti-lo. Este processo ocorre para todo o material fonológico ²⁶⁸. À medida que o vocabulário se estende e que a sua prática de ouvir e produzir aumenta, assim o falante adquire automática e implicitamente a frequência estatística e as probabilidades sequenciais dos padrões fonológicos da L2. O grau de rapidez com que este conhecimento implícito tem lugar depende da proximidade entre a L1 e a L2 (Ellis 1997:127). Mas, tendo em consideração os dados da aquisição do português L1 e da aquisição da escrita, podemos acrescentar que esse grau de rapidez depende também de um continuum de dificuldade de aquisição, comum aos falantes de L1 e de L2, que espelha, certamente, o grau de universalidade e de marcação de um determinado traço ou de uma estrutura.

Quando comparados com as “mãos inábeis” de Rita Marquilhas, os informantes do *PESTRA* cometem muito menos erros. Quando comparados com “as mãos pouco hábeis” de Afrânio Barbosa, cometem bastante mais. No entanto, é preciso ter em consideração que no caso dos estrangeiros não se trata de adquirir os princípios abstractos que regem os sistemas de escrita segmental, visto que eles já tinham sido adquiridos por todos, aquando da aquisição da escrita das respectivas L1 (e reutilizados em outra

²⁶⁷ A oscilação entre *perguntar* e *preguntar* é até objecto da base X do *Relatório do Acordo de 1945*. As formas *preguntar* e *prêguntar* são nesta consideradas como «meras representantes de variações fonéticas». Ver Castro, Duarte & Leiria 1987:171.

²⁶⁸ Ver tratamento quantitativo das várias estruturas silábicas, realizado por d’Andrade e Viana (1993) e referido antes.

ou outras L2), mas sim de adquirir as estruturas e as distinções específicas do português. Isto aplica-se até mesmo aos chineses, visto que todos já tinham aprendido a escrever em inglês. Por isso, e como já salientei, o número de desvios observados em formas escritas, que dependem em boa parte de memória visual, pode ser enganador quanto ao **grau de aquisição do sistema fonético-fonológico do Português**, que, por sua vez, depende sobretudo da sua representação na memória fonológica a longo prazo.

Lahiri & Marslen-Wilson 1991 afirmam que temos uma representação com baixa especificação da forma lexical. Isto afigura-se como plausível se pensarmos na enorme variação, no plano fonético, a que, mesmo dentro de uma mesma variedade de uma mesma língua, uma forma está sujeita em função do contexto linguístico e do falante que a produz. Se o falante só dispõe de uma representação pouco específica, então também é plausível inferir que pequenas diferenças entre línguas próximas não estarão fortemente representadas no léxico mental. Isto tem certamente consequências na compreensão, aprendizagem e manutenção de línguas como o português, o espanhol, o italiano, e, em certa medida, o francês. Daí resultará, por um lado, uma competência acrescida de um falante de uma delas para a compreensão das outras, mas, em contrapartida, maiores dificuldades para a manutenção delas em separado. Lembremos ainda que aspectos estruturais e perceptuais são processados a um nível mais superficial e que o sistema fonético-fonológico da L1 está altamente automatizado. Daí resultará, por outro lado, maior dificuldade na reestruturação deste sistema em direcção ao da L2.

Dito de outro modo, o número relativamente semelhante de **formas aproximadas detectadas na escrita**, nos vários *sub-corpora*, não significa necessariamente que o estágio de aquisição dos diferentes sistemas fonético-fonológicos da interlíngua dos diferentes grupos seja equiparável tanto para a compreensão como para a produção de discurso *on-line*. O contacto repetido com a forma escrita das palavras ajudará a aquisição automática e implícita da frequência estatística e das probabilidades sequenciais dos padrões fonológicos da L2, mas isso só indirectamente se reflectirá na estabilização do sistema e nas suas potencialidades para a codificação e descodificação de **discurso on-line**.

A prová-lo está o perfil linguístico de cada grupo de informantes, estabelecido a partir da prova de compreensão do oral realizada na mesma ocasião que os documentos que constituem o *corpus* em estudo, e de que dei conta em 5.2. *Informantes e materiais*). Como se poderá verificar no quadro em que se dá conta do perfil linguístico dos informantes seleccionados, no que respeita a compreensão do oral, 94 % dos informantes espanhóis, e 81 % dos franceses mostraram ter compreendido perfeitamente toda a informação veiculada no texto estímulo; esse montante desce para 46,2 % no caso dos suecos, e 39,4 % no dos alemães; mas no caso dos chineses não representa senão 4,1 % dos informantes; ou seja, apenas 2 dos 49 informantes chineses mostraram ter compreendido todos os detalhes de uma história que foi lida três vezes; além disso, apenas outros 8 deram provas de ter compreendido uma boa parte.

Motivada por esta enorme disparidade entre o conhecimento da forma escrita das palavras (e, pelo menos, o de um significado e de certas propriedades sintácticas que lhe estão associadas), que permite aos falantes de chinês compreender e produzir textos escritos em português, mas não produzir e, sobretudo, compreender discurso *on-line*, realizei um **trabalho experimental** que tinha como principal objectivo testar o desempenho de um grupo de aprendentes falantes de línguas que não utilizam (ou não utilizam exclusivamente) o alfabeto latino ²⁶⁹. O grupo era constituído por um falante de hindi, por um de uma variedade de árabe (de Oman), por dois falantes de japonês, e por quatro falantes de chinês (um de mandarim e três de cantonês), todos com igual tempo de estudo de português (90 horas lectivas).

A experiência envolvia um conjunto de tarefas: **1.** ouvir (duas vezes) e escrever em seguida cada uma das 125 palavras de uma lista resultante de um texto previamente desmontado; **2.** ler as palavras da lista e sublinhar as desconhecidas; três semanas depois, **3.** ler um breve resumo do mesmo texto, ouvir ler o texto duas vezes; ouvir o texto, segmentado em sequências curtas, cada uma lida duas vezes pelo experimentador, e escrevê-las, em seguida; **4.** ler o texto e sublinhar todas as palavras ou sequências desconhecidas.

²⁶⁹ Deste trabalho experimental, dá conta Acosta & Leiria 1997.

O desempenho dos chineses foi muito mais fraco que o dos outros sujeitos envolvidos no estudo no que respeita as tarefas que envolviam descodificação oral. No caso destes, não foi muito significativa a diferença de resultados entre a tarefa **1.** (escrever palavras descontextualizadas) e **3.** (escrever palavras em contexto); os sujeitos, quando confrontados com os estímulos visuais, declararam conhecer um número muito superior de palavras em relação àquelas que mostraram ser capazes de escrever correctamente quando confrontados com o estímulo auditivo; verificaram-se desvios até mesmo em palavras muito frequentes e com as quais os sujeitos tinham contactado repetidas vezes (ex.: *agua* por <agora>, *aou* por <ao>, *andes* por <antes>, *banha*, *belho*, *banhu* por <banho>, *viso*, *lixo* por <disso>, *oci* por <hoje>, *óptima*, *o pimto*, *outíma* por <última>); os desvios concentram-se em palavras formadas por um número de sílabas e de morfemas igual ou superior a dois; verificou-se uma tendência para reproduzir o esquema silábico CV, ignorando o ataque ramificado e a coda (com consequências notáveis a nível da morfossintaxe: singular vs. plural, e distinção entre formas verbais conjugadas e infinitivo); o número de desvios foi superior sempre que nas sequências consonânticas intervinham pares distintivos inexistentes em chinês (ex.: *felante*, *falentes* por <frente>; *outar* por <outra>; *fulutas*, *furitas* por <frutas>; *presa*, *perlas*, por <beleza>; *exeprencia*, *siplicia*, *cibriencia*, *espelancão*, *explia*, por <experiência>); o traço de vozeamento mostrou ser causa de desvios em todos os casos, muito especialmente no par oclusivo dental (ex.: *andes* por <antes>; *serdesçe*, *serdese*, *serdesa* por <certeza>; *condar* por <contar>; *quadro* por <quatro>; *domar* por <tomar>; *furdar*, *vou dar* por <voltar>, *prados* por <pratos>; *apetide*, *apedite*, *à petile*, *aprdida* por <apetite>; *malda*, *marda*, *malde*, *marde*, por <malta>); em relação às fricativas, o problema não foi menor (ex.: *parssiço*, *greciso*, *pessou*, por <preciso>; *xatise*, *chetise*, *jedise*, *chetis*, *chedisa*, *cerdisa* por <chatice>); e, como se pode constatar nos exemplos acima, o sistema vocálico foi o que ocasionou maior número de formas desviantes, mas, embora eles se estendam à vogal tónica, foi o vocalismo átono que originou a maior parte deles.

Perante a disparidade entre estes dados obtidos por exposição a um estímulo auditivo e o reduzido número de desvios ortográficos e de formas aproximadas presentes no material escrito, parece não

restarem dúvidas quanto ao papel que a memória visual deste grupo linguístico desempenha na aprendizagem da forma escrita das palavras e quanto ao impacto que o uso ou não de um sistema alfabético tem no processamento lexical. O que mais me surpreende é que, ao olharem para estas formas tão desviantes, os seus autores parecem não conseguir confrontá-las com a lista de palavras do português que têm armazenada na memória visual ²⁷⁰. Se isso lhes fosse possível, ao verificarem que tais formas não fazem parte dela, seriam capazes, pelo menos nalguns casos, de as reinterpretar e rescrever aproximando-as da forma armazenada na memória visual. Essa manifesta impossibilidade resulta com certeza de as conexões entre a imagem gráfica (e o significado que lhe está associado) e a imagem sonora serem muito ténues ou inexistentes, e de o tempo de que dispõem para o processamento (tanto palavra a palavra como *on-line*) não ser suficiente para as aproximarem e relacionarem ²⁷¹. Isso explica por que é que os chineses, dispendo de um vocabulário sensivelmente igual aos outros, de certo modo até mais rico, obtiveram resultados tão insatisfatórios na compreensão oral. Na verdade, como vimos no Capítulo 2, era esta a hipótese de Meara:

My guess is that lexical access is in fact a key element in most language skills, playing a particular important role in listening comprehension and in oral fluency. Learners whose lexical access skills are poor would be expected to perform badly on tasks which involve real-time processing. (Meara 1993:293)

6.3.1.3. *Desvios na flexão de nomes, adjectivos e verbos*

²⁷⁰ Gomes de Morais & Teberosky (1994:23), sobre «**erros e transgressões infantis na ortografia do PB**», afirmam: «os sujeitos ouviam frases pronunciadas segundo o dialecto culto do pesquisador mas, ao notá-las, pareciam basear-se em suas próprias pronúncias.»

²⁷¹ A propósito da relação entre alfabeto e ortografia, observa Cagliari (1993/1999:174-175): «Conhecer o alfabeto é um ponto de partida, mas **saber lidar com a ortografia** é fundamental, quer para escrever, quer para ler. [...] Depois que a ortografia passou a comandar a escrita alfabética, quem escreve coloca as letras nas palavras como um chinês coloca os traços nos caracteres de sua escrita. Ambos se guiam, basicamente, pelo significado e não pelos sons da fala. Dizemos que o alfabeto é um sistema fonográfico, porque partimos dos sons vocálicos e consonânticos para escrever, mas somente terminamos a tarefa quando constatamos que, apesar das variações possíveis de pronúncia, a escrita de uma palavra é única». Também é provável que a variação morfossintáctica a que as palavras estão sujeitas seja um factor de grande perturbação para os chineses, dificultando a consciencialização de que «a escrita de uma palavra é única.»

Saber uma palavra também é saber as diferentes formas que resultam da sua flexão. No que respeita os **nomes e adjectivos**, e como pode ver, na *TABELA 9* em 6.3.1., é muito reduzido o número de desvios desta natureza, sendo isto particularmente verdade no caso dos *sub-corpora* alemão, sueco e chinês. As ocorrências desviantes na **flexão de número**²⁷² são de nomes e adjectivos terminados em **-ão**, **-l**, **-r**, ou **-s**. Nos alemães, quatro ocorrências de *bars* por <bares> e duas de *português* por <portugueses>, *avós* por <avós> e *pasteis* por <pastéis>, nos suecos, e *português* por <portugueses> e *patrões* por <patrões>, nos chineses, constituem a totalidade deste tipo de desvios destes aprendentes. Uma característica do português, que o distingue das outras línguas românicas, a supressão do *l* e do *n* intervocálicos, reflecte-se na flexão em número e está na origem da maior parte dos erros dos espanhóis e franceses. Além de *refeições* por <refeições>, nos espanhóis; *país* por <países>, *razões* e *comparações* por <razões> e <comparações>, e *familiais* por <familiares>, formas encontradas no *sub-corpus* francês, as restantes ocorrências são de nomes e adjectivos terminados em *-l* (ESP: *racionaes*, *descomunales*, *espanholes*, *faciles*, *interminavels*, *personales*, e *sociales* por <racionais>, <descomunais>, <espanhóis>, <fáceis>, <intermináveis>, <pessoais>, e <sociais>; FRANC: *esteriles*, *hoteles*, *hotels*, *horriveles* e *papels* por <estéreis>, <hotéis>, <horríveis> e <papéis>).

Formas verbais desviantes são em maior número e atingem valores muito próximos em todos os *sub-corpora*. A maior parte dos desvios resulta de:

1) confusões entre conjugações diferentes (FRANC: *reflectei*, *vivei*, e *utilizia* por <reflecti>, <vivi> e <utilizava>; ALEM: *corriu*, *escolhei*, e *percebiu*, por <correu>, <escolhi>, e <percebeu>; SUEC: *escolhei*, *morem* e *gostiamos* por <escolhi>, <morem> e <gostávamos>; CHIN: *chameu*, e *comivam* por <chamou> e <comiam>);

²⁷² «[D]e um ponto de vista formal, **a variação em número é sistemática e obrigatória**. À excepção de um pequeno conjunto de formas lexicalmente condicionadas, qualquer palavra de natureza nominal é susceptível de participar na oposição singular/plural, quer se trate de uma especificação por concordância, como nos adjectivos. Tratando-se de um processo sistemático e obrigatório, **a realização de contrastes de número é adequadamente caracterizada como um processo de flexão**.» (Villalva 1994:213) (bold no original)

2) regularização de paradigmas (ESP: *fazerá* por <fará>; FRANC: *pedo* por <peço>; ALEM: *pedo* por <peço>; SUEC: *explicastes* e *percebistes* por <explicaste> e <percebeste>, *quere* por <quer>, e *hão (muitas pessoas)* por <há>; CHIN: *ouvo* por <ouço>, *havam* por <havia>);

3) em verbos irregulares, derivação de formas a partir do radical inadequado (ESP: *possas* por <podes>; ALEM: *possam* por <podem>; e *punhessem* por <pusessem>; SUEC: *possa* por <pode>; CHIN: *possa* por <pode> e *foríamos* por <iríamos>) ²⁷³; verbos de alternância vocálica (ESP: *sento*, *sintiu*, *sugero*, *sufre*, e *servem* por <sinto>, <sugiro>, <sofre>, e <servem>; FRANC: *consiguem*, *preferam*, *sento*, *sintia*, e *servem* por <conseguem>, <preferam> <sinto>, <sentia> e <servem>; ALEM: *prefero*, *sentam-se* e *sintia-me* por <prefiro>, <sintam-se> e <sentia-me>; CHIN: *sento* e *senteu-se* por <sinto> e <sentiu-se>); e resultam ainda de confusão entre formas dos verbos *ver* e *vir* (ALEM: *visse* por <viesse>; CHIN: *viemos* por <vimos>).

De facto, e **parecendo confirmar aquilo que defendem muitos investigadores, os aprendentes não transferem a morfologia flexional** ²⁷⁴; **ou seja, os erros não reflectem os sistemas da L1** ²⁷⁵. Por vezes, têm origem numa representação desviante a nível da base lexical que, essa sim, pode ser determinada por características da L1 ou de outra língua conhecida; é o caso de vocábulos que, em português, sofreram a supressão do *l* e do *n* intervocálicos, mas não nas outras línguas românicas, assim como de verbos que pertencem a diferentes conjugações na L1 ou em outra conhecida pelo aprendente e em português.

6.3.1.4. *Desvios na atribuição de género*

²⁷³ Em qualquer dos casos do verbo *poder* incluídos aqui, não há nada no contexto que faça supor que se trata de um erro de morfossintaxe, ou seja, de um Presente do Conjuntivo em vez de Presente do Indicativo.

²⁷⁴ «[I]nflexional morphology does not transfer» (Eubank 1996).

²⁷⁵ No entanto, há dois ou três casos de formas verbais do *sub-corpus* sueco para as quais tenho dificuldade em encontrar outra explicação que não a da transferência de morfologia flexional. Veja-se: «Álcool e condução é duas coisas que não é bom para mexer. Mas os portugueses não **comprender** essa coisa.» (S40). Sendo **-r** o morfema de Presente do Indicativo na L1, e não havendo uma razão aparente para a selecção do Infinitivo neste contexto, parece-me pertinente pôr a hipótese de que, porque a forma terminada em **-r** existe em Português, o aprendente possa usá-la com o valor de Presente do Indicativo.

Considerar a atribuição de género como um desvio formal não é uma decisão isenta de controvérsia. Mas, de facto, embora o género tenha sempre uma base semântica, existem poucas línguas com sistemas estritamente semânticos, sendo os sistemas mais comuns os formais de base semântica (Corbett 1991:63, citado por Villalva 1994:276). Em português, «a atribuição de um valor de género aos nomes resulta da intersecção de critérios semânticos (predominantes) e de critérios formais, que têm uma considerável capacidade preditiva e integra um igualmente considerável conjunto de excepções. A atribuição de um valor de género aos adjectivos é sintacticamente determinada.» (Villalva 1994:234-235). Visto que «os critérios semânticos não explicam a especificação de género em nomes que referem entidades inanimadas» (Villalva 1994:229), é com base nesses critérios formais, que, uma vez conhecidos, e apesar das excepções, fornecem, em muitos casos, a um falante estrangeiro orientação para atribuir género a essas entidades, que incluo aqui este tipo de desvios. Incluo também os desvios de concordância em género sempre que, e como se poderá ver pelos exemplos dados a seguir, eles parecem ser motivados por uma atribuição desviante de género ao nome envolvido.

No caso dos falantes de **línguas românicas**, a motivação para o reduzido número de desvios é, em quase todos os casos, bastante transparente: nomes inanimados que têm um género diferente na L1 (ESP: *os árvores, as costumes, um garagem, as legumes, a leite*; FRANC: *os arvores, o calma, o paisagem, o ponte, o viagem, a limite*), ou a que o aprendente, por falta de apoio da L1 ou da L2, atribui um género formal incorrecto (ESP: *da nada, o flexibilização*; FRANC: *a amanhã, do Camara, o chamuss, os férias, a pastel*).

Como é sabido, as **línguas germânicas** têm, neste ponto, muito pouca orientação a oferecer aos seus falantes que seja transferível para o português ²⁷⁶. Mas, embora o número de desvios na atribuições de género

²⁷⁶ Em **Alemão**, os nomes podem pertencer a um de três géneros: masculino (*der Vater*), feminino (*die Mutter*) ou neutro (*das Kind*). Certos sufixos determinam o género dos nomes. Mas a única forma segura de saber o género de qualquer nome inanimado é através do artigo. No entanto, Salmons (1993:411) afirma: «German gender assignment has been shown to be largely predictable on the basis of phonological shape (e.g. final and initial segments or clusters), semantic features, and morphological features...». Em **Sueco**, o sistema é bastante simples. Os nomes distribuem-se por duas classes distinguíveis pelos artigos **en** e **ett**. (*en flicka, en pojke, en person, en bok, en stad, en blomma; ett brev, ett rum, ett apple, ett ar, ett*

seja relativamente elevado, os falantes de **alemão** e de **sueco**, salvo raras exceções (ALEM: *os camisas, outro cantina, um praia, nas campos*; SUEC: *um característica, um vida, o empresa*), parecem ter aprendido a regra básica em português. De facto, a maior parte dos desvios são em vocábulos terminados em **-e** (ALEM: ex: *o árvore, o base, o carne, o cidade, os cores, as transportes, Rua do Sorte*; SUEC: ex.: *as cafés, um chatice, um cidade, um ponte, um qualidade, um responsabilidade, a restaurante*), em **-ão** (ALEM: ex.: *o alimentação, os condições, o razão*; SUEC: ex: *o combinação, o construção, o direcção, o situação, o solução, o televisão, o mão*), em vogal nasal ou em consoante (ALEM: ex: *o vantagem, o imagem, o paisagem, o paragem, o capital, a funeral, uma favor*; SUEC: ex.: *os embalagens, uma jardim, o paragem, as lugares, uma país*), ou regidos por regras especiais (ALEM: ex: *uma dilema, a dia*; SUEC: ex: *uma mapa, a telefonema, a problema*). O mesmo se passa em relação aos desvios de concordância com adjectivos ou determinantes (ALEM: ex: *pequenos dores, estes cores, muito sorte, muito saudade; este afirmação, condições subhumanos, nosso conversação; muita carisma, outras problemas*; SUEC: ex: *carne magro, velho cidade, muitas filmes, novos oportunidades; muitas aviões, muitos reclamações, alguns sensações, este situação; mesma problema*).

Ainda que a L1 não tenha nada para oferecer aos **chineses**, o número de desvios, também neste caso, é relativamente pequeno. Poucos são nos vocábulos terminados em **-o** e em **-a** (*o cadeira, o maneira, os ruas*), assim como em **-ão** (*às patrões, no refeição*); registam-se alguns em vocábulos terminados em **-e** (*os árvores, o cidade, o humidade, um parte*), em consoante (*uma lugar, as vegetais, a país*) e em vocábulos que constituem excepções (*a clima, a delta de Cantão, uma dia*). O mesmo se passa em relação aos desvios de concordância com adjectivos ou determinantes (*muito desilusão, muito saudade, pouco livradade, boas hotéis, várias iates, outra programa*).

Como se pode ver pelos exemplos, os desvios incidem muitas vezes nos mesmos vocábulos, independentemente da L1. Do total de 319 desvios relativos à atribuição de género, 76, ou seja, aproximadamente 24 %, concentram-se em apenas cinco vocábulos: *cidade, problema, viagem*, e em

trad....) (*uma rapariga, um rapaz, uma pessoa, um livro, uma cidade, uma flor; uma carta, uma sala, uma maçã, um ano, uma árvore*).

outros dois, a que ainda não fiz referência, mas que, de facto, também são comuns a todos os *sub-corpora*. Trata-se do muito frequente sobrecomum **peessoa** e de **gente**. Dou aqui alguns exemplos.

O que me parece que se pode inferir desses exemplos é que, no léxico mental de alguns dos aprendentes, os nomes **sobrecomuns** *peessoa* e *criança* estão representados como se fossem **comuns de dois** (*o/a peessoa, o/a criança, como o/a artista ou o/a turista*). No caso dos comuns de dois, sendo o masculino, em português, tal como nas outras línguas românicas, o valor não-marcado, é ele que é seleccionado sempre que se pretende obter

- E15** *eu acho que <m> nem quando uma peessoa es velha pode saber se ele escolheu bem o mal*
- F32** *Desde algunos anos, os peessoas preferiram os viagens em grupos...*
- A07** *Todos vezes no primero mês quando eu encontrei peessoas portuguesês eu tinha dificuldades <para> a entender e a falar com eles.*
- A36** *Peessoas primitivos existem em todos os países, mas eu gosto mais de viver em Portugal, porque tudo é muito mais 'humano' entre as peessoas.*
- A45** *Quando pedo as peessoas, dizem que eles não gostam o metro e o autocarro em Lisboa.*
- S01** *... muitas peessoas dão importância a sua aparência porque eles querem ter uma personalidade diferente de outras peessoas.*
- S41** *Hoje muitas peessoas só comem legumes. Eles pensam que é bom para a saude...*
- C08** *...os peessoas contam nós sobre cultural Portugal, e eu gosto muito...*
- C29** *O mar estava calmo. Os peessoas tomavam alguns banhos.*
- C30** *quando era criança, não podia sair de casa <so> brincava com outro criança...*
- C32** *Um criança tem a sua forma de ver o mundo [...] porque ele vivia no mundo só pouco tempo.*
- F32** *Respeito os gostos de todos os gentes, mas...*
- A07** *Em Alemanha os gentes precisam mais ropas ...*
- A49** *Quando estamos nas transportes públicos, toda a gente é simpático, e nos temos um lugar sentado em todos nossos viagens.*
- C18** *os cidades: Porto, Braga etc têm lugares do trabalho. O gente ganhou dinheiro facil.*
- C29** *Muito gente sentiram-se no cadeira e beberem um sumo e tomarem café.*

uma referência genérica. Daí *os pessoas, pessoas portuguesas, pessoas primitivos* e o pronome pessoal masculino *eles*. Esta representação de *pessoa* é, por sua vez, generalizada ao vocábulo *a gente* (*as gentes, os gentes*), que, frequentemente, no léxico dos aprendentes, corresponde a *as pessoas*, e não ao valor pronominal que tem actualmente em português ²⁷⁷. De facto, e apesar da frequência altíssima no *input*, *a gente* nunca é usado no *PESTRA* com este valor.

6.3.1.5. Conclusão

Da observação dos dados até agora analisados, ou seja, relativos a **desvios formais**, podemos desde já retirar algumas conclusões:

1. factores intralexicais combinados potenciam as dificuldades de aquisição; (veja-se traço de vozeamento, ataque ramificado, coda, nasalidade, /r/ intervocálico, e o vocalismo átono, quando combinados, criam problemas à estabilização do sistema fonético-fonológico);

2. representações desviantes a nível do léxico têm consequências na gramática; (veja-se os nomes sobrecomuns *pessoa* e *criança*, representados como se fossem comuns de dois, e as consequências que isso tem na morfossintaxe).

3. os desvios em vocábulos muito frequentes representam uma boa percentagem do total; (veja-se erros de ortografia e de acentuação em formas dos verbos e em outros vocábulos de alta

²⁷⁷ Lopes 2003 é um estudo do percurso histórico da «inserção de *a gente* no quadro pronominal do português». A autora, aplicando pressupostos funcionalistas, generativistas e variacionistas, analisa «a atuação de fatores linguísticos e extralinguísticos na alteração categorial de nome para pronome» numa amostra constituída por documentos do PE, PB e PA (Moçambique) que vão do século XIII ao século XX. Em relação ao PE, a autora conclui que, «embora a gramaticalização tenha se iniciado na mesma fase histórica» [século XVII-XVIII], «nos textos escritos por portugueses, o emprego da nova forma pronominal (*a gente*) é menos significativo» e apresenta «um comportamento mais conservador em termos da implementação da mudança», enquanto que os textos brasileiros e moçambicanos têm «um comportamento mais inovador, no que se refere à frequência de uso da nova forma pronominal.» No caso do PE, verifica-se uma «maior probabilidade do uso de *a gente* com formas da terceira pessoa do singular» e também uma «maior tendência ao emprego de *a gente* pronominal em sujeitos genéricos».

frequência; veja-se, entre outras, as formas desviantes de *dinheiro*; veja-se a atribuição desviante de género a *cidade*, *problema*, *viagem*, *pessoa* e *gente*);

4. factores interlexicais, de que se destaca a relação tipológica, afectam certos aspectos de uma palavra; (veja-se o peso que a proximidade entre as línguas tem na transferência de conhecimentos a nível da ortografia e do sistema fonológico).

Mas, de facto, como nos diz Corder 1983, o papel desempenhado pela L1 é bastante mais penetrante e subtil do que tradicionalmente se tem acreditado. A prová-lo estão os dados dos chineses. Estes, se não têm conhecimentos linguísticos prévios directamente transferíveis para a L2, **transferem muitas das estratégias e das competências gerais adquiridas aquando da aquisição da L1.** Isso fica claro no número insignificante de erros ortográficos e de acentuação. Responsável por isso é a memória visual treinadíssima de que dispõem, é a capacidade de atenção ao pormenor, à excepção, ao caso a caso... Verificámos que aproximadamente 50 % dos desvios formais advertidos no *sub-corpus* chinês resultam da componente fonético-fonológica do léxico. Constatámos também que esse número é só a ponta do *iceberg* que a memória visual não conseguiu esconder completamente. Michel Paradis (1994) falou-nos sobre os tipos de memória e de conhecimento envolvidos na aquisição e uso de diferentes componentes do léxico: os aspectos fonológicos são compatíveis com a memória e com o conhecimento procedimental. É esse conhecimento procedimental, indisponível por causa da distância linguística, que o conhecimento declarativo, o saber caso a caso, vem compensar. Mas, infelizmente, só na compreensão e na produção escritas, esse recurso dá frutos. De facto, **quanto maior é a distância linguística entre a L1 e a L2 mais o aprendente se apoia em conhecimento explícito ou declarativo e maior será a diferença entre os seus processos de aquisição e de uso e os processos usados pelos falantes nativos.**

6.3.2. Vocabulário indisponível e estratégias compensatórias

Para suprirem faltas no seu léxico, os aprendentes usam várias estratégias, muitas delas coincidentes com as dos falantes nativos, ou mesmo das línguas quando encaradas numa perspectiva diacrónica. Neste caso, para além da mudança semântica ou gramatical, são frequentes o recurso a **empréstimos** de outras línguas, a que se seguem fenómenos de relexificação, e a criação de **palavras novas**. Mas o facto de um significado ser codificado numa língua por uma palavra ou por duas ou três, é relativamente arbitrário, dependendo isso, em parte, da sua relevância para a comunidade falante. Assim, outra estratégia também muito produtiva consiste em, a partir de itens lexicais disponíveis, recorrer a processos aditivos, formando compostos, locuções e outras **combinatórias**. Como vimos, estes itens compostos põem problemas de duas ordens: quanto ao modo como estão representados no léxico mental e quanto ao seu grau de fixidez e idiomaticidade. No entanto, e porque, como vimos em 6.3.1, neste estágio, a interlíngua dos falantes comporta sobretudo combinatórias com um grau de fixidez relativamente alto, elas serão tratadas aqui, embora, em alguns casos, as combinatórias aproximadas possam eventualmente resultar de desvios em propriedades lexicais. A *TABELA 13* dá conta do número de ocorrências, em cada um dos *sub-corpora*, relativo a **empréstimos**, criação de **neologismos a partir de empréstimos**, e criação de **neologismos formais**, e também de **combinatórias aproximadas**.

Tabela 13

<i>VOCABULÁRIO INDISPONÍVEL</i>	espanhol	francês	alemão	sueco	chinês
Vocábulo:					
empréstimos	72	51	30	25	8
neologismos a partir de empréstimos	64	52	13	9	1
neologismos formais	15	25	26	44	40
total	148	126	66	75	39
%	11,4	9	3,9	3,8	1,9
<i>Combinatórias aproximadas</i>	68	86	106	92	118
%	5,1	6	6	4,5	5,7

6.3.2.1. *Vocábulo*

Zobl 1980 define **empréstimo** como a substituição de um item lexical de uma língua por um outro sem que isso interfira na estrutura sintáctica. Como vimos, também Poulisse 1993 considera este recurso como uma estratégia de substituição. No caso de criação de palavras novas, que resultam de empréstimos ou de aproximações a vocábulos existentes na variedade alvo, Poulisse diz haver um processo mais complexo. Trata-se, como também vimos, de aliar à substituição uma estratégia (substituição + estratégia). Daí resultam **neologismos a partir de empréstimos** ou relexificações, ou seja, empréstimos, da L1 ou de outra, modificados no que respeita a representação fonológica (ou, como neste caso, gráfica) de modo a ficarem mais de acordo com aquilo que o aprendente considera ser próprio da língua alvo, e **neologismos formais**, ou seja, formas desviantes mas construídas com materiais dessa mesma variedade (*interpretora* ou *embarar* por *intérprete* ou *partir*)²⁷⁸.

Os resultados deixam patente, mais uma vez, o papel da distância linguística. E, neste caso, não só na aprendizagem mas também na *performance*. De facto, e como seria de esperar, no caso dos **espanhóis**, todos os 72 empréstimos são desta língua. Embora haja outros (*afrontar, alrededores, desafortunadamente, camarero, cola, compartir, contestar, seguridad, influir, impulsar...*), os que registam maior número de ocorrências são vocábulos muito frequentes (*así, mas, no, e o por assim, mais, não e ou*). Não são, certamente, vocábulos desconhecidos; **a indisponibilidade aparente resulta sim de estes vocábulos estarem muito activos na L1 e de, conseqüentemente, serem muito difíceis de controlar. Trata-se, portanto, de uma questão de performance e não de aprendizagem.** E isso distingue os empréstimos dos neologismos a partir de empréstimos (*confraternidade, ciudade, gerações, indudável, saludavel, vontade, melhorar...*). O uso de uma estratégia de relexificação, que nalguns casos (*gerações, saludavel, vontade*) deixa intocadas diferenças fundamentais entre as duas línguas (a manutenção em Castelhana do **-n-** e do **-l-** intervocálicos), mostra que não é uma incapacidade de controle mas sim a convicção de que, de facto, eles fazem parte da L2. Neologismos formais não são muitos e, tal como acontece com algumas das formas aproximadas,

²⁷⁸ A tipologia de desvios de PPOM só distingue *empréstimos* de *neologismos de forma*.

alguns deles (*ensangados* e *arascar-se* por *zangar-se* e *desenrascar-se*) parecem reflectir o contacto muito esporádico com os itens, em particular, na sala de aula.

Apesar de, como referi, não ter incluído na amostra documentos de informantes que tivessem declarado ter um conhecimento mais do que elementar de outras línguas românicas, a dificuldade em manter separados os léxicos de duas línguas tão próximas como são o português e o espanhol manifesta-se também nos outros *sub-corpora* ²⁷⁹. Dos 51 empréstimos a que recorrem os **franceses**, 28 são do espanhol (*cambiar, legumbres, espacio, olvidar, o, salí, vivir...*), 2 do italiano (*divertente* e *punto*) e só 21 são do francês (*autre, contre, interessant, minute, mais...*). No entanto, conscientes da distância relativa entre a sua L1 e o português, os franceses preferem aplicar uma estratégia e produzir neologismos a partir de empréstimos (*atiram, convir, nome, arriva, traversam, augmentam, regrettará* por *atraem, convidar, chama, chega, atravessam, aumentam e lamentará*) e neologismos formais (*emborava* por *partia*), alguns deles relacionados com a questão da atribuição de género (*o diploma, o escolho, a interpretora, os turistos* por *o diploma, a escolha, a intérprete, os turistas*).

Quanto mais afastadas são as duas línguas mais o aprendente descobre a impossibilidade de recurso a empréstimos. Mas qualquer língua conhecida pelo sujeito pode ser fonte de empréstimo. Por isso, os **alemães** não pedem emprestadas palavras à sua L1, mas a outras línguas românicas, tanto para uso directo (16 ao francês: *contre, habitudes, saxophone, tristesse, utile...*; 9 ao espanhol: *camino, similares, probas...*; e, um único informante, 5 ao italiano: *pulir, punta...*), como para criar neologismos (*letra, governmento, suportadores* por *carta, governo* e *apoiantes*). Por outro lado, os neologismos formais continuam a reflectir dificuldades relacionadas com a língua alvo (*azeito* e *problemo*, por *azeite* e

²⁷⁹ Ringbom 1985 analisa dados de finlandeses, que têm o sueco como L2, a aprenderem inglês. «Hybrids, blends and relexifications all provide concrete evidence of **interaction between two foreign languages** in the reader's mind. Hybrids are forms consisting of morphemes from different languages. [...] There are a few occurrences of blends, where a Swedish ending is inserted into what is otherwise an English word [...] In relexifications, a Swedish word has been activated in the learner's mind and then 'relexified' i.e., modified phonologically to make it better conform to what is perceived as English norms» (Ringbom 1985:48).

problema; vegetarianos por vegetarianos) e o contacto pouco frequente ou exclusivamente obtido através de *input* oral (*maloucora por loucura*).

Também os **suecos** não recorrem ao léxico da L1, mas sim, ainda que moderadamente, não só às línguas românicas (ao espanhol: *servício e mi*; e ao francês: *bombe, concours, sud, simple..*) mas também ao inglês (*japanese, portuguese, international e typical*) tanto para empréstimos como para neologismos a partir de empréstimos (*doencia, diferenças, invitar, estrange*). Mas, além da formação de palavras novas, que relevam de questões relacionadas com o português (*dio por dia, metade por metade, tinha por tua*), o que mais caracteriza os dados deste grupo é a criação de um relativamente grande número de verbos da primeira conjugação (*escolhar, proibir, respostar, traduzar, viver por escolher, proibir, responder, traduzir e viver*), seguindo a estratégia da regularização pelo mais frequente e coincidindo com o facto de, actualmente, o português dispôr dessa possibilidade.

Os dados dos **chineses**, se mostram claramente que está afastada a possibilidade de recorrer à sua L1, também mostram que nenhuma outra língua está muito activa. Apenas 8 ocorrências podem ser consideradas como empréstimos do inglês (3 de *example*, 2 de *portuguese*, 1 de *problem* e 1 de *principal* por *director*), havendo apenas a registar um caso que parece ser neologismo a partir de empréstimo do inglês (*estranços por estrangeiros*). De entre os neologismos formais destaco alguns nomes que se prendem com a questão do género (*um fenómeno, a lancha, a genta e a garfa* por *um fenómeno, o lanche, a gente, e o garfo*), sem, no entanto, e tendo em consideração o grande número de desvios verificados no vocalismo átono, descartar a possibilidade de ser esse o motor destas formações. A forma *implorador* por *importador* pode, em contrapartida, radicar em questões relativas à aquisição da sílaba com ataque ramificado e ao traço de vozeamento. O verbo *maiorar* (por *aumentar*) mostra a capacidade para gerar verbos da primeira conjugação, mas o verbo *vistar* (por *visitar*) deixa mais uma vez a dúvida: trata-se de uma formação a partir de 'visto' ou de uma questão fonético-fonológica?

Estes dados desviantes de verbos da primeira conjugação têm o mérito de nos mostrar que certos processos relativos à formação de palavras, que constituem conhecimento procedimental, começam a ser

adquiridos. Embora as regras que governam os processos de formação de palavras em L1 restrinjam as hipóteses que os sujeitos põem acerca das regras lexicais de L2 (Hancin-Bhatt & Nagy 1994), todos parecem ter descoberto a possibilidade de formar verbos de tema em **-a** (*emborar, respostar, traduzar, maiorar...*), o que, por outro lado, nos pode levar a concluir que não só a produtividade da regra foi adquirida mas também que ela conduz a um processo de sobre-generalização ²⁸⁰.

Creio, no entanto, que devemos ser cautelosos quanto à extensão deste conhecimento procedimental. Graça Rio-Torto 1996 chama a atenção para a importância da história da língua na formação de palavras e afirma que certas palavras não são produtos derivacionais do português, mas sim palavras prefixadas no latim e preservadas em português. O mesmo acontece com as outras línguas românicas e com outras línguas herdeiras do latim, nomeadamente as germânicas. Esta observação faz-nos pensar que, perante os dados, temos de ser muito cuidadosos ao pormos a hipótese de os aprendentes terem adquirido determinados mecanismos para formar derivados. Em muitos casos, eles limitam-se certamente a transpor para a interlíngua uma forma correspondente da L1 ou de outra língua que conheçam.

No que respeita a morfologia nominal, é Marta Baralo (1997:69) quem faz uma observação muito pertinente, ao distinguir entre aquilo que é parametrizado, e que pode ser adquirido em L2, e aquilo que é idiossincrático, e que escapa à competência do não nativo até mesmo quando este conseguiu chegar a um domínio bastante profundo da L2. Nominalizações parametrizadas são, por exemplo, nomes deverbais, como *geração* ou *evolução*, formados com o sufixo *-ção*, enquanto que *destino*, *exagero*, *consumo*, *desaparecimento* e *divertimento* são formações idiossincráticas. De facto, a amostra espanhola forneceu-nos *exageração*, a francesa *consumação*, a alemã *desaparição*, a sueca *divertição*, e a chinesa

²⁸⁰ «It is not within the scope of this paper to investigate the various factors facilitating or hindering the acquisition of this process, such as L1 influence and the TL complexity. However, it is interesting to note that learners coming from different language groups exhibited similar preferences on the elicitation tasks. It is reasonable to assume, therefore, that at the advanced level of acquisition that this study was investigating, L1 influence in the application of TL word formation is marginal» (Olshtain 1987:231). Creio que o mesmo se pode dizer no nosso caso.

destinação. Com efeito, **a distinção entre as diversas possibilidades que uma dada língua oferece e a aplicação caso a caso que os seus falantes fazem delas é responsável por um grande número de desvios, sobretudo quando se trata de falantes de línguas muito próximas** (e até mesmo de falantes nativos ...) ²⁸¹.

6.3.2.2. *Combinatórias*

Por último, na mesma *TABELA 13*, apresentam-se as ocorrências de **combinatórias aproximadas**, que resultam de processos de adição de vocábulos da língua alvo.

Numa perspectiva diacrónica, em muitos casos, o que parece ter acontecido foi a associação de um significado específico a uma sequência frequente, cuja interpretação se situa num *continuum* que vai de mais literal a mais idiomática. E, tal como acontece com o vocabulário em geral, o que caracteriza as combinatórias aqui consideradas é a sua grande heterogeneidade. Elas tanto podem codificar categorias nocionais (tempo, espaço, causa, consequência...) como funcionais (saudar, agradecer, despedir-se, fazer um pedido...). O que há de comum entre elas é o facto de todas resultarem de processos aditivos. Do ponto de vista do aprendente, em muitos casos, já não se trata simplesmente de a um dado vocábulo associar um significado, mas sim de adquirir uma sequência com um certo grau de arbitrariedade, gramaticalidade e idiomaticidade, em muitos casos, tanto numa perspectiva paradigmática como sintagmática, muito próxima de outra ou outras.

Michael Lewis 1997b diz ser possível identificar quatro tipos de itens lexicais ²⁸²: «Type 1: a. *Words* b. *Polywords*; Type 2: *Collocations*; Type 3: *Institutionalised utterances*; Type 4: *Sentences frames and heads*.» **Words**,

²⁸¹ A prová-lo estão todos os documentos informativos ou explicativos, traduzidos para português, que acompanham produtos estrangeiros, entre os quais se destacam os documentários apresentados nos canais de televisão por cabo.

²⁸² A primeira classificação surge em Nattinger 1980 e compreende seis tipos de **“lexical phrases”**: *polywords*, *phrasal constraints*, *deictic locutions*, *sentence builders*, *situational utterances*, *verbantim texts*; Nattinger & DeCarrico 1992/1997 distinguem quatro tipos: *polywords*, *institutionalized expressions*, *phrasal constraints*, e *sentence builders*. Pereira 1994 e Nascimento & Pereira 1995 falam de *grupos cristalizados*, *semicristalizados*, e *co-ocorrências privilegiadas*.

segundo este autor, foram sempre reconhecidas como unidades independentes. Uma frase ou um enunciado pode mudar completamente se se mudar uma palavra (*Could you pass my pen/calculator, please?*). Palavras simples podem aparecer como itens independentes (*Stop!, Sure!, Please!*). Apesar de todos os itens lexicais, excepto *words*, serem constituídos por mais do que uma palavra, aqui o termo **polywords** restringe-se a frases curtas que têm um certo grau de idiomaticidade (*by the way, on the other hand*), e que aparecem até mesmo em dicionários de bolso. **Collocations** são pares ou grupos de palavras que co-ocorrem muito frequentemente, dependendo disso do tipo de textos. Embora possam ser formadas por mais do que duas palavras e pela maior parte das categorias gramaticais, são, em muitos casos, pares formados por um verbo e por um nome, ou por um adjectivo e um nome (*to raise capital, a short-term strategy*). **Institutionalised utterances** são mais típicas do oral do que da escrita, e, geralmente, desempenham uma função mais pragmática do que referencial; podem ser frases completas (*I'll get it. There's a call for you.*) ou cabeças (*If I were you, I'd...*). Muitas delas contêm uma alta proporção de palavras deslexicalizadas assim como um alto grau de idiomaticidade. **Sentences frames and heads** são o equivalente na escrita das *institutionalised utterances*. Podem ser curtas e simples (*secondly... and finally*) ou mais longas (*We come now to a number of important reservations...*). São estas últimas que nos servem de orientação para descodificar textos longos ou que nos permitem estruturar e articular os nossos próprios textos ²⁸³.

Para a análise dos dados que se segue, usarei esta proposta de classificação de Lewis 1997b com ligeiras adaptações inspiradas em Nattinger & DeCarrico 1992/1997 e em Pereira 1994 e Nascimento & Pereira 1995. Assim, correspondendo a *polywords*, falarei de **sequências cristalizadas** (*ainda que, com certeza, hoje em dia, mais ou menos, mesmo que, por causa de, por exemplo, por isso...*). Correspondendo a *collocations*,

²⁸³ Sobre **sentences frames and heads** diz Lewis 1997:259: «The frustration of reading a student's essay and thinking 'I know what you mean, but that's not the way to say (=write) it', is most frequently caused by the student's failure to use this type of lexical item. [...] This type of lexical item remains outside the field of competence even of many native speakers; but it can be of great use to, for example, students of academic or professional English.» Ver *Introdução*.

distinguirei entre **sequências semicristalizadas**, ou seja, aquelas que admitem variação de itens dentro de uma estrutura (cf. *phrasal constraints* de Nattinger & DeCarrico 1992/1997) que se mantém constante (*todos os dias / meses / anos; há oito / quinze dias / horas / minutos / muito tempo; à / ao noite / tarde / sábado / meio-dia; de manhã / tarde / noite; dois / três meses/ dias antes/depois; à espera / procura, ...*), ou em que um item se mantém constante e a estrutura varia (*às / por / outras / muitas vezes,...*); **co-ocorrências lexicais** (*animais domésticos, aquecimento central, batatas fritas, velhos amigos, contactos sociais, comida portuguesa, situação económica, sociedade multicultural, totalmente diferente, dar um passeio, fazer anos, gastar dinheiro, passar férias,...*)²⁸⁴; e **cabeças de frase** (*acho que; penso / pensava que; espero / esperava que; não sei se; ainda bem que; gosto muito de/que; quando uma pessoa...*) (cf. *sentence builders* de Nattinger & DeCarrico 1992/1997). A *institutionalised utterances* chamarei **fórmulas**, independentemente de só serem usadas na oralidade ou também na escrita (*faz favor; muito obrigado; que pena!; até à próxima; um abraço; obrigado pela tua carta; tenho muitas saudades tuas; eu nem queria acreditar!*). *Sentence frames and heads* designarei por **estruturadores textuais**, quer sejam mais frequentes na escrita, na escrita oralizada ou na oralidade (*antes de mais/ em primeiro lugar / para começar... em seguida / para continuar... por último / para finalizar / para terminar...; por um lado... por outro lado; não só... mas também...*).

Observemos agora os dados, não sem antes fazer uma pequena observação. Do facto de certos tipos de itens lexicais apresentarem menos desvios do que outros não se pode concluir, como é evidente, que eles apresentem menos dificuldades do ponto de vista da aquisição. Significa sim que, por razões que se prendem com as características da amostra e com o estágio de aquisição dos aprendentes, eles foram muito menos usados. Por essas razões, espera-se encontrar um número relativamente pequeno de ocorrências desviantes *de fórmulas, estruturadores textuais, cabeças de frase e co-ocorrências lexicais*; de *fórmulas* porque, embora tenham sido usadas algumas, em particular nas cartas, como vimos, elas

²⁸⁴ As regências, que Pereira 1994 classifica como *co-ocorrências gramaticais*, serão incluídas em *Desvios na selecção do léxico*.

são mais típicas da oralidade; de *estruturadores textuais* porque, por um lado, trata-se de textos relativamente curtos, por outro, porque neste estágio de aquisição é provável que os aprendentes não tentem usá-los; de *cabeças de frase* porque os aprendentes parecem concentrar-se num número bastante restrito deste tipo de itens que usam repetidas vezes; quanto às *co-ocorrências lexicais* é também provável que o tipo de texto produzido e o estágio de aquisição não seja muito propiciador do uso de itens deste tipo, para além das formadas por nome e adjetivo. Restam as *sequências cristalizadas* e *semi-cristalizadas* que, como se pode ver pelos exemplos acima, têm função referencial e codificam categorias nocionais, muito em especial, a localização temporal ²⁸⁵.

Tal como se esperava, o maior número de ocorrências desviantes de **combinatórias aproximadas** são *sequências cristalizadas* e *semi-cristalizadas*. É nestes dois tipos de itens lexicais que se concentra o maior número de ocorrências de combinatórias formalmente aproximadas ou usadas indevidamente. Elas estão relacionadas com diversas áreas semânticas e exprimem relações de causa/razão/condição – consequência, relações pessoais de reciprocidade, relações espaciais, mas são, sobretudo, conectores de base adverbial, nominal, e prepositiva, ou locuções que exprimem relações temporais. Vejamos alguns exemplos de cada um dos *subcorpora*. Do espanhol: *a esperar* (à espera), *senão que* (a não ser que), *nas noites* (à noite), *com tal de* (de modo a), *desde agora* (desde já), *em tanto que* (enquanto que), *na veira* (perto de), *em todos partes* (por todos os lados); do francês: *ao fim* (afinal), *até fora* (até à rua), *mais e mais* (cada vez mais), *em fim* (finalmente), *juntos* (uns com os outros), *ir e volta* (ida e volta); do Alemão: *às manhãs* (de manhã), *seja que for* (seja como for), *um outros* (uns aos outros); do sueco: *de pé* (a pé), *em ponto* (a horas), *no domingo* (ao domingo), *como primeiro* (em primeiro lugar), *50 anos atrás* (há 50 anos), *na próxima semana* (na semana seguinte); do chinês: *não como* (ao contrário

²⁸⁵ Muitos dos documentos produzidos pelos informantes e incluídos no *corpus* são narrativos. Kumpf (1984:133) afirma que a conversação e a narrativa existem como modalidades discursivas básicas em todas as línguas naturais e que a narrativa é particularmente interessante, sendo a melhor fonte de informação para o estudo do Tempo e do Aspecto, enquanto que a conversação, embora o possa ser também, é mais abundante em formas de ‘irrealis’, sendo por isso mais adequado a estudos sobre o Modo.

de), *todas maneiras* (de qualquer maneira), *só pouco tempo* (há pouco tempo), etc.

Mas de facto, as **combinatórias que codificam relações temporais** representam aproximadamente 50 % do total de ocorrências de combinatórias aproximadas.²⁸⁶ No *sub-corpus* espanhol representam 46,4 %, no francês 38,8 %, no alemão 56,8 %, no sueco 49,4 % e no chinês 51,4 %. A observação dos exemplos acima torna evidente a enorme variedade de sequências básicas deste tipo e a importância que elas têm no discurso. Neste estágio de aquisição, em que os aprendentes evidenciam uma grande apetência pela narração de episódios ou histórias curtas, a deixis temporal surge como um dos factores estruturantes do seu processo de comunicação.

As restantes sequências desviantes são algumas **fórmulas** do oral pouco frequentes (ALEM: *Toda a gente levantem os braços!* por *Mãos ao ar!*), ou muito automatizadas na L1 (ESP: *Até pronto!* por *Até breve!*), ou são remate de cartas (FRAN: *Vou parar...* por *Vou terminar...*; ESP: *Subscrevo-me em baixo*); e são também uma ou outra **cabeça de frase** (ALEM: *que tu sabes*, ou *como tu sabes* por *como sabes*; *ouvi que* e *ouvi de* por *ouvi dizer que* e *ouvi falar de*). São igualmente **co-ocorrências lexicais** de baixa frequência (ESP: *carretão de bois* por *carro de bois*; FR: *escola-primeira* por *escola primária*; ALEM: *lojas de ecologia* por *lojas de produtos dietéticos*; SUEC: *anjo de protecção* por *anjo da guarda*, *cartão de conduzir* por *carta de condução*, *casa de estacionamento* por *parque de estacionamento*; *casa de ladrões* por *prisão*; CHIN: *a velha palavra* por *provérbio*); estas últimas envolvem já, segundo Poullisse 1993, estratégias de reconceptualização.

Há ainda um certo número de desvios, mas, em alguns casos, com um número relativamente grande de ocorrências, em **sequências muito frequentes, mas marcadas** dentro da língua portuguesa. É o caso de *ir-se embora* (*ir-se*, *fugir embora*, e também *emborar*) *tenho saudades tuas* (*tenho saudades de ti*), *um amigo meu* (*um amigo de mim*, *um amigo de me*); e de *à espera* (*a esperar*, *a espera*) e *à procura* (*a procurar*). Estas últimas juntamente com as locuções temporais *há x* (*há muito tempo*, *há dois*

²⁸⁶ Das 153 combinatórias registadas na lista de 2217 palavras do PF, os espanhóis produziram 37 entradas de combinatórias aproximadas, os franceses 38, os alemães 42, os suecos 41, e os chineses 64.

meses...), *x depois (uma semana depois, dois dias depois...)*, e *x passado (o mês passado, o ano passado)* parecem constituir, de entre os usados neste estádio, os itens de mais difícil aquisição.

6.3.2.3. Conclusão

Da observação dos dados analisados relativos a **vocabulário indisponível**, podemos concluir que:

1. quanto mais uma língua é próxima da L2 mais ela é a fonte preferida para empréstimos; de facto, não só o grupo falante de espanhol é aquele que mais recorre a esta estratégia de substituição, mas também os outros grupos recorrem preferencialmente a esta língua; este é um fenómeno altamente sujeito a variação e muito dependente da situação, sendo, por isso, um fenómeno relacionado com a aprendizagem mas também com a *performance*;

2. os aprendentes parecem estar conscientes da distância linguística, e, à medida que ela vai aumentando, vão-se movendo da estratégia de **substituição** para a de **substituição + estratégia**; este movimento envolve não só a substituição de empréstimos por neologismos a partir de empréstimos, mas também uma preferência crescente por material da L2 que se consubstancia num aumento de neologismos formais;

3. as combinatórias mais sujeitas a desvios são aquelas que, de um ponto de vista paradigmático, têm mais conexões com outras e em relação às quais é mais difícil usar estratégias de evitação; ou seja, as sequências cristalizadas e semi-cristalizadas, que, além de, por vezes, terem um alto grau de idiomaticidade (ver, por exemplo, *à noite, de manhã, ao sábado*), também concorrem com outras formalmente muito semelhantes (*à noite/de noite, de manhã/amanhã*); de facto, em muitos casos, trata-se de uma estratégia de substituição que tem na origem, não a indisponibilidade do item, mas provavelmente uma dificuldade em

estabelecer as suas propriedades semânticas (ver, por exemplo, *a pé/de pé*).

4. factores linguísticos intralexicais, nomeadamente a marcação, afectam a aquisição (ver nomes deverbais idiossincráticos).

5. mais uma vez, itens lexicais muito frequentes reúnem uma boa percentagem de desvios; isto aplica-se tanto a itens simples (ver dificuldade de controle, por parte dos espanhóis, de vocábulos de alta frequência) como a combinatórias (ver sequências cristalizadas e semi-cristalizadas).

Estas são algumas conclusões a que este parágrafo nos permite chegar e que se vêm juntar às mencionadas no final do parágrafo anterior. Mas, antes de avançar, gostaria de, neste ponto, fazer um balanço. Para isso, observemos a *TABELA 14* que dá conta do total de desvios analisados até aqui: desvios formais em vocabulário disponível, e desvios resultantes de estratégias para compensar vocabulário indisponível.

Tabela 14

	espanhol	francês	alemão	sueco	chinês
desvios formais em vocábulos disponíveis	53,9	41,6	33,9	30,3	20,4
desvios resultantes de indisponibilidade lexical	15,5	15	9,9	8,3	7,6
% total	69,4	56,6	43,8	38,6	28

O traçado da curva de desvios em função da distância linguística é muito nítido. Enquanto que os espanhóis totalizam já quase 70 % do total de desvios anotados no *corpus*, os chineses, no outro extremo, ainda não chegam aos 30 %. Isto significa, entre outras coisas, que **quanto mais próxima é a L1 do aprendente da língua em aquisição mais ele recorre a estratégias de substituição e de substituição + estratégia e, em compensação, menos a estratégias de reconceptualização**. Ora, como faz notar Poulisse (1993), são estas últimas que implicam mudanças na mensagem pré-verbal e que, por serem mais exigentes do ponto de vista do

processamento, retardam o débito linguístico e podem afectar consideravelmente a estrutura do enunciado ou da frase.

Por outro lado, tendo em consideração que, quase na totalidade, desvios formais em vocabulário disponível e desvios resultantes de estratégias para compensar vocabulário indisponível não interferem em relações sintagmáticas, na estrutura sintáctica e em aspectos semânticos, e, por conseguinte, afectam exclusivamente o item desviante mas não o enunciado ou a frase, fica clara a diferença entre a interlíngua dos vários grupos linguísticos. Ou seja, **à medida que a distância linguística entre a L1 e a L2 vai aumentando os desvios vão afectando mais a componente semântica e, conseqüentemente, as relações sintagmáticas e sintácticas do léxico.** E é por isso que, no caso das línguas românicas, já tratámos uma boa parte dos desvios, no caso das germânicas, ainda não chegámos a meio da tarefa, e no caso dos chineses, está quase tudo por fazer...

6.3.3. *Desvios na selecção do léxico*

No capítulo 2, reflectimos sobre aquilo que, no léxico mental, está representado numa entrada lexical. Vimos que cada entrada é bastante rica e que estabelece uma rede de relações paradigmáticas e sintagmáticas com outras, criando uma poderosa rede de conexões. No caso de falantes de mais de uma língua, temos ainda de considerar as conexões entre os léxicos das duas línguas, ou eventualmente mais, conexões essas que dependem das relações tipológicas entre as línguas, do nível de proficiência que o falante tem na(s) língua(s) desfavorecida(s) e também da idade em que foi ou foram aprendidas.

Dentro de um sistema linguístico, as fronteiras semânticas de uma entrada são definidas pelas suas relações com outras. Isto significa, entre outras coisas, que, no âmbito de uma mesma língua, em função da idade, do grupo sociolinguístico do falante, e até mesmo de idiossincrasias pessoais, essas fronteiras podem apresentar diferenças, geralmente pouco acentuadas, e que a introdução de mais uma entrada no sistema conduzirá à reestruturação deste. No caso da aprendizagem de uma L2, em muitos casos, não só o sistema da L1 e da L2 não são coincidentes no que respeita essas fronteiras, o que implica a definição de novas fronteiras semânticas

dentro de cada micro-sistema, mas pode implicar também uma reestruturação da estrutura conceptual associada à L1.

Acontece que, durante o processo de reestruturação, certas propriedades semânticas que o aprendente atribui a um dado item podem apresentar um ou mais desvios no que respeita o seu significado referencial, extensões metafóricas, e adequação pragmática. Esses desvios resultam do cruzamento entre relações paradigmáticas e relações sintagmáticas das entradas, inter e intralexicais. Eles fazem-se sentir, com certeza, sobretudo nos usos menos prototípicos e estão relacionados, em parte, com as diferentes escolhas que cada língua faz para expressar sentidos idiomáticos. Mas, se aceitarmos que a **idiomaticidade** é um fenómeno básico e gradual na linguagem, como propõe a linguística cognitiva, fica de parte a tentativa de estabelecer uma fronteira entre significado referencial, extensões metafóricas, adequação pragmática, valores afectivos e adequação pragmática.

Por outro lado, também parece razoável admitir, como faz Goldberg 1995, que, no nosso léxico mental, não existe uma divisão estrita entre **léxico e construções**, e que estas relacionam o plano semântico com o plano sintáctico. Isto significa, no caso da L2, que o aprendente, no momento de formular a mensagem, activa uma determinada entrada que tem conexões paradigmáticas e sintagmáticas com outras da L2 e, certamente também, com outras da L1 (ou de outra língua conhecida). Assim, o desvio pode manifestar-se não só no vocábulo em causa mas também na construção de que ele é parte.

Investigação sobre **associação de palavras** tem mostrado que determinados itens lexicais privilegiam as conexões paradigmáticas, enquanto que outros privilegiam as sintagmáticas. À maior parte dos nomes e adjectivos, por exemplo, associamos preferencialmente outros nomes e adjectivos, enquanto que a verbos associamos outros verbos mas também, frequentemente, *chunks* que incluem outras categorias, nomeadamente preposições e ou conjunções. Ou seja, muitos nomes e adjectivos parecem ter mais autonomia no nosso léxico mental, com certeza graças à força do seu conteúdo referencial; enquanto que outras categorias, pela sua menor carga semântica, não só precisam de mais apoios ao nível da representação no nosso léxico mental, associando-se para isso a outras em sequências,

mas, por outro lado, também estão mais disponíveis para a polissemia e para usos menos prototípicos, podendo chegar a desempenhar, em variedades linguísticas bastante elaboradas, funções meramente sintáticas. Algumas preposições (como *a* e *de*) são disto o melhor exemplo.

Em conformidade com isto, contrariamente aquilo que parece sugerir Laufer 1997, na análise dos dados que se segue, não se estabelece uma distinção entre propriedades semânticas de um item lexical, relações paradigmáticas com outros, e o seu comportamento numa frase ou num enunciado. Assume-se que os processos de **reestruturação e reconceptualização**, em direcção à língua alvo, envolvem, necessariamente, a criação de hipóteses desviantes quanto às propriedades semânticas atribuídas aos itens lexicais, que se manifestam na interlíngua, e que podem ser visíveis também a nível sintáctico ²⁸⁷. E, em conformidade com o que ficou dito em capítulos anteriores, **espera encontrar-se um maior número de desvios em itens lexicais mais dessemantizados, mais polissémicos e nos seus usos menos prototípicos.**

Como vimos, por vezes, quando um item lexical não está disponível no vocabulário do falante de L2, ele socorre-se de uma estratégia de substituição que consiste em tomar de empréstimo, de um modo geral inconscientemente, um item de outra língua. Para colmatar essa falta, ele pode também socorrer-se de uma estratégia de aproximação ou de generalização usando um item disponível que faça parte da sua interlíngua e que considere apropriado. No entanto, e tendo em consideração que saber uma palavra envolve saber a sua adequação a vários contextos, um dado item pode fazer parte do vocabulário do aprendente mas, devido à representação que dele tem, não estar disponível para um determinado uso. Nesse caso, ele será substituído por outro, considerado como adequado. O resultado desta estratégia pode confundir-se com estratégias de aproximação ou de generalização. Ou seja, **o facto de um item não ser usado não significa necessariamente que ele não faça parte do**

²⁸⁷ «We work from the assumption that the syntactic properties of a lexical item can largely be predicted from its semantic description. [...] Imagine a mature speaker learning a new word. [...] he is able to predict its syntactic properties on the basis of the semantic specification. [...] There will of course be a few residual exceptions in syntax (items which have idiosyncratic properties, that have to be learnt point-by-point) just as there are often irregularities in morphology» (Dixon 1977: 24-25).

vocabulário do aprendente; significa sim que, caso faça parte, ele não foi considerado como adequado ao contexto em causa.

Como se pode verificar na *TABELA 15*, e como seria de prever, tendo em conta o tempo de estudo e a distância linguística, alguns dos *sub-corpora* apresentam um grande número de desvios, o que inviabiliza, no espaço desta dissertação, uma análise e um comentário muito detalhado da sua totalidade. Por isso, comentarei brevemente os resultados aqui apresentados. Farei um comentário um pouco mais detalhado dos desvios em verbos mais frequentes, uma vez que é à volta dele que se organiza a frase ou o enunciado. E é esta certamente a razão por que a maior parte dos trabalhos de investigação nesta área se tem centrado no verbo. Além disso, e ao contrário do que acontece com outras classes abertas, nomeadamente com nomes e adjectivos, há um pequeno número de verbos bastante frequentes que concentra um grande número de desvios.

Tabela 15

<i>DESVIOS NA SELECÇÃO DO LÉXICO</i>						
	espanhol	francês	alemão	sueco	chinês	Total
em nomes	30	36	55	55	44	220
em adjectivos	13	10	22	28	15	88
em verbos	60	116	149	185	220	730
em preposições	72	109	199	187	193	760
em outras categorias	39	45	72	83	158	397
recategorização	6	10	25	50	54	145
	220	326	522	588	684	2340
	%	16,6	22,9	29,5	28,5	33,1

A *TABELA 15* dá conta da totalidade dos desvios na selecção dos itens lexicais e dos seus reflexos nas construções. À primeira vista, **o que ressalta da leitura da tabela na sua globalidade, é a inversão do sentido dos resultados em relação àquilo que acontecia com desvios formais e vocabulário indisponível.** Embora no âmbito de cada um dos *sub-corpora* as percentagens sejam relativamente próximas, há agora uma grande diferença entre 220 desvios dos espanhóis e os 684 dos chineses. Esses desvios, como a tabela também mostra, têm a sua menor expressão em

nomes e em adjectivos, enquanto que os verbos e as preposições representam 53 % do total dos desvios do *PESTRA* considerados em léxico deficitário.

6.3.3.1. *Desvios nos nomes*

Como se esperava há uma grande dispersão no que respeita o uso desviante de nomes, tanto dentro de um *sub-corpus* como no conjunto do *PESTRA*. É essa provavelmente a razão por que, para além dos tradicionais trabalhos sobre falsos amigos, não se encontram trabalhos de investigação sobre aquisição de nomes ²⁸⁸. Vejamos alguns casos exemplares:

- E17** *Espero ter as **qualidades** necessárias para trabalhar no seu* [habilitações]
sector de fotografia.
- E20** *... algumas vezes tínhamos tempo para fazer compras (livros,* [lenços de seda]
*postais, **lenços-de-seda**).*
- E30** *na cantina universitária [...] formamos fila para comprar o* [senha]
***bilhete**. [...] quando um pobre estudante consegue ao fim ter a*
senha na mão...
- E36** *Uma coisa muito diferente entre os dois países é o horário das* [refeições]
***comidas**.*
- F18** *Fico muito contente de ter recebido as tuas **novidades**.* [notícias]
- F24** *Mênfis, Tebas, Gisé [...] Todos estes nomes são maravilhas para* [ouvidos]
*as minhas **orelhas**.*
- A01** *Mas infelizmente a aparência conta muito e existe um grande* [competição]
***concurso** entre as mulheres de quem é a mais linda, a mais*
rica...
- A18** *... agora as minhas **crianças** são adultos, este tempo já está* [filhos]
passado.
- A12** *Primeiro depois eu levantei-me e houve **lume** na nossa casa, mas* [fogo]
os bombeiros foram rapido.
- S06** *A **cafeteira** tinha três empregados, uma **menina** e dois* [cafeteria/café]
homens... [rapariga]
- S14** *As empresas têm de prometer os empregados novos aumentos de* [ordenado]
***salário**...*
- S44** *É sempre uma **pergunta** de dinheiro* [questão]
- C19** *Esta **despesa** inesperada vem à minha casa.* [despesa]

²⁸⁸ Cohen 1989 é um trabalho não sobre a aquisição, mas sobre a perda de léxico produtivo. O autor estuda duas crianças (9 e 13 anos), falantes de Hebraico e Inglês que, tendo vivido um ano no Brasil, têm o PB como L3. O estudo foi feito nove meses depois de abandonarem o Brasil. Cohen concluiu que o mais jovem perdeu mais palavras do que a sua irmã de 13 anos. Concluiu ainda que ambas perderam mais **nomes** do que itens de qualquer outra classe.

C37 *Beber café que é um **habitante** dos portugueses.*

[hábito]

C31 *Aqui, não como outros países europeu, não há uma **paisagem** [clima ?] depressa de vida, mas está cheio de paz.*

Alguns desvios parecem ser motivados por semelhanças ao nível da forma entre dois vocábulos. São exemplos, lençol por lenço, cafeteira por cafetaria, despensa por despesa, e habitante por hábito. Os dois primeiros são nomes que, embora pertencendo ao mesmo campo semântico, designam entidades cujas fronteiras semânticas são facilmente definíveis; no primeiro caso, trata-se de objectos semelhantes, mas de dimensões e funções diferentes; no segundo caso, cafeteira por cafetaria, é provável que a confusão resulte da dificuldade de estabilização dos ditongos. De qualquer modo, o falante nativo escolheria café. No caso de despensa por despesa, não há qualquer relação semântica entre eles, e é até provável que o aprendente nem sequer conheça a palavra despensa; trata-se, certamente também, de uma confusão, de uma casualidade, provocada também por aspectos formais, ou seja, pela dificuldade na estabilização da nasalidade. Quanto ao último caso, habitante por hábito, pode acontecer que também resulte de uma questão formal: da dificuldade de aquisição do segmento fonológico /t/.

As ocorrências desviantes exemplificadas por bilhete – senha, pergunta – questão, comidas – refeições, novidades – notícias, concurso – competição, crianças – filhos, lume – fogo, menina – rapariga, são as mais frequentes e prendem-se com relações paradigmáticas com outros, e com as suas fronteiras dentro de um mesmo campo semântico. O aprendente usa uma estratégia de aproximação ou de generalização. No entanto, o facto de não usar o item adequado não significa que ele não faça parte do repertório do aprendente. Significa sim que, devido à representação que dele tem, ele não está disponível para o uso em causa ou não é considerado como adequado (veja-se, em E30, senha).

A minha preferência, neste contexto, de ordenado por salário, que pode até não ser a do leitor falante nativo como eu, é um daqueles casos de que falei antes relacionado com fronteiras pouco acentuadas e que depende de idiossincrasias pessoais. O mesmo não se pode dizer do tipo de desvios exemplificado pelo uso do vocábulo paisagem em C31. Aqui trata-se de uma extensão metafórica que a língua portuguesa não escolheu fazer, tendo preferido atribuir esta função a clima, ambiente ou atmosfera, em qualquer dos casos, palavras do campo semântico de ar.

Afirmar antes que há uma grande dispersão no que respeita o uso desviante de nomes. No entanto, há um item, a que já dedicámos a nossa atenção a propósito da atribuição de género aos nomes, e que, mais uma vez, concentra um grande número de desvios. Trata-se de gente, usado como sinónimo de pessoas. Observemos alguns desses usos:

- E1 ... *sem o progresso* *estavamos ainda a caçar o mamute para comer [...] e*
 3 *contestaste que naquele tempo* **a gente** *vivia feliz, com certeza.*
- F42** *As vezes, suportar os vapores dos carros, o barulho e o humor da gente,*
augmentam 'o stress' das *pessoas* *que terminam por fazer depressões.*
- A2 ... **a gente** *alemá e muito hospitaleira e não e preciso ter vergonha de*
 2 *perguntar o camino*
- S4 *Muito antigamente* **a gente** *comiam só para viver.*
 1
- C1 *Os cidades: Porto, Braga etc têm lugares de trabalho.* **O gente** *ganhou dinheiro*
 8 *facil.*

Os exemplos acima são uma das 4 ocorrências desviantes de um total de 9 no *sub-corpus* espanhol, uma das 12 em 20 no francês, uma das 9 em 17 no alemão, uma das 9 em 20 no sueco, e a única em 21 no chinês. Estes desvios resultam certamente de o português, diferentemente das outras línguas românicas, ter especializado este nome para desempenhar uma função que corresponde ao pronome pessoal *nós*, o que constitui uma função [+ marcada] em relação às outras línguas do mesmo grupo. Como vimos, alguns dos aprendentes germânicos têm alguns conhecimentos de outras línguas românicas, enquanto que os chineses não. Isso talvez explique a quase inexistência de erros nos chineses. As restantes ocorrências dos chineses são em sequências indiscutivelmente correctas (*toda a gente, muita gente, pouca gente, e cheio de gente*). De facto, o colectivo *a gente* só tem um valor semelhante ao de outras línguas românicas quando acompanhado de um quantificador.

6.3.3.2. *Desvios nos adjectivos*

Embora sendo uma classes aberta, os adjectivos totalizam um número relativamente baixo de desvios, em relação às outras classes. Isso deve-se provavelmente ao facto de representarem menos de 1/3 de entradas de nomes, em cada um dos *sub-corpora* (ver TABELA 7), já que os aprendentes parecem confiar mais em adjectivos muito frequentes que usam, preferencialmente, com função predicativa.²⁸⁹

²⁸⁹ Destacam-se: *bom* 162, *bonito* 53, *diferente* 79, *grande* 128, *importante* 72, e *melhor* 82. Recordo aqui uma Nota do Capítulo III: «Not all languages have the major word class Adjective....» (Dixon 1977) e «An areal typological characteristic of European languages is that all the four open parts of speech (N, V, ADJ, ADV) are well developed, in the sense that

- E07** *Recordei [...] o ultimo viagem a Coimbra onde passamos um mes **inolvidável**.*
- E27** *...não tinha um **formoso** vestido, nem um formoso penteado.*
- E18** *...eu gostaria de ser directora duma Empresa de hotelaria para poder estar com tantas pessoas de **distintos** países ...*
- F23** *...dormem nas praias e, ao fim das férias, ficam vermelhos como gambás **cozinhad**s.*
- F51** *Mas como geralmente elas não têm ordenado bastante **importante** para comprar uma casa, são obrigadas viver num apartamento barato onde há só um pouco de espace privado.*
- A47** *Quando as pessoas moravam no campo, a vida normalmente não era **luxuriosa**.*
- A44** *Estrangeiros [...] fazem muitas vezes trabalhos que os povos **domesticos** não querem fazer.*
- S20** *As noites eu conduziria o carro com música muito **forte** e com janelas abertas.*
- S41** *Muito antigamente a gente comiam só para sobreviver. E a comida era muito **invariável** e sem pimentos, molhos legumes e era as vezes cru*
- C11** *Quando nós chegamos à paragem [...] viemos a rapariga a pé, tremendo e **rígida**.*
- C33** *Portanto doença do coração, tensão em alto, cancro, aparecem sempre, e é **popular**.*

Estes são alguns exemplos e a eles se aplica o que foi dito para os nomes. Os casos seleccionados do *sub-corpus* espanhol pretendem ilustrar aquilo a que chamaria *selecção improvável* e que só acontece com este grupo, tanto nesta classe como em outras. São escolhas possíveis, mas sentidas como desviantes em relação às dos falantes nativos para o mesmo contexto.

No entanto, ao contrário do que acontece em L1 e em início de aprendizagem informal de L2 (cf. Axelsson 1994), neste caso há uma maior riqueza de adjectivos que pode estar relacionada com o contexto de

they are distinguished formally in a clear way, and that they form large classes with several thousand members.» (Viberg 1993:341)

aprendizagem – formal e em imersão – bem como com o facto de se tratar de material escrito.

6.3.3.3. *Desvios nos verbos*

A TABELA 16 distingue entre os desvios observados em verbos em geral e os

Tabela 16

VERBO	espanhol	francês	alemão	sueco	chinês	Totais
geral	41	48	92	95	128	404
<i>ser, estar e ficar</i>	13	54	37	61	74	218
omissão / inserção do pronome de flexão reflexiva	6	14	15	24	39	98
total	60	116	149	185	241	730

desvios nos verbos *ser, estar, e ficar*. Regista ainda em separado, e não serão comentados aqui, os desvios resultantes da supressão ou, mais raramente, da inclusão, do pronome de flexão reflexivo (*alimentar* por *alimentar-se*, ou *sentar* por *sentar-se*) do pronome inerente (*chamar* por *chamar-se*, ou *tornar* por *tornar-se*) e do anticausativo (*sentir* por *sentir-se*, ou *aborrecer* por *aborrecer-se*).

Como se pode constatar, os desvios em verbos representam mais de 25 % do total dos desvios nas várias categorias. No que respeita os **verbos em geral**, os resultados são bastante claros, se tivermos em consideração a distância linguística: resultados semelhantes entre si nas duas línguas românicas (41 e 48), semelhantes também nas duas línguas germânicas (92 e 95), e relativamente mais elevados no grupo chinês (128). Quanto à selecção de verbos que lexicalizam distinções aspectuais, como *ser, estar, ou ficar*, o grupo espanhol distingue-se claramente de todos os outros, que apresentam resultados relativamente semelhantes entre si. Mas procuremos agora algumas explicações para estes resultados.

Nos Capítulos 1 e 3, a propósito da importância que têm os estudos tipológicos para a investigação em aquisição, referi os trabalhos de Viberg (1993) e a sua proposta de **verbos básicos** (13) elaborada a partir da comparação de onze línguas da Europa. O Quadro 2 abaixo, extraído de

Viberg (1993a:346), dá conta da distinção entre *verbos nucleares*, aqueles que reflectem tendências universais e que são os tipologicamente menos marcados dentro do seu campo semântico, e *verbos específicos* das línguas da Europa (“*areally specific basic verbs*”), o verbo de cópula BE e os modais CAN e MUST, ou, o caso de HAVE, específico apenas de uma parte delas (“*subareally specific basic verb*”) ²⁹⁰.

Quadro 2

Basic verbs in European languages						
A. Reflecting universal tendencies: the nuclear verbs						
<i>Motion</i>	<i>Possession</i>	<i>Production</i>	<i>Verbal communication</i>	<i>Perception</i>	<i>Cognition</i>	<i>Desire</i>
GO (10) COME (8)	GIVE (11) TAKE (11)	MAKE (10)	SAY (11)	SEE (11)	KNOW (9)	WANT (8)
B. Areally specific basic verbs			C. Subareally specific basic verb			
<i>General Dynamic</i>		<i>Modal verbs</i>		<i>Possession / State</i>		
BE (11)		CAN (11) MUST (9)		HAVE (8)		

(The figures within parentheses state the number of languages in which a verb with the given meaning belongs to the twenty most frequent ones. Languages: English, German, Swedish; French, Spanish, Italian, Rumanian; Russian, Polish; Finnish, Hungarian.)

No Capítulo 3, referi o estudo comparativo de Viberg 1993a com crianças de seis anos, falantes de Sueco L1 e de Sueco L2, em que este investigador conclui que **as crianças falantes de L2 tendem a usar mais os verbos nucleares e menos os verbos básicos mais específicos da**

²⁹⁰ «The most frequent verb in practically all European languages is the copula **BE**, which clearly is such an areal-specific characteristic. In many languages outside Europe, 2 noun phrases are simply juxtaposed in equational sentences or linked with a morpheme that lacks many or all verbal characteristics (Li and Thompson, 1977). The existence of a particular group of modal verbs is another important areal-specific characteristic of European languages. As is stated by Palmer (1986), similar meanings are often expressed by bound inflection (mood) or clitics and particles in other languages. The second most frequent verb in a large group of languages in the center of Europe is **HAVE**. But actually, this verb is completely missing from the Celtic languages at the western fringe of Europe and from languages to the east such as Finno-Ugrian languages.» (Viberg 1993a:348).

língua sueca. Se observar o *ANEXO 13*, que dá conta dos verbos mais frequentes no *PESTRA* e no *PF*, verificará que, nos nossos dados, os resultados não são muito claros. De facto, os falantes de português L2 parecem confiar bastante nos verbos nucleares e, em geral, usam-nos com mais frequência do que os falantes de L1; isso é tendencialmente verdade para GO – *ir*, MAKE – *fazer*, SAY – *dizer*, SEE – *ver*, KNOW – *saber*, e WANT – *querer* e TAKE – *tomar*. No caso de GIVE – *dar* os estrangeiros usam sensivelmente menos, e no caso de COME – *vir* usam notoriamente menos. Pelo contrário, os falantes estrangeiros usam abundantemente alguns dos verbos básicos específicos das línguas europeias, como BE – *ser*, CAN – *poder* e HAVE – *ter*, e até mesmo verbos típicos das línguas românicas ibéricas, como *estar* e *ficar*. Se considerarmos *dever* como o correspondente de MUST, ele nem sequer aparece entre os 30 mais frequentes.

Estes resultados, ligeiramente divergentes em relação aos de Viberg (1993), estão certamente relacionados, antes de mais, com o diferente desenvolvimento cognitivo de crianças e de adultos. Mas, para termos uma visão mais correcta da questão, relacionemos o número total de ocorrências com o número de ocorrências desviantes nos 20 mais frequentes verbos no *PF*, ou seja, nos verbos básicos da nossa língua.

A *TABELA 17* mostra-nos claramente que a maior parte dos desvios não se concentra nos 10 verbos que ocupam a segunda metade da tabela. Na verdade, **os 10 mais frequentes verbos no *PESTRA* são suficientes para absorver 364 dos 397, ou seja, a quase totalidade dos desvios nos 20 mais frequentes verbos no *PF*.** Esses 10 mais frequentes, tanto no *PESTRA* como no *PF*, são quase todos ***verbos específicos das línguas da Europa ou específicos das línguas românicas ibéricas***: os modais, o par *ter/haver*, e os aspectuais, muito em especial a oposição aspectual *ser/estar*, específica das línguas românicas ibéricas. Entre os 10 seguintes incluem-se alguns verbos de movimento. Como vimos no Capítulo 1, as línguas românicas ibéricas seguem para estes um diferente padrão de fusão de significados que não só as distingue das línguas germânicas como

Desvios nos 20 mais frequentes verbos no PF								
VERBO	espanhol	francês	alemão	sueco	chinês	Total	Nº no PESTRA	Nº no PF
ser	2	19	23	19	32	78	1	1
ter / ter de	2	2	6	14	22	46	2	2
estar	8	26	12	25	19	89	3	3
poder	0	1	6	2	4	13	4	10
fazer	3	0	5	4	10	22	5	7
haver	3	3	3	2	2	13	6	6
ir	2	2	4	2	6	16	7	5
gostar	2	1	1	1	3	8	8	14
ficar	6	7	2	8	21	40	9	15
querer	0	0	5	11	2	18	10	8
saber	0	0	1	1	2	4	11	9
achar	0	0	0	0	0	0	12	16
dizer	1	1	1	1	0	4	13	4
ver	0	0	3	0	0	3	15	13
chegar	0	0	1	2	0	3	16	19
trabalhar	0	0	0	0	0	0	21	17
dar	2	0	0	1	1	5	23	12
começar	0	0	0	1	0	1	30	20
vir	0	3	0	1	2	6	40	11
andar	0	1	2	1	3	7	64	18
	31	66	75	96	229	397		

as coloca numa categoria à parte dentro das línguas indo-europeias. Essa diferença é, por sua vez, também geradora de um certo número de desvios, nos *sub-corpora* germânicos e chinês (ver Talmy, em 1.3.1.).

Começemos por observar alguns desvios no uso de *ser*, *estar* e *ficar*.

- E13** *Eu perguntei-te como podias **ser** tão seguro disso, e respondeste que não podiam ser infelizes por causa de uma coisa que não conhecem.*
- E25** *Estive na festa até as quatro da manhã. Depois fomos a tomar o pequeno almoço num café que **está** no Saldanha.*
- E33** *Uma das coisas que mais me impressionou quando cheguei a Portugal (e **ainda me está a impressionar**), foi a capacidade que os portugueses têm para fazer filas organizadas.*
- F09** *Lembras-te como nós **estávamos** boas amigas. Nós fazíamos tudo juntas. Nós **éramos** na mesma turma durante seis anos , nós íamos juntas ao cinema ...*
- F35** *Mas, parece-se que comer mais em Portugal do que na Suíça **está** uma festa. É verdade, comer parece ocupar um lugar mais importante na vida dos portugueses.*
- F17** *Gostávamos de estudar a psicologia. O que gostava o mais, **estava** fazer inqueritos na rua: **estava** na paragem do autocarro e tinha de observar como as pessoas atravessaram a rua. Para ti, tinhas de olhar as pessoas e de ver*

como **estavam a** utilizar as novas máquinas do metro.

- F45** O problema de trânsito **fica** hoje uma dentro as maiores preocupações dos políticos de todos os países da Europa. [...] Lisboa **fica** uma cidade antiga, velha e a maior parte das suas ruas são estreitas e não podem suportar todos aqueles novos carros!
- A11** Ainda bem que tu esperaste mas é uma pena que nos não tivemos suficiente tempo para converçar. É verdade que eu **era** com muita vontade a ver ti outra vez....
- A25** No último mês vi 'grande festival das tunas internacionais. [...] Os meus amigos portugueses **ficavam** vestidos com fatos. As mantas eram pretas e as camisas eram brancas. [...] **Era** uma grande bicha em frente do Coliseu mas finalmente conseguíamos entrar.
- A33** ... os portugueses não são muito pontual. [...] Agora já me habituo com pessoas atrasados e também não **estou** pontual.
- S10** Agora estou sentado numa café em Estocolmo...[...] **Fica** muito estranho que mudar numa país que está muito diferente do que Portugal
- S34** Eu acho que as pessoas gordos **estão** mais felizes do que as pessoas magros por isso praticamente todas as pessoas sérios **estão** magros também.
- S26** À duas semanas atrás vi um filme que se chama 'Em Nome do Pai'. Este filme **estava** um dos melhores que eu tinha visto em toda a minha vida.
- S31** Já havia uma fila de aproximadamente 50 pessoas que estavam a esperar. Esperávamos e esperávamos. [...] **Éramos** fartos de esperar!
- C01** ... tinha um amigo que quase nunca estudava mas passou o exame porque no dia anterior ao exame, ele **ficava** aborrecido, portanto, abriu o livro e estudava só aquele tema.
- C12** Eu soube o Marco Paulo **estava** muito famoso, tinha um programa chamou-se "Tenho dois amores".
- C20** ... agora também não tenho muito dinheiro para se emprestar, muito desculpe! Acho que possa pedir ao J., ele **está** solteiro e nos últimos anos tem trabalhado em Macãu.

O facto de a distinção *ser/estar* ser específica das línguas românicas ibéricas pode explicar a dificuldade sentida pelos franceses. O facto de as línguas germânicas não só não terem esta distinção mas privilegiarem o **tempo** e não o **aspecto** justifica a dificuldade sentida por alemães e suecos, dificuldade essa que se estende também aos casos em que o aspecto é codificado na morfologia, em especial à oposição perfectivo / imperfectivo. Este problema, como é sabido, persiste até estádios de aquisição muito avançados em falantes deste grupo de línguas (ver, mais adiante, em 6.3.4 *O léxico para além do léxico*). Quanto aos chineses, se relacionarmos os

desvios com a frequência (ver ANEXO 13), veremos que usam ligeiramente menos *ser* e *estar* (mas mais *ficar*) mas que não apresentam um número maior de desvios nestes verbos. Estes resultados podem ter duas explicações. Por um lado, a especificidade de *ser* enquanto verbo das línguas da Europa explica provavelmente o menor número de ocorrências. Por outro lado, o facto de o chinês ser uma língua que privilegia o **aspecto** e

C32 *Um português com um chinês de ver o mundo é defrente. O português no seu país é muito livradade, sabe as coisas do mundo é muito. Mas o chinês no seu país só tem pouco livradade. Sabe as coisas do mundo também só pouco. Por isso o português vê o mundo é muito pequeno e fácil.*

C37 *Eu sempre dormi na sala de aula, antigamente, porque não gostei essa lição. [...] Agora quando tenho sono, bebo um pequeno café. Depois pouco tempo, **sem sono**, é verdade!*

C46 *Eu moro muito longe da universidade, portanto, tenho que levantar-me muito cedo, porque tenho que gastar sobre uma hora de minha casa para a universidade, tenho medo de chegar à aula tarde. Carros, carros, **muito chato!***

não o tempo explica com certeza o baixo número de desvios no que respeita até mesmo a selecção de verbos como estes, ou seja, que lexicalizam o aspecto, e que são específicos de um sub-grupo de línguas ²⁹¹.

Na verdade, dos 74 desvios advertidos nestes três verbos, 53 dizem respeito a desvios na selecção de *ser*, *estar* ou *ficar*. Apesar da dificuldade que constitui interpretar produções tão desviantes como aquelas em que ocorrem desvios que podem ser considerados como resultantes da omissão ou da inserção abusiva destes verbos, dos 21 desvios considerados como tal, apenas 5 são inserção inadequada de *ser*, e os outros 16 são 12 omissões das formas *é* ou *ser* e 4 da forma *fica*. No entanto, esta dificuldade que, como seria de esperar, se manifesta exclusivamente no *sub-corpus* chinês, parece ser rapidamente ultrapassada, visto que este tipo de desvios surge sobretudo em documentos também muito desviantes a outros níveis, mas não em produções que denotam uma maior competência ²⁹².

²⁹¹ Não me alongarei mais sobre a questão do **aspecto verbal** e sobre as diferentes perspectivas adoptadas por cada uma das línguas representadas no *PESTRA* por tê-la tratado em Leiria 1991. Ver também 6.3.4. *O léxico para além do léxico*.

²⁹² Em trabalhos futuros, como referirei mais adiante, conto estudar que tipos de desvios surgem em cada documento e, a partir daí, procurarei abordar a questão dos **estádios de aquisição**.

Os desvios no âmbito dos **verbos modais** parecem estar sobretudo relacionados com a presença, no espaço dos hiperónimos *MUST* e *CAN*, de outros verbos que disputam o espaço semântico, respectivamente, a *dever* e *poder*. No primeiro caso, isso prende-se com a existência neste sub-sistema, em português e espanhol, e ao contrário do que acontece no par francês – italiano, de *ter de / que* a disputar terreno com o verbo *dever*. E é exactamente por isso que, como se pode ver no *ANEXO 13*, este verbo não está entre os 13 mais frequentes, mas ocupa a vigésima quarta posição no *PF* e a trigésima primeira no *PESTRA* ²⁹³. No espaço de *CAN*, o quadro complica-se ainda mais, porque a *poder* fazem concorrência *conseguir*, *ser capaz de* e *saber*.

Como se pode ver nos exemplos da página seguinte, os desvios dos franceses e alemães resultam, de facto, desta difícil partilha entre *dever* e

- E23** *Tiveste sorte! A tua carta chegou em quatro dias, o que **tem de** ser incrível para os nossos países.*
- F21** *Por hoje, é tudo. Dentro de alguns minutos, são horas de trabalhar. **Devo** acabar a carta.*
- F23** *Ao ano passado passei 2 semanas em Paris para visitar a torre Eifel, o museo Le Louvre... voltei mais cansada do que quando comencei as férias. E para lembrar-me do que visitei, **devo** ler os livros que comprei.*
- A08** *Neste momento, tenho uma namorada bonita, com um rosto aberto e uma beleza natureza. Lembras-te na N. Ela é actriz no teatro e <por> <consegue> **pode** representar muito bem*
- A43** *... na Alemaha muitos jovens querian estudar e depois muitas pessoas estão desemprego. Meu irmão [...] vai fazer um outro curso, mas não recebe dinheiro. Meus pais **devem** pagar o curso.*
- S14** *... as vezes os jovens já têm todas as coisas necessárias porque os pais deram-lhes antes. Estes jovens costuma de agir um pouco egocênticos. Eles nunca têm suficiente. Gostam de mostrar as suas existências luxas mas “não **temos de** esquecer que trabalharam muito para ter estas vidas”.*
- S27** *E agora os problemas começou. Ele pensou que Deus queria o matar mas não **podia**. Ele achou que era imortal, portanto o titulo ‘Sem medo de viver’.*
- C07** *Quando eu vim Lisboa, não conheço nada e não **pude** falar português...*
- C37** *Em Macau ou na China, pouca gente gosta de beber café. Quando cheguei a Portugal, no primeiro mês, vi muitos portugueses que bebiam café no qualquer tempo. Muito estranho, sentava, eles não têm medo de cafeína. À noite eles **podem** dormir?*

²⁹³ De qualquer modo, as 12 ocorrências de *dever* no *sub-corpus* português têm valor deontico e não epistémico. O mesmo acontece com a quase totalidade das ocorrências no *PESTRA*.

- C44** *Já em Lisboa o trânsito é muito muito caótico, por isso o autocarro não consegue continuar pelo caminho. [...] Ele tem que tomar o metro para Rossio, logo que ele entra a estação do metro, não **pode** ver o chão da estação, só **pode** olhar as cabeças dos passageiros.*

ter de / que. No caso dos suecos, como mostra o exemplo, estes podem até invadir o espaço de *dever* e *poder*. Quanto aos chineses não foi possível encontrar exemplos de desvios no espaço de *MUST*. De facto, há 10 ocorrências de *dever* e 22 de *ter de/que* sem erros. Os 22 desvios contabilizados na tabela em *ter, ter de/que* resultam de problemas no par *ter - haver*. Como se sabe, este último é objecto frequente de estratégias de evitação que se manifestam até em falantes nativos de outras variedades do português, como é o caso do PB.

O menor número de ocorrências de *vir*, que o coloca em quadragésimo lugar no *PESTRA*, enquanto que no *PF* ele ocupa a décima primeira posição, pode dever-se a dois factores. Os verbos *GO (ir)* e *COME (vir)* são ambos verbos deícticos de movimento, ocupando assim o mesmo campo semântico. Além disso, o português, quando comparado com outras línguas europeias, respeita muito mais o seu uso prototípico ²⁹⁴, o que, se por um lado reduz os contextos de uso de *vir*, por outro cria no aprendente a desconfiança em relação a este mesmo verbo.

- A12** *Ao dia ontem houve algumas coisas inesperadas. Primeiro depois eu levantei-me houve lume na nossa casa, mas os bombeiros **foram** rápido.*
- S21** *Quando era criança [...] As vezes eu **fui** para Portugal com minha família porque meu país vive ao Algarve.*
- C13** *Moro em Lisboa, vivo no apartamento no centro de Lisboa. [...] Depois de estudei em escola de Macau, eu **vou** para Portugal.[...] Por isso, em este Verão, **vais** a Lisboa e fazemos uma viagem em Lisboa! Ok?*

²⁹⁴ Batoréo 2000 discute largamente esta questão e faz um estudo interlinguístico deste par. «Um verbo deíctico é um verbo de movimento definido em função dos intervenientes do acto de interacção verbal, isto é, o seu significado prototípico implica que o Locutor e/ou o Alocutário se coloquem na Origem e/ou no Alvo. Em Português, os verbos deícticos podem ser abordados em dois pares *ir/vir* e *levar/trazer*.» (Batoréo 2000:511).

Assim, é provável que o aprendente, como se pode ver acima, neste estágio de aquisição, prefira *ir*, o verbo mais frequente do par. E é de facto o que acontece, já que a quase totalidade das 16 ocorrências desviantes de *ir* se deve à não selecção de *vir*.

Outro verbo que, em português, tem o seu campo bastante reduzido é o correspondente ao hiperónimo TAKE. De facto o verbo **tomar** caracteriza-se, em relação às outras línguas, por uma redução, até mesmo a nível prototípico, das suas possibilidades de uso ²⁹⁵. Exactamente por isso ele não é um dos verbos mais frequentes em português, ocupando a quinquagésima nona posição na lista ordenada por ordem decrescente do PF, mas a quadragésima primeira no PESTRA.

- F37** *Tomamos pequenas refeições aqui e ali mas não são sempre nutritivos. E como nos somos sempre a apressa-nos, não **tomamos** tempo para comer sentado num sitio que pode descansa-nos.*
- A10** *O ladrão saiu do carro e coreu para mim. [...] Ele olhou atrás e quando veio a policia ele **tomou-me** e puxou-me numa loja.*
- A11** *O homem respondeu que morava perto o podia arranjar facilmente um martelo e um prego para reparar o meu sapato. Por isso eu acordei e ele corriu **tomando** o salto estragado para sua casa ao lado daquela rua.*
- S31** *Éramos fartos de esperar! **Tomámos** um intervalo para ir a um café onde bebemos café e comemos pasteis de nata.*

Pelo contrário, o menor número de ocorrências de **dar** no PESTRA do que no *sub-corpus* português e no PF explica-se certamente pela não extensão do seu sentido prototípico a muitos usos periféricos. Os portugueses fizeram bastante mais uso deste verbo, embora sem que, nessas extensões, na quase totalidade dos casos, ele tenha perdido a sua estrutura argumental prototípica. São exemplos do seu uso pelos portugueses: *dar 100\$00, dar conselhos, dar prazer, dar origem, dar a vez,*

²⁹⁵ No Esquema Imagético do verbo **tomar**, em Português, o evento é volitivo e inceptivo. No PE, ele «caracteriza-se pela representação reduzida do significado prototípico, limitando-se praticamente a instâncias sinónimas de ‘ingerir alimento ou medicamento’ [como *tomar chá, tomar um comprimido*]. Uma maioria significativa de usos de *tomar* abrange instâncias abstractas...» [como *tomar banho, tomar uma decisão, ...*]. «Repare-se que o verbo *tomar*, em P, não está marcado quanto ao Modo, como acontece com os seus equivalentes noutras línguas. Na necessidade de exprimir o conceito de ‘*tomar+Modo*’, o sistema da língua utiliza a estratégia específica da lexicalização que resulta na criação de outros verbos – marcados quanto ao Modo – tais como, por exemplo, *agarrar* ou *pegar*.» (Batoréo 2000:522).

dar uma queda, mas também *dar nas vistas*, *dar cabo de* e *dar-se conta de*. Mas é nos documentos espanhóis que se pode encontrar em maior número sequências mais periféricas (*dar-se conta de*, *dar a conhecer*, *dar para comer*, e *não dar para sair*). De facto, no PESTRA, só se registaram 4 ocorrências desviantes, sendo 2 dos espanhóis (ver TABELA 16).

- E19** ... se alguém quiser ir comigo não faz mal só que, antes de sair-mos, gostava de lhe **dar** as minhas recomendações.
- S26** O governo e a policia achavam que esteve uma boa idea para **dar** a culpa dos 'hippies' mas eles sabem que esteve um outro grupo de terror que fez.
- C06** Eu conheço uma amiga, que é muito rica. E ela tem muitos amigos [...] Ela paga sempre. É verdade, ela **dá**-lhes as vidas muito preenchidas.

São também os espanhóis quem usa, até mais que os portugueses, abundantemente o verbo **fazer**. De facto, foram contadas 50 sequências diferentes com o verbo *fazer* (*fazer pontes*, *fazer exame*, *fazer compras*, *fazer férias*, *fazer projectos*, *fazer frio*, *fazer anos*, *fazer uma ideia*, *fazer-se com (farinha)*, *fazer o possível*, *fazer parte de*, *fazer votos*, etc...), enquanto que no *sub-corpus* português elas não passam de 20.

- E23** Os portugueses não precisam nem de vistos, nem de vacinas. Ao chegares, o guarda **fazerá** um carimbo no teu passaporte. Assim, poderás ficar no México 90 dias.
- A38** Quem sempre usa o metro ou os autocarros não pode ser numa boa forma. Então – não só ver televisão pode **fazer prazer** mas também **fazer esporte** e ginásticos.
- S19** Primeiro vou levá-lo a Sesimbra [...] Depois vamos atravessar as Arrábidas [...] e no pico mais alto vamos **ter um piquenique** com pão com manteiga, queijo fresco, fiambre, uns rissois e café.
- C18** Quando eu ganho mais dinheiro, da-ti o capital para **fazer** uma loja. Na vila é difícil [...] Mas **fazer** uma loja na vila está barata. Eu penso **fazer** uma loja no cidade Porto ou Braga.

Podemos ainda constatar que muitos desvios acontecem por **generalização** de certos verbos frequentes a usos menos prototípicos para

os quais a língua alvo deu preferência a outros. É o caso de certos predicados complexos, tais como *fazer um carimbo* por *pôr um carimbo*, *de fazer uma loja* por *abrir uma loja*, *fazer prazer* por *dar prazer*, e *ter um piquenique* por *fazer um piquenique*. Eles ilustram a dificuldade de traçar fronteiras entre verbos de elevada frequência, graças exactamente às suas ocorrências mais ou menos dessemantizadas, como **dar, fazer, pôr e ter**. Em qualquer dos casos, a sua extensão a valores menos prototípicos é comum a várias línguas²⁹⁶, no entanto, mesmo em línguas relativamente próximas, eles não são fiáveis para o aprendente (veja-se, por exemplo, *dar atenção*, *faire attention* e *pay attention*).

Como vimos no Capítulo 1, a linguística tipológica defende que a relação entre forma e função não é arbitrária, e que as línguas diferem quanto ao grau de saliência que atribuem a um determinado significado. Em conformidade com essas hierarquias, os padrões de lexicalização variam consoante as línguas de modo consistente e regular em cada uma delas ou dentro de um grupo tipológico. Assim, como nos mostrou Leonard Talmy, uma outra característica das línguas românicas, que as distingue de muitas outras línguas indo-europeias, nomeadamente das línguas germânicas, e também do chinês, é o diferente padrão de lexicalização que rege os verbos. É à luz desta diferença que podem ser interpretados desvios em verbos um pouco menos frequentes, mas que, em alguns casos (*chegar* e *andar*) ainda fazem parte dos 20 mais frequentes no PF. O verbo *andar*, como se pode verificar no ANEXO 18, não faz parte dos 50 mais frequentes no PESTRA, no entanto, ocupa a décima oitava posição no PF²⁹⁷.

A10 *Peço mil desculpas para o meu atraso. Mas, faz favor, ouve que tenho explicar. Queríamos encontrá-nos no jardim das estrelas. Era um bom dia. <Bom tempo> Sai a casa punctual e normalmente tinha bastante tempo **andar** ao encontro.*

²⁹⁶ Ver Duarte 2000, sobre “predicados complexos”.

²⁹⁷ O verbo **andar** parece ter muitos sentidos não prototípicos; daí a sua alta frequência (ver *andar a*, *andar de*). Estas extensões são fonte de desvios (**E36** *A alimentação e a saúde vão tão ligadas que não podemos esquecer-nos da importância dela na hora de comer*; **F43** *Os governos têm como tarefa de encontrar e aplicar medidas para controlar a poluição. [...] Outra sugestão seria a criação de pistas para as pessoas irem de bicicleta*. **A48** *O facto é, que todos os carros estão estacionado ou corrido no mesmo tempo, seguidamente é bastante difícil encontrar um lugar livre. Na Alemanha, prefero **ir da** bicicleta, no campo e também nas cidades, mas em Lisboa não é muito agradável.*)

- A10** *Eu **ultrapassei** a Rua do Sorte quando um acidente aconteceu. [...] e no momento em queria **conseguir** o outro lado da rua o ladrão encontrou-se numa semáfora. O ladrão saiu do carro e correu para mim. Era um choque, olhando para ele eu não consegui fazer nada.*
- A12** *Desculpe, desculpe, desculpe. Eu sei que ontem esperaste duas horas ao nosso arvore no parque. Depois eu tinha chegado, tu **correste** sem ouvir os meus desculpes.*
- A20** *Resumido, não há dinheiro para emprestar. [...] Ou venda o teu ‘Porsche’ ou o ‘Rolex’. Ninguém precisa nesta altura estes simbolos do capitalismo. [...] Esperando que tu possas **chegar** o dinheiro que precisas. O teu amigo ...²⁹⁸*
- A24** *Os jovens deste bairro arranchavam uma marcha em vestidos tão giros que eles **recebiam** o 1º prémio na <competição> competência com os outros bairros!*
- S39** *Na Suécia é cerveja muito cara. Muitas pessoas fazem cerveja em casa e depois **conduzim** o resultado para os amigos.*
- C19** *Por outro lado, sei que tens de **receber** dinheiro para resolver o problema. Eu conheço um amigo que trabalha num banco.*
- C21** *Se tiveres 20 dias para viajar, pode **viajar** Macau, Hong Kong e China. [...] Em Macau, há o transporte suficiente para **viajar** a cidade ou para **viajar** outros países.*

Os exemplos e os comentários acima dão apenas uma pálida ideia da riqueza do *corpus* no que respeita desvios na selecção dos verbos.²⁹⁹ Aludi somente aos verbos básicos visto que o objectivo era mostrar que, como prevê a abordagem linguística tipológica, **quanto mais específico de um grupo de línguas ou de uma língua é um verbo mais difícil será a aquisição de todas as suas propriedades semânticas.** De facto, **os desvios radicam em tendências universais vs. características particulares, podendo estas últimas ser idiosincrasias do português ou da L1, e situar-se tanto ao nível prototípico como periférico. Por outro lado, mais uma vez, fica claro que, contrariamente àquilo que nos pareceria óbvio, o maior número de desvios se concentra em itens lexicais muito frequentes.**

²⁹⁸ «...se procurássemos um verbo que traduzisse a ideia de ‘meta’ como por exemplo através da expressão ‘atingir determinado sítio’, nenhum melhor do que o verbo *chegar*» (Teixeira 1995:136)

²⁹⁹ Além destes, seriam dignos de um comentário extenso, entre outros, *ver – olhar, ouvir – escutar, levar – trazer, procurar – encontrar, querer – gostar, saber – conhecer*, ou o verbo **arranjar** que dada a sua abrangência (*arranjar o carro, arranjar bilhetes, arranjar um emprego, arranjar problemas, arranjar uma doença...*), faz supor ao aprendente que as suas possibilidades de extensão são quase infinitas (*arranjar trânsito, arranjar uma marcha*).

6.3.3.4. Desvios em preposições

O facto de as preposições constituírem uma classe de palavras fechada formada por um número bastante restrito de itens permite observar com alguma facilidade o *continuum* que vai das semanticamente mais robustas, como é o caso de *até*, *desde* ou *sem*, até às mais débeis porque (e por isso...) mais disponíveis para a polissemia, podendo chegar a desempenhar funções meramente sintácticas, como é o caso de *a* e *de*. De qualquer modo, o conteúdo semântico das palavras gramaticais parece remeter para o conhecimento linguístico e não para o conhecimento do mundo ³⁰⁰. Na verdade, se procurarmos na nossa representação mental o sentido prototípico de cada um destes marcadores, teremos muito mais dificuldade em captar o destas últimas do que o das primeiras. É natural, portanto, que o aprendente precise de muito mais estruturas exemplares das semanticamente menos robustas e mais polissémicas para reestruturar o seu sistema interlinguístico e estabelecer o sistema alvo. A TABELA 18 dá conta do uso de preposições, no *sub-corpus* português e no *PESTRA*, sem ter em consideração as ocorrências desviantes.

Tabela 18

PREP.	port.	espanhol	francês	alemão	sueco	chinês	Total	Média
de	692	582	566	480	361	461	2450	490.
em	293	294	327	381	365	334	1701	340.2
para	153	163	200	153	170	152	838	167.6
a	230	165	104	111	102	136	618	123.6
com	122	102	90	108	145	76	521	104.2
por	37	58	35	41	48	34	216	43.2
sem	18	20	23	17	25	12	97	19.4
sobre	5	11	12	23	30	10	86	17.2
até/até a	12	14	13	13	9	12	61	12.2
desde	9	7	10	4	0	1	22	4.4
	1571	1416	1380	1331	1255	1228	6610	1322

A tabela está ordenada da preposição mais usada (*de*) para a menos usada (*desde*) no *PESTRA*. Do confronto das colunas ‘port’ e ‘média’, podemos concluir que os portugueses usaram mais preposições (1571) do que os estrangeiros (média do total: 1322). Mas, um olhar comparativo mais

³⁰⁰ Agradeço a Manuela Vasconcelos esta reflexão.

atento levar-nos-á a concluir que, nalguns casos, usaram até mais ou um pouco mais (*em, para, por, sem, sobre e até*). A diferença para menos na média do *PESTRA* é da responsabilidade, numa pequena medida, de **com** (port: 122; *PESTRA*: 104.2) e **desde** (port: 9; *PESTRA*: 4.4), e, em grande medida, de **a** (port: 230; *PESTRA*: 123.6) e de **de** (port: 692; *PESTRA*: 490).

A *TABELA 19* apresenta resultados globais quanto a desvios em preposições, sem ter em conta a sua origem no que respeita a L1 dos aprendentes. Vemos que os casos de **omissão de preposição** (280) são os que motivaram maior número de desvios. Em contrapartida, os aprendentes procederam 143 vezes à **inserção** de uma qualquer preposição onde os falantes nativos não teriam usado nenhuma.

Tabela 19

	Ø	a	de	em	para	por	com	sobre	desde	(outra)
usou ...	280	92	138	95	72	45	35	6	0	1
onde devia ter usado outra										
usou outra	143	160	193	76	112	31	32	7	0	14
onde devia ter usado ...										
	432	252	331	171	184	76	67	13	0	15

Mas, de facto, são as preposições **de** e **a** aquelas que provocaram maior número de desvios, sendo, no entanto, a preferência abusiva por **de** (138) e por **a** (92) menor do que a sua omissão (*de* 193, *a* 153). Mas a observação ficará distorcida se não relacionarmos o **número total de ocorrências** de cada uma delas com o respectivo **número de desvios**. É disso que trata a *TABELA 20*, ordenada por ordem decrescente da percentagem de desvios.

Tabela 20

PREP.	TOTAL DE OCORRÊNCIAS	TOTAL DE DESVIOS	% DE DESVIOS
a	618	252	40,7
por	216	76	35,2
para	838	184	22
sobre	86	13	15,1

de	2450	331	13,5
com	521	67	12,9
em	1701	171	9,9
até /até a	61	2	3,3
sem	97	0	0
desde	22	0	0

De acordo com esta, a preposição *a* é, de facto, a que põe mais problemas à aprendizagem. Os resultados mostram-nos que em quase 41 % dos casos em que ela está envolvida há uma qualquer forma de desvio: omissão, inserção, ou uso desviante. Seguem-se *por* e *para*, com 35 % e com 22 %. Na mesma tabela, a preposição *de*, contrariamente àquilo que pareciam indiciar os resultados mostrados na *TABELA 19*, apresenta uma baixa percentagem de erros. De facto, o número total de desvios deve-se à sua altíssima frequência na língua. O mesmo se passa com *em*: regista um alto número de ocorrências mas relativamente poucas são desviantes.

PÁGINA EM BRANCO PARA A TABELA 21

Passando à *TABELA 21*, vê-se de que modo cada um dos grupos contribuiu para os resultados globais. Da sua observação ressalta que, no caso dos **espanhóis**, se verifica uma concentração de desvios em α , motivada tanto pelo seu uso abusivo como pela sua omissão; poderia

esperar-se que os **franceses** seleccionassem *por* e *para* de forma desviante, mas, de facto, os dados não o confirmam; o que parece é terem uma preferência excessiva por *de* e por *em*, em detrimento de *a*; para os **alemães**, a mais problemática parece ser *de*, seguida de *em*, mas a dificuldade alastra e atinge *por* e *para*; e, nos **suecos**, ela inclui também *com*; os **chineses**, apresentam um quadro ligeiramente distinto, ou, se quisermos, mais desviante, situando-se a questão fundamentalmente na omissão de quatro destes marcadores (*a*, *de*, *em* e *para*). Ou seja, como seria de esperar, também neste caso a distância linguística se faz sentir; à medida que ela vai aumentando maior número de diferentes preposições vai sendo afectado por desvios. Os exemplos que se seguem, com a preposição *a*, por ser aquela que apresenta maior percentagem de ocorrências desviantes, não pretendem mais do que ilustrar o problema.

E09 se viesses eu podia marcar um jantar para poder convidar /<-lhes>/ a todos os meus amigos.

E11 *Eu lembro-me muito da Missão, [...] das mães que sofrem porque não tem comida nem medicamentos para dar o seus filhos.*

E16 *gostava de ter um emprego no que tivesse a oportunidade de poder acabar o mestrado, e que pelo menos o ordenado fosse de 300 000 escudos [...] Bom se calhara aspiro muito.*

E25 *Estive na festa até as quatro da manhã. Depois fomos a tomar o pequeno almoço num café que está no Saldanha.*

E27 *foi para mim uma de esas noites inesquecíveis e que poucas vezes voltam a repetir-se. Ali conheci a pessoas que não voltei a encontrar.*

E41 *há uma tendência geral ao flexibilização do mercado do trabalho duma parte, e de outra, liberalizar aos governos, da carga dos pagos das reformas.*

F03 *Para ela, ha tantas coisas a vêr no mundo que ela não pode acabar de viajar.*

F12 *Sou aqui, com minha carta, para te pedir desculpa. Como sabes muito bem, devia vir para tua casa e gentar contigo. Não consegui apanhar o comboio para chegar na tua aldeia.*

F28 *Eu acho que as pessoas que ‘acreditam’ no destino servem-se do mesmo conceito como pretexto: é uma desculpa para fugir da Libertade demasiada difícil a suportar.*

A02 *Há um mês o meu amigo Thomas passava as suas férias em Portugal. Ele ficava ao apartamento da minha namurada e mim.*

A10 *Quando entrei ao banco, porque precisava suficiente dinheiro para sair contigo, normalmente é muito caro sair contigo...*

A42 *A última possibilidade para a melhorar é talvez, que as pessoas, que trabalhem em Bars, Discos etc. tenham mais cuidado com o álcool mas eles*

pensam sempre só a ganhar mais dinheiro.

- S14** *No este tempo hão muitos jovens que pensam nas suas carreiras profissionais. [...] As empresas têm de prometer os empregados novos aumentos de salário [...] e promovê-los depois dois anos.*
- S20** *Mas agora temos uma semana para fazer um passeio cá em Portugal. Primeiro telefono o meu amigo P. Para contar a notícia.*
- S22** *Depois de Estocolmo pudemos ir à norte da Suécia para vermos o sol de meia-noite. [...] Outra vez em Estocolmo podemos tomar o barco à Gotland, uma ilha fora da Suécia.*
- C18** *Quando você trouxe-me no lago perto Campos. Você disse-me: 'A vida verdade!' A vida está bonita. e os habitantes são boa.*
- C23** *Eu sou de Pequim e gosto muito de ir de viagem. A China é muito grande, então eu posso ir muitos lugares com meus amigos. Já fui muitos lugares, por exemplo, A Parede Longo, a Palácio de Verão...*
- C45** *Quando cheguei, em primeiro mês, morei numa casa que estava em Odivelas. *Provavelmente, durei quase uma e meia hora para chegar universidade. [...] Então eu mudei a casa a Lisboa.*

Os desvios de que estes exemplos são uma amostra têm, com certeza, e fundamentalmente, na sua origem *transfer* da L1 e ou hipóteses incorrectas a que a fraca transparência dos dados do PE conduz os aprendentes. É sabido o papel que tem o ensino na chamada de atenção para certos aspectos menos salientes da língua em aprendizagem que, de outro modo, poderiam passar despercebidos. Essa chamada de atenção, se, por um lado, pode, em parte, conduzir a usos abusivos, por outro, estimula a reestruturação constante da interlíngua aproximando-a da variedade alvo.

Os efeitos do ensino na aprendizagem, e, é claro, de *input* abundante, podem ser notados exactamente no caso das preposições quando se compara o sistema de falantes de L2 que aprenderam nestas circunstâncias com aqueles que, pelo contrário, dispuseram de um ambiente linguístico muito mais pobre. Isso reflecte-se, por exemplo, nos crioulos de base lexical portuguesa. De facto, estes não seleccionaram da língua fonte todas as preposições mais frequentes mas só as mais salientes. Como resultado desta selecção, a preposição *a* não passou para os crioulos (para exemplo com o crioulo da ilha do Príncipe, ver Muysken 1988). Do mesmo modo, no português do Brasil (PB) e no português de África (PA) ela é em muitos

contextos substituída por *em* ³⁰¹. Hanna Batoréo (2000:444) diz mesmo que *a* é uma preposição típica do PE. Perpétua Gonçalves vê a questão de um modo um pouco diferente, e que chama a atenção para o papel do ensino.

Gonçalves (1989 e 1996) estudou as ocorrências da preposição *a* num *corpus* constituído por documentos escritos e orais, produzidos por **estudantes universitários**, nos quais “parece possível encontrar algumas tendências estáveis da norma culta do PM” [português de Moçambique]. Como é sabido, ainda que a preposição *a* em PE reja obrigatoriamente o OI, há também alguns OD que podem ser regidos por ela. Da análise dos dados fornecidos pelo *corpus*, Gonçalves conclui que os falantes de PM generalizam esta possibilidade e inserem a preposição *a* junto de constituintes com a função gramatical de OD sempre que este possui o traço [+ humano] (Gonçalves 1989:16) ³⁰². Este comportamento é explicado pela autora como sendo uma reinterpretação do papel desta preposição: ela passa a ser um marcador relacionado com a semântica do complemento seleccionado pelo verbo, e não um marcador sintáctico com a função de introduzir este complemento, como acontece em PE. Isto significa, portanto, que, mesmo neste caso, em que, em PE, ela tem uma função meramente sintáctica, lhe é atribuído um conteúdo semântico por parte destes falantes de L2. Tratando-se, como nota a autora, de motivação que reside na língua-alvo, é de considerar a hipótese de que o mesmo se verifique no PESTRA. Por outro lado, como referi acima, é de considerar que o facto de os informantes deste estudo, tal como os do PESTRA, terem tido acesso a ensino formal e demorado da língua seja responsável pela manutenção, ainda que reinterpretada, da preposição. Mas observemos agora o comportamento de cada um dos grupos.

A TABELA 22 diz-nos que em 93 (18+39+29+3+4) casos a preposição *a* foi substituída por outra e em 67 pelo marcador \emptyset . Diz-nos ainda que à medida que a distância linguística vai aumentando também vai subindo,

³⁰¹ Cabral 2005, que estudou um *corpus* constituído por documentos produzidos em contexto escolar por falantes de português de Angola, afirma: “[E]ntre as proposições que desencadeiam o maior número de desvios, a primeira prende-se com o limite (*a*), a segunda com a imprecisão (*em*) e a terceira com a origem (*de*). A tendência de a preposição *em* estar a especializar-se para contextos característicos de outras preposições, nomeadamente *a*, *para*, e *de*, pode ser o resultado da sua perda de características específicas, ou seja, do seu esvaziamento de sentido...” (Cabral 2005:174).

³⁰² Entre outros, são exemplos fornecidos pela autora: *o professor alerta aos alunos; a filha do imperador amou ao Manuel; a natureza não pode dominar ao homem*.

Tabela 22

devia ter usado <i>a</i> mas usou	Ø	DE	EM	PARA	POR	COM	SOBRE	DESDE	
espanhol	9	1	3	4	1				18
francês	1	2	21	2	1				27
alemão	7	6	4	8		3			28
sueco	15	6	6	14	1	1			43
chinês	35	3	5	1					44
	67	18	39	29	3	4			160

ligeiramente, o número de ocorrências desviantes. A TABELA 23, em contrapartida, dá-nos informação quanto ao uso abusivo da mesma preposição. Como se pode verificar, é relativamente menor o número de ocorrências desta natureza, e menos nítida a relação L1 - número de ocorrências desviantes. O que é relativamente mais nítido é a estratégia seguida por cada um dos grupos, sobretudo quando lemos as duas tabelas em conjunto.

Tabela 23

usou <i>a</i> mas devia ter usado	Ø	DE	EM	PARA	POR	COM	SOBRE	DESDE	
espanhol	10	2	1	4					17
francês	4	5	3	4		1			17
alemão	7	1	11	8			1		28
sueco	5	1	1	6					13
chinês	7			10					17
	33	9	16	32		1	1		92

Os **espanhóis** parecem ser apanhados num dilema que equilibra omissões (9) e inserções (10). Por outro lado, no *sub-corpus* espanhol, os desvios relativos a inserção da preposição com o OD (*convidar a, conhecer a, liberalizar a ...*) são comparáveis aos dos informantes do estudo de Perpétua Gonçalves. No entanto, e ao contrário do que acontece no caso destes últimos, a L1 dos espanhóis pode ser responsável por este tipo de ocorrências. Ela tem evidência que reforça os casos excepcionais existentes

em PE, já que é frequente em espanhol a preposição *a* introduzir o OD, quando este possui o traço [+ humano] ³⁰³.

No caso dos **franceses**, as 17 ocorrências abusivas de *a* parecem estar muito relacionadas com a L1 (*ser difícil a, conseguir a, confrontar-se a, pensar a, convidar a, etc...*). A substituição de *a* por *em* aproxima-os de outras variedades nacionais do português, nomeadamente, do PB e do PA. De facto, há nos dados 21 ocorrências de *em* por *a* com verbos de movimento de alta frequência (**chegar** *na casa, na aldeia, no mercado, na estação; ir em Egipto, numa loja, neste país, no Algarve, no aeroporto, no centro, no cinema, etc...; voltar* *na agência*). Dado o grande número de casos, poderia pôr-se a hipótese de estes informantes terem tido uma exposição demorada ao PB, mas, de facto, não há mais nada nos documentos que apoie essa hipótese, e, além disso, os desvios repartem-se por nada menos do que 15 informantes. A hipótese mais viável será a dificuldade de estabelecer correspondências claras entre o sistema preposicional francês e o português.

Ao contrário dos franceses, os **alemães** mostram uma certa preferência pela preposição *a*; usam-na 7 vezes em contextos em que não devia ocorrer (*talvez encontras a muita gente, consegui a apanhar o autocarro*, por exemplo); usam-na 11 vezes em vez de *em* (*entrar ao banco, ficar ao apartamento, pensar ao peso, ter dificuldades a entender*) e 8 em vez de *para* (*ir de carro ao trabalho, ir a Portugal, olhar aos hábitos, pedir a passar, sorrir a mim, haver coisas a ver*), embora, em alguns casos, também a substituam por outras (*telefonar com, estar ligado com; começar de, continuar de; ir em, chegar em, estender-se em*), ou a omitam (*responder alguém, perguntar alguém*).

Como se pode ver na TABELA 21, enquanto os alemães apresentam um quadro de resultados bastante confuso, o dos **suecos** é muito mais claro. Apesar de alguns casos de *em* por *a* (*assistir na conferência, ir no Algarve, etc...*), o mais frequente é omitirem a preposição *a* (15 ocorrências) (*telefonar o meu amigo, prometer os empregados, ...*), ou substituem-na por *para* (14 ocorrências) (*autorizar 73 mil pessoas para ficar, dar moedas para*

³⁰³ Ver, em Bosque & Demonte (dir.) 1999, Cano Aguilar, 'Los complementos de régimen verbal'.

os pobres, ficar para apanhar sol, mostrar para os vizinhos...), embora também haja alguns casos de a por para (6 ocorrências) (tomar o barco à Gotland, vir a Lisboa, ...).

Aquilo que é anunciado nos dados suecos surge com mais clareza ainda no quadro do grupo dos **chineses**: a omissão frequente de *a* e uma relação relativamente forte entre esta e *para* (*ir ao escritório, mudar de casa a Lisboa, o dinheiro corre em abundância ao governo...*). Os casos em que é omitida a preposição *a* correspondem, *grosso modo*, àqueles em que outras variedades nacionais do português e os franceses usam *em*, ou seja, com os verbos de movimento muito frequentes (*ir casa de, ir viajar Coimbra, chegar Lisboa...*).

Poderá, portanto, pôr-se como hipótese, sujeita a confirmação pela análise detalhada dos dados, que, no caso do marcador que rege os complementos dos verbos de movimento, há uma ordem de aquisição de que o primeiro estágio é o marcador \emptyset , o segundo estágio será a preposição *em*, e a preposição *a* será o terceiro. A distinção entre *a* (*ir a casa*) e *para* (*ir para casa*) é, certamente, adquirida depois e constituirá um quarto estágio.

Não me parece que valha muito a pena continuar com uma apresentação dos dados seguindo este modelo para todas as preposições. Ir além desta muito insatisfatória enumeração de desvios, só será possível se desenvolvermos uma investigação sintáctico-semântica minuciosa, tendo em conta todas as restrições contextuais, nomeadamente, os verbos de que elas regem complementos. Fica registada, e deve ser investigada, a hipótese de Perpétua Gonçalves, segundo a qual os falantes não nativos reinterpretem o valor de certas preposições. Essa investigação deve ter como objectivo final contribuir para o estabelecimento de estádios de aquisição. Mas, só por si, ela constituiria um trabalho de grande envergadura que, infelizmente, não cabe no âmbito desta obra.

Antes de abandonar as preposições, gostaria ainda de salientar que, **mais uma vez, e isto aplica-se a todas elas, são os verbos mais frequentes que mais contribuem para os desvios em preposições.** Acabámos de ver que, no que respeita a preposição *a*, um número bastante grande de desvios acontece com verbos de movimento de altíssima frequência (*ir, chegar, voltar*); se espreitarmos para a preposição *de*, a mais frequente, veremos que, dos 143 casos de omissão, 46, o que equivale a 32

%, são relativos a *gostar de* (ESP: 4 de 5; FRANC: 0 de 2; ALEM: 15 de 38; SUEC: 15 de 42; CHIN: 12 de 43) e 19, o que equivale a 13,3 %, são relativas a *precisar de* (ESP: 0 de 5; FRANC: 0 de 2; ALEM: 7 de 38; SUEC: 6 de 42; CHIN: 7 de 43). Ou seja, **mais de 45 % das omissões da preposição *de* estão relacionadas apenas com dois verbos.**

6.3.3.5. *Desvios em outras categorias*

Em Mateus *et alii* (1989:176) pode ler-se que, em português, «importa considerar como categorias sintáticas principais as seguintes categorias lexicais: ... (i) N(ome); (ii) V(erbo); (iii) ADJ(ectivo); (iv) ADV(érbio); (v) PREP(osição)». As autoras acrescentam que «as categorias N, V, ADJ e PREP são usualmente caracterizadas através de uma dada combinação de valores dos traços 'N' (=nominal) e 'V' (=verbal)». Da **categoria ADV** diz-se em nota que ela «é, historicamente, uma categoria derivada, não caracterizável homogeneamente através de uma combinação de traços N e V. Assim, a generalidade dos deícticos de lugar e de tempo do Português actual têm a sua origem em expressões nominais ou preposicionais latinas contendo demonstrativos (cf. *aí, cá, hoje, agora*). Outros advérbios formaram-se a partir do antigo ablativo instrumental de adjectivos latinos (cf. *bem, mal, longe, tarde*). Alguns advérbios formaram-se a partir da forma não marcada de adjectivos (cf. *muito, pouco, alto, baixo*).»³⁰⁴

Vem isto a propósito dos diversíssimos comportamentos e valores que os itens lexicais tradicionalmente incluídos na categoria ADV adquirem ao nível da frase/do enunciado ou do texto/do discurso e da conseqüente dificuldade de tratamento que envolvem tanto a nível da descrição como num trabalho como este³⁰⁵. Ou seja, cada um deles exige um tratamento minucioso (microbiologista, como diria Meara) impossível de realizar no âmbito desta investigação.

Exemplos de itens lexicais de alta frequência geralmente incluídos na categoria ADV usados de forma desviante no *PESTRA* são *ali, aqui, cá*, e

³⁰⁴ Ver, no Capítulo 3, a nota 177.

³⁰⁵ «[A]s gramáticas enquadram tradicionalmente entre os advérbios uma quantidade enorme de palavras que, apenas em algumas ocorrências particulares e em alguns ambientes sintáticos, atendem aos critérios tradicionais. Tratar do 'advérbio' é, antes de mais nada, tomar consciência desses equívocos, constatando a diversidade de emprego dessas expressões.» (Ilari 1996:108).

então. Ana Cristina M. Lopes 1996 chama a atenção para a necessidade de levar a cabo investigação centrada «na caracterização semântico-pragmática de alguns itens da língua portuguesa cuja classificação gramatical tem suscitado perplexidade ao longo do tempo, dado o seu funcionamento / comportamento linguístico não homogêneo». A inexistência de estudos sobre o item **então**, «tão frequente na interacção verbal quotidiana», levou a autora «a escolhê-lo como primeiro dado a analisar no conjunto das partículas discursivas do português» que tenciona estudar. (Lopes 1996:177).

Lopes 1996 questiona a classificação de Cunha e Cintra 1984 como advérbio de tempo e como ‘palavra denotativa de situação’, mas faz notar que os autores «tiveram consciência do valor não uniforme de *então*, dado que o incluíram em duas categorias gramaticais distintas.» Parte da análise de um conjunto de exemplos para identificar e caracterizar «os valores semânticos e pragmáticos do item em apreço» e propõe três valores:

(1) um valor temporal, como em “*O João contou tudo à irmã. Sentiu-se, então, muito aliviado*”, em que, «na interpretação mais plausível, é possível uma paráfrase em que a sequencialidade temporal se alia à causalidade» (*e depois / por causa disso*), não excluindo embora «a possibilidade de uma paráfrase do tipo *nesse momento, nessa altura*».

(2) um valor argumentativo, como em “*Queres entrar em Medicina? Então, estuda!*”, comutável por *nesse caso e sendo assim*; é ainda parafrasável por *O que conclus dá?*, como em “A: – *Cheiras a tabaco*. B: – *E então?*”; e comutável por *portanto* «quando introduz uma proposição conclusiva cuja evidência é garantida por uma assunção prévia», como em “*Esta inferência pode ser cancelada sem contradição. Então, é uma implicatura*”. E

(3) um valor de marcador de estruturação conversacional; trata-se de «uma partícula discursiva que ocorre exclusivamente na oralidade e funciona como uma espécie de marca de pontuação, operando, assim, ao nível da actividade enunciativa e não dos conteúdos proposicionais expressos», podendo, em certos casos, desencadear a interacção, «com uma função fática ritualizada, circunscrita a uma troca de abertura» (“*Então, como vai isso? / Então?*”). No entanto, ele «nunca perde em absoluto um valor anafórico / de retoma, [...] funcionando como mecanismo de

sinalização de continuidade discursiva» e «mesmo quando desencadeia a interacção, com uma função meramente iniciativa, pressupõe trocas verbais anteriores entre os interlocutores» (Lopes 1996:186).

A autora conclui o seu texto afirmando que a dificuldade de classificação gramatical «pode ser reequacionada: o primitivo advérbio temporal adquire, no âmbito textual, um valor conjuncional, no sentido amplo do termo: introduz funcionalmente uma conclusão, prefacia e articula constituintes discursivos». Mas, antes, chama a atenção para «as afinidades semânticas entre construções temporais, causais e condicionais» e para «a existência de uma rede cognitiva que articula de forma muito estreita os três domínios semânticos» (Lopes 1996:188).

O quadro abaixo dá conta de uma parte dos usos considerados desviantes (ou ligeiramente desviantes) no PESTRA, e que representam aproximadamente 15 % dos desvios incluídos “em outras categorias”³⁰⁶.

- F03** *Para ela, há tantas coisas a vêr no mundo que ela não pode acabar de viajar. **Então**, quando ela fala, é sempre muito interessant.*
- F31** *.... uma pessoa que tem os seus pensamentos próprios é uma pessoa livre. E a liberdade pede responsabilidade. **Então** que é tão facil de não ser livre...*
- F37** *Agora, a vida muda muito, as empregadas são muito caro **então**, temos que fazer as coisas da casa sozinhos. O restaurante é uma facilidade para nos mas isso também pode ser muito caro ao fim. **Então** nos não comemos correctamente.*
- F44** *Para mi o turismo de massas não tem só inconvenientes mas tenho medo que seja o caso um dia. **Então** preservem a natureza!*
- A08** *Somos mais velhos do que antes. **Então** na minha cabeça eu sou o mesmo jovem dantes. [no entanto]*
- A08** *Oxalá seja possivel um trabalho na universidade. **Então** vamos mergulhar num futuro feliz...*
- A38** *Quem sempre usa o metro ou os autocarros não pode ser numa forma boa. **Então** – não só ver televisão pode fazer prazer mas também fazer esporte e ginásticos.*
- A45** *Referendo-se ao transito nas grandes cidades, como por esémplo Lisboa, há dues possibilidades! Primeiro o trânsito ‘pessoal’ e segundo o trânsito ‘publico’ **então** o transporte publico. [ou seja]*
- S16** *...muito obrigada pela fotografia. [...] Lembro-me que naquela noite jantávamos num restaurant na baixa. Depois íamos ao concerto onde aquele tipo tirou a fotografia. O concerto era fabuloso. **Então** Katja temos de encontrar-nos.*

³⁰⁶ ESP – 0 (zero) usos desviantes em 17 ocorrências; FRAN – 7 em 19; ALEM – 6 em 11; SUEC – 6 em 10 ; e CHIN – 10 em 21.

- S25** *A gente agora fala muito sobre um filme que se chama “A lista de Schindler”. Eu também vi esse filme e gostei imenso. É uma história verdadeira e conta de uma tragédia que aconteceu só há 50 anos. **Então** há pessoas que ainda vivem e sabem o que aconteceu.*
- C13** *... eu vou à universidade em Lisboa. [...] Embora sei que é muito difícil [...] E eu queria ser o Médico. Como o teu pensas? Escreves-me na próxima vez! **Então**, eu penso que vou passar em Lisboa quase 10 anos (é muito tempo!).*
- C23** *E estas férias eram as melhores férias da minha vida. Agora, fico em Lisboa, não há muitos amigos, **então** não posso ir viagem muitas vezes.*

Embora os dados acima mereçam uma análise mais detalhada, em que sejam relacionados com usos não desviantes e com usos dos mesmos informantes, o que parece ressaltar é que **os aprendentes atribuem a *então* preferencialmente funções de articulador de constituintes discursivos, bem como funções argumentativas e conclusivas, daí resultando várias ocorrências em que o falante nativo preferiria certamente *por isso*** ³⁰⁷, e dando a entender que lhes passa despercebida a componente de sequencialidade temporal inerente a este marcador. Esta interpretação, por parte dos falantes estrangeiros, pode estar relacionada com o escasso número de ocorrências, no *input*, em que *então* assume exclusivamente o seu primitivo valor temporal. De facto, para a análise do funcionamento de *então* como marcador de estruturação conversacional, Ana Cristina M. Lopes recorreu ao *corpus* constituído pelas entrevistas transcritas do *Português Fundamental* publicadas no tomo 1 do vol. 2 (*Métodos e Documentos*) e concluiu que, neste *corpus*, «só 2 % das ocorrências de *então* assumem um valor temporal; 15 % das ocorrências actualizam um valor argumentativo»; e que «a maior percentagem, aproximadamente 80 %, corresponde ao marcador de estruturação conversacional...» (Lopes 1996:185) ³⁰⁸.

Também, segundo Rodolfo Ilari, «os dêiticos invariáveis a que se tem chamado tradicionalmente de ‘advérbios de lugar / advérbios de tempo admitem usos bastante variados, muitos deles difíceis de classificar.» No

³⁰⁷ A observação dos dados mostra que 60 informantes usam *por isso*; 22 usam *então* desviante; 8 destes 22 também usam *por isso*.

³⁰⁸ Segundo a autora, «*então* não é a única expressão que admite usos temporais e usos argumentativos: *agora*, *sempre*, e *logo* também exibem idêntica ‘flutuação’ semântica.» (Ana C M Lopes 1996:187). Ver também Frota 1992 e Moreira da Silva 1995. No entanto, os dados em apreço mostram que os aprendentes se restringem aos seus usos temporais. Mas, ver ***sempre***, em 6.3.4 *O léxico para além do léxico*.

entanto, é possível reconhecer-lhes «propriedades sintáticas e distribucionais próprias que os distinguem dos outros ‘advérbios’ ». Isso justifica o tratamento destes **deícticos** como uma classe à parte, bem configurada. (Ilari 1996:110-111). E, embora no *PESTRA* os usos destes itens se restrinjam quase exclusivamente aos seus usos mais prototípicos, parece, portanto, legítimo tratá-los juntamente com os deícticos demonstrativos e possessivos.

- E06** *Querido amigo A.. Ante-Ontem recibí a **vossa** carta de Espanha com a fotografia que fiz a professora [...] Por todo quero agradecer-lhe a sua carta e a fotografia.*
- E22** *Quando chegamos a Elcot Center todas as pessoas dividiram-se em pequenos grupos. [...] Primeiro fomos para os jogos mecânicos; e **alí** nós falamos com pessoas de diferentes partes do mundo*
- E26** *A festa foi na casa de um amigo. A casa fica em Telheiras, perto do “Carrefour”. É uma casa muito grande e **alí** as festas são muito divertidas.*
- F04** *Os meus amigos costumam fazer comentários irónicos sobre a minha barriga volumosa.[...] Então, antes de rirem do meu aspecto físico, lembrem-se **disso**: muito riso, pouco siso.*
- F09** *Tenho de voltar também para o Luxemburgo no dia 12 de Fevereiro e ficar **alí** durante três semanas porque a minha mãe é muito doente.*
- F14** *Exmos. Senhores... Eu li o **seu** anúncio no jornal “O Público” de ontem, no qual dizem procurar um empregado de mesa. Fiquei muito interessado neste trabalho.*
- F45** *Primeiro, todas as grandes cidades atiram muitos trabalhadores [...] Segundo, considero que <essas> **aquelas** cidades não são preparadas para recibir tanta gente. Lisboa fica uma cidade antiga, velha e <as> a maior parte das suas ruas são estreitas e não podem suportar todos <esses> **aqueles** novos carros!*
- A14** *Querido amigo! Espero que tudo esteja a correr bem por **alí**...porque por aqui podia ser melhor! Como tu sabes tenho cassado aqui no Brasil e moro cá há dois anos.*
- A36** *Quando eu ouvi que durante da revolução em Portugal quase ninguém morreu, eu pensei que **este** mostra muito bem a mentalidade dos Portugueses que é um povo tranquilo e positivo no pensamento.*
- A50** *Gosto muito de Berlim, a antiga-nova capital da Alemanha. O meu irmão morava **alí** durante 10 anos.*
- S11** *Tambem podias acabar a escola agora, ou seja, fazer um intervalo, e pedir dinheiro emprestado da **sua** família para **sua** viagem.*
- S37** *Pense nas outras pessoas que podem morrer por causa dum ‘accidente álcool’! Pense no **teu** vizinho!*
- S40** *Álcool e condução é dois coisas que não é bom para mexer. Mas os portugueses não compreender **essa coisa**.*
- S43** *Nós podíamos ajudar com construção de autostradas, com novas pontes, com*

*manutenção de barcos velhos, com equipamentos da polícia e mais. Nós não podíamos fazer **este** na Suécia, porque todo o país está pronto.*

- C03** *Um dia, fui às compras e vi um relógio muito bonito [...] Estava interessada em comprar um e esse parecia me convir. Pedi ao dono da loja para me mostra, mas ele disse que não se vende. Foi uma pena porque gostei muito **disso**.*
- C21** *Recebi a tua carta e fico muito contente para saber a **sua** visita. [...] Aqui vai uma lista de respostas da as **suas** perguntas.*
- C25** *Quando ele viu a pintura, e disse-me que ele nunca tinha recebido uma prenda tão bonita como **isto**, e iria guardar-a para sempre. As palavras do P fazia-me tão alegre.*
- C29** *Ontem à noite, voltei em casa. e depois liguei a televisão. Nesse momento, mostrou uma óptima paisagem sobre a San. **Aquela** cidade ficava no Sul da China. [...] Eu gostei muito aquele espectáculo. Por isso continuei a ver esse programa. A guia disse que no verão têm estado óptimo tempo, muitos estrangeiros foram **ali**.*

A análise permite concluir que o conjunto de desvios em deícticos constitui aproximadamente 25 % do total incluído naquilo a que chamei “em outras categorias”, sendo os demonstrativos, em particular as suas formas neutras, seguidos por *aqui, ali, cá e lá*, os responsáveis pela maior fatia. No entanto, como salientei em 6.2.1 *Vocábulos*, quando comparei o número de ocorrências no *sub-corpus* português e no *PESTRA*, o número de ocorrências desviantes em demonstrativos é relativamente insignificante em relação ao seu total (7,5 %) e menor que em marcadores espaciais (12,8 %). Quanto aos desvios em possessivos eles estão muitas vezes relacionados com o sistema das formas de tratamento, verificam-se nos pares *teu, tua / seu, sua*, e representam também uma percentagem muito baixa em relação ao total de 755 ocorrências no *corpus*.

Foram incluídos também neste ponto os desvios em **quantificadores**. Como se pode ver abaixo, e apesar de a expressão da quantificação recorrer a muitos mais processos do que o uso de advérbios, quantificadores indefinidos e de numerais ³⁰⁹, são muitos os itens envolvidos.

³⁰⁹ Para uma introdução a processos de **quantificação** das línguas naturais, ver *Discursos 4*. Este número contém, para além do texto introdutório do seu editor (Peres 1993), mais quatro textos sobre esta questão.

- E17** Gosto muito de fotografia e é por isso que pensei em responder ao seu anúncio para emprego no sector de fotografia. Sempre estou a tirar fotografias e gosto muito de **todo** o emprego relacionado com a fotografia porque penso que a fotografia é uma arte.
- E18** De facto, eu gostaria de ser directora duma Empresa de hotelaria para poder estar com **tantas** pessoas de distintos países, e conhecer como vivem em cada um delos.
- E32** Porque rara vez numa fila alguém protesta.[...] Se o autocarro demora meia hora para chegar, ainda bem que chegou! Que o empregado fecha a porta para não deixar entrarem as pessoas que estão a esperar na fila já há tempo, também não parece zangar **nenhum**.
- F08** Eu tive <tão> **tanto** prazer em ler a tua carta. [...] Penso **bem** em ti, cumprimentos ao teu marido, beijinhos aos teus filhos.
- F13** Este manha, estava pronta uma hora antes de sair de casa, todas as coisas começaram bem. As malas fechavam desde ontem, em fim **todo** estava bem. [...] E melhor de pensar a amanhã e ver se o final consigo a apanhar este avião com todos os papels. Espero que não esta **tão** zingado e que vais ficar melhor.
- F34** Aconteceu-me, em Lisboa, ficar uma vez numa situação quase igual àquela que a Suzana ficou. Tinha chegado a Portugal há pouco tempo, e, como <todo> **cada** estrangeiro, não conhecia todas as maneiras da vida lisboeta.
- F53** O objectivo da publicidade é criar um sentimento de insatisfação, um sentimento que falta sempre **uma** coisa necessaria para bem aproveitar a vida.
- A19** na universidade tinha uma amiga que também estava a estudar arquitectura comigo. [...] Mas no fim foi muito difícil para ela compreender que as construções tinham de ser barato e facil. Não era importante fazer desenhos muito giro, mas era important para arranjar **todo** muito bem.
- A22** Uma coisa especial e muito importante em Alemanha é sem dúvida um dicionário. Assim podes traduzir <todo> tudo sozinho porque sabes que os alemães **apenas** falam portugueses...
- A26** Encontrei-me com outros amigos e íamos todos à Praça do Comercio onde havia um grande espectáculo com muita música brasileira e **tanta** gente.
- A29** Mas se isso fosse tudo eu não vinha a Portugal. Há <muita mais> cada vez mais coisas ca em Lisboa que eu gosto de aproveitar o tempo libro que tenho aqui **tão bastante**.
- A35** Eu comecei de rir porque não percebi nada e a minha amiga não percebeu que eu estivesse a rir. 'Mas porque, isto é muito bom! Eu gosto também! Porque é que estás a rir **tão muito**?'
- A39** Nunca vi <tão muitas> tantos restaurantes, cafés, tascas etc. do que em Portugal. Pode comer em **cada** lugar. Sempre há uma pequena pastelaria.
- A45** Em Lisboa o sistema para pulir a cidade e muito bom bom porque quase **cada** rua no centro é pulita durante a noite
- S01** E muito normal pessoas que tem uma idade 15-18 comecem vestir-se diferente de os seus colegas na escola etc. **Alguem** vai-ser um Punk, **alguens** vão vestir-se numa maneira mais boémio, ou como as pessoas num grupo da música que eles gostam.
- S02** Vou falar sobre homens. Uma coisa que é importante para mim é que são

*simpáticos, limpos e cheiram bem (não, ótimo...). Isto é só pela primeira vista. Depois disto há muito mais coisas claro [...] Há alguma coisa melhor do que quando um homem cheira **tão** bem?*

- S03** *Há umas semanas encontrei-me com um rapaz Português. [...] Eu acho que ele é muito inteligente, um pouco tímido e que ele não tem <tão> **tã boa** confiança em si como ele pretende.*
- S07** *Era uma vez quando chegava no aeroporto de Lisboa.[...] Durante este tempo e agora também aeroporto de Lisboa estava com obras. Mas era **ninguém** informação sobre isto. [...] Ali havia **ninguém** sinal que mostrar aonde ia?*
- S09** *Filip e Kajsa têm muitos amigos, adoram os seus professoras e tomam banhos na piscina (na escola) muitas vezes **cada** semana.*
- S10** *Agora estou sentado numa café em Estocolmo [...] todas as coisas (quase) aqui são <muito> **mais** diferente **do que** Portugal.*
- S20** *Talvez queira ficar um dia em Albufeira. [...] Iriamos às praias diferentes e com certeza eu fico **tudo** vermelho.*
- S35** *Por exemplo quando eu entrar uma pastelaria boa em Portugal eu nunca vi **tão muitos** bolos e pasteis.*
- S45** *Agora a poluição de cidade é muito forte. È mais forte do que **apenas** anos atrás. Por exemplo quando meu avô era criança ninguém tinha automóvel e agora nós temos muitos carros. Talvez seja um carro por um homem e é **tanto!***
- C07** *... conheço um rapariga è portuguesa, ela é neta de dona da casa, ela não é muito bonito, mas é simpática. Quando eu vim Lisboa, não conheço **nada** e não pude falar português, ela ajua-me fui banco, escola ect.*
- C27** *Entramos-nos a sala grande de nossa faculdade, ouvi a música 'Rock & Roll' <com> muito alto. Os estudantes estavam dançar, ficavam muitos malucos. Todas usaram ropas <muitos> muito especial, **alguém** ficou como a bebê, **alguém** ficou como o 'Cowboy'.*
- C31** *Depois de chegar a Portugal, eu tenho a primeira impressão interessante pelo aspecto das filas interminável de pessoas. Sempre as vejo nas ruas, nas paragens, nos restaurantes de frango, nas cantinas, nas bilheteiras, <qualquer> **quaisquer** lojas e etc.*
- C44** *Se não houvesse **muitos** engarrafamentos, a vida nas grandes cidades não tornava difícil.*
- C45** *Todos os dias, tenho que andar de eléctrico e metro a escola. Mas as vezes, o eléctrico é **menos**. Uma vez, esperei o eléctrico há quase uma hora.*

Os desvios em itens lexicais que exprimem a quantificação representam aproximadamente 30 % do total dos desvios incluídos em “outras categorias”. Os suecos, seguidos dos chineses e dos alemães, são os grupos que evidenciam maiores dificuldades. Os pares *todo* /

tudo ³¹⁰, **todo / cada, e nenhum / ninguém são os que mais contribuem para esse número**. Porque neste estágio os aprendentes recorrem bastante a processos de intensificação, igualmente responsáveis por um grande número de ocorrências sentidas como desviantes são os pares *muito/tão* e *muito, -a, -os, -as/tanto, -a, -os, as* ³¹¹; digo sentidas porque, em muitos casos, elas são a transposição para a escrita de ocorrências que, no oral, por serem acompanhadas de entoação marcada, são gramaticais. Não se trata, portanto, de uma questão exclusivamente lexical, já que, por vezes, bastaria que a frase fosse pontuada com ponto de exclamação para que se tornassem aceitáveis ³¹².

Também o reduzidíssimo número de ocorrências desviantes de **conectores** foram incluídas neste ponto. Como disse na análise das ocorrências *PF*, os aprendentes limitam-se a um conjunto formado por quatro ou cinco conectores de alta frequência (*que, porque, quando, por isso*). Os poucos desvios, como se pode ver nos exemplos abaixo, ocorrem sobretudo em itens que compartilham o espaço semântico com outros (*quando/logo que/enquanto; embora/ainda que/mesmo que/apesar de*).

- E08** *Assim que temos de vender o andar para poder ter pago tudo. **Enquanto** tenha novidades já te contarei.*
- F07** *Lembro me deste dia. Ficavamos durante todo o dia na praia. O sol era quente **mesmo que** estivessemos em dezembro.*
- A03** *Um dia fui à Baixa para comprar um casaco. O tempo estava muito quente e não senti-me muito bem. **Quando** tinha comprado o casaco queria voltar para casa. Por isso tina de apanhar o 44. [...] Mas o que podia fazer? Nada **mas** esperar.*
- A09** *Eu esperei uma hora e tu não chegaste. Então tive dúvidas: ‘esta hora é correcto?’. Não sabia nada. Hoje de manhã, **quando** tomei o pequeno-almoço, ouvi as notícias de televisão e no fim o homem disse: ‘São 11.00 horas, bom dia.’ Eu vi o meu relógio e foram 12.00 horas.*
- A29** *Sentar-se por uma bicazinha num café favorizado é o aspecto mais importante sobre a vida cultural em Portugal. <Embora / quer/ esteja a > **Embora** esteja a chover eu vou quatro vezes por dia para um café.[...] **Mesmo que** os preços sejam um pouco caro para a minha bolsa e <adoro ir> gosto de ir para ao ‘Bar*

³¹⁰ Para os usos de **todo / tudo**, ver Bechara 1999:196-199.

³¹¹ Veja-se também **F24** *Sem esquecer a religião egípcia que era **muita** complicada e **muita** rica.*

³¹² Também segundo Ilari «um outro conjunto de palavras tradicionalmente classificadas como advérbios que colocam em xeque os critérios tradicionais», e que também se justifica tratar como classe bem configurada, são os intensificadores. (Ilari 1996:110-111).

Tradicional’.

- A35** *Eu comecei de rir porque não percebi nada e a minha amiga não percebeu **que** eu estivesse a rir. ‘Mas porque, isto é muito bom! Eu gosto também! Porque é que estás a rir tão muito?’*
- S07** *Então, o meu avião que tinha chegado de Copenhague chegou na hora h. Ainda todas as pessoas divertiram-se, talvez **porque** os copos que eram servidos durante a viagem.*
- S30** *Quando eu tenho um encontro com um/uma português/a eu sempre ficava esperar por ele/ela. [...] E os desculpes que eles têm são ‘boas’, - Ah, desculpe! eu esperei por o autocarro um meia-hora! – Ah, desculpe não procurei um lugar de meu carro! [...] E todos são verdades, em Portugal não tenho dificuldades de venho atrásada por um encontro **porque** o transito.*

É interessante constatar que, apesar de serem muito frequentes na sala de aula exercícios para treino das concessivas, os conectores que as introduzem têm, no *PESTRA*, tal como acontece no *sub-corpus* português, uma frequência muito baixa (*embora*: PT 4, ESP 7, FR 4, AL 4, SU 2, CH 9; *ainda que*: PT 2, SU 2; *mesmo que*: PT 1, ESP 2, FR 3, AL 1, SU 1). Quanto a *apesar de*, que exige uma oração infinitiva, enquanto os portugueses o usam 9 vezes, ele é usado apenas 1 vez por espanhóis e chineses, 2 por franceses, e nenhuma por suecos; os alemães, usam 4 vezes não *apesar de*, mas *apesar disso*, ou seja, substituem a oração infinitiva, típica do português, por um sintagma preposicional³¹³. Também as ocorrências de *quando* (por um lado, PT 31, ESP 39, FR 36; por outro, AL 63, SU 62, CH 61) e os seus usos desviantes (AL 4 e SU 4) poderão indiciar estádios de aquisição diferentes entre os dois grupos falantes de línguas românicas e os outros três; estas diferenças poderão estar relacionadas, e assim deverão ser investigadas, com o contexto mais vasto da expressão do tempo e do aspecto (ver o exemplo acima de A03).

Em resumo, foram incluídos neste ponto itens lexicais para que, em muitos casos devido à sua polivalência, abordagens mais recentes no âmbito da gramática descritiva têm feito propostas de classificação diferentes das da gramática tradicional. Como salientei antes, essa polivalência, do mesmo modo que causa dificuldades à descrição, também exige, ao nível da aquisição e uso, um tratamento minucioso e que ultrapassa os limites da semântica lexical. Daí que, a análise aqui

³¹³ Exemplo: **A26** ... *a festa do fim do ano 1993 [...] Não dormi muito durante essa noite mas **apesar disso** gostava a noite, a festa e o início do meu ano em Portugal.*

apresentada não pretenda mais do que avaliar quais os itens lexicais que concentram maior número de desvios nesta fase de aquisição. No entanto, esta análise exploratória parece evidenciar, ao nível do léxico mental, uma relação estreita entre os diferentes itens lexicais que exprimem, entre outras, relações deícticas, por um lado, e de quantificação, por outro; ou seja, uma abordagem, para o ensino do léxico, por categorias nocionais, tal como a que a descrição linguística mais actual propõe, e que envolve também outros processos linguísticos, parece ser muito mais adequada do que a que tem sido praticada até agora, que incide, por exemplo, em demonstrativos, possessivos, e indefinidos.

6.3.3.6. *Recategorização*

Do mesmo modo que certos itens lexicais põem problemas ao tratamento linguístico, também oferecem dificuldades aos aprendentes no que respeita às suas propriedades sintácticas. Ou seja, o aprendente tem uma representação de um dado item cujas **propriedades sintácticas** não coincidem com a dos falantes nativos e usa um vocábulo pertencente a uma determinada categoria como se ela pertencesse a outra. É o caso do par *tão/tanto*, tratado em quantificação. No entanto, e como se pode ver nos exemplos, esse tipo de desvios está, em muitos casos, relacionado com outras questões, entre outras com a morfologia derivacional e com características de certos verbos.

E08 *Como já sabes, estamos morando em Lisboa há pouco tempo. As pessoas são simpáticas. [...] Mas também tenho que dizer-te que ter de começar de novo, por uma parte é **bem** mas por a outra é difícil.*

E36 *Uma coisa muito diferente entre os dois países é o horário das comidas. [...] O jantar nunca é antes das nove e meia ou dez horas, por tanto vão deitar-se com os estômagos cheios demais. Dormem **mau**, mas eles nunca tentaram de mudar as coisas.*

F32 *Em 1936, chegaram os férias com pagamentos e desde este ano, o **numeroso** de viagens sempre cresceu.*

A08 *Neste momento, tenho uma namorada bonita, com um rosto aberto e uma beleza **natureza**.*

A18 *Acho que cada profissão tem atrações e tem <'contras'> **atrasados**.*

A36 *Os Alemães são sempre eficiente, também quando eles matam! Os Alemães gostam muito este adjectivo de ser 'eficiente', mas eu penso que, por*

exemplo, na economia é muito **bem**, mas nas guerras e revoluções é muito **mal**...

- A42** Na maioria são jovens, que são culpados do acidente. Eles não pensam e bebam os copos. E depois acham **bom**, conduzir com álcool.
- A47** Ultimamente, o trânsito em Lisboa tem crescido enorme. [...] O <elargamento> **crescido** rápido das cidades é a coisa para muitos outros problemas: muitas pessoas faltam habitação ou vivem em condições subhumanos.
- S21** As vezes fui na praia com minha família, meus primos, tias e avôs. Eu não recordo um dia de **chover** ou muito vento durante verão quando era criança.
- S36** Receita pelas magras pessoas [...] Frite o bife no muito óleo. Põe a nata em cima do bife e ao fim põe muito sal e pimenta no bife. O bife sabe muito **bom** e tem, com certeza, muitas calorias.
- S40** Depois de bebeu, talvez não sentem-ses diferentes, mas são. Vocês não reagirem com o mesmo **rápido**, e virem pior.
- S46** ... andar nas ruas de Lisboa [...] Andar nas ruas pode ser as vezes uma verdadeira aventura. Mas é uma aventura que faz **bom**, pelo menos para mim, É esta vida que eu gosto, e por isso não quero mudar para uma cidade pequeno.
- S47** Por exemplo quando uma pessoa faz uma grande **errada**, a policia pode tirar o cartão de conduzir, porque na Suecia a policia faz isto! Talvez os portugueses pensem que eu esteja maluco, talvez seja verdade!
- C04** Estamos nesta sociedade melhor já entramos numa época nova, há muito <acontecime> **acontecido**, ninguém acredita, é muito estranho e **maravilha**, também é injusto.
- C13** E em Lisboa, muitas coisas são bonita. Especialmente as igrejas, gosto muito! E também posso estudar **histórica** português ao vivo!
- C26** O meu amigo [...] pediu nos passe à frente na bicha. Eu fiz, embora eu tenha **envergonhada**, e o meu amigo também, eu estava muito **felizmente**.
- C30** Normalmente televisão <é muito divertido> dá nos **divertido**. Mas a televisão <há> também há algum coisa <mal> mau.
- C40** E também não tem forma para atriar os patrões dexarem em Macau. Acho que é muito **mal** para o futuro de Macau. Tem sorte, o turismo está bem **desenvolvimento** em Macau, e agora o governo de Macau precisa continuar para o progresso.
- C41** Portugal ainda é um dos países europeus que tem menos **desempregos**.
- C47** Engarrafamento é um problema mundial [...] Eu acredito que o maneira que resolver o problema é de construir mais túnel e maiorar as ruas. Este maneira é feita em Campo Grande e Campo Pequeno, tem um resultado **bem** porque agora há menos tempos para passar naquelas zonas.
- C49** Não gosto muito viver em bairro antigo.[...] Claro, num bairro antigo há muitos monumentos, prédios antigos [...] Mas viver lá não é **mal**, mas mais ou menos, há dificuldades.

Os pares *bom bem* e *mal/mau* parecem pôr problemas a todos os grupos, excepto aos franceses. Mas um olhar mais atento mostra-nos que os desvios ocorrem, sobretudo, em contextos com os verbos *ter, ser* e *estar*. Este facto pode estar relacionado, como vimos antes, com a especificidade destes verbos, numa perspectiva tipológica (ver Viberg 1993a). De facto, eles são particularmente frequentes no caso dos chineses, e são típicos de um estágio de aquisição em que há também muitos desvios de outras naturezas. Por outro lado, eles parecem ocorrer quando uma dada construção pode ter qualquer conexão com outra ou outras frequentes e que, por isso mesmo, fará parte do léxico mental do aprendente (ver Goldberg 1995). Por exemplo: *é muito bom! / sabe muito bom!; faz bom tempo / faz bom; é uma maravilha / é estranho e maravilha; estou bem, felizmente / eu estava felizmente; sou envergonhada / tenho envergonhada*. Esta hipótese parece ser confirmada pela sequência rasurada em C30, “<é muito divertido>”, e substituída por “dá nos divertido”.

É também em textos globalmente mais desviantes, mas não só, que ocorrem desvios entre pares de nome / adjectivo (*natureza/natural; rapidez/rápido*); mas mais frequentes são os desvios nos pares nome / adjectivo verbal (*erro/errado; divertimento/divertido; desenvolvimento/desenvolvido; acontecimento/acontecido*). Na verdade, as formações com o sufixo *-mento* parecem pôr alguns problemas. Mais uma vez as rasuras em A47, “<elargamento>” substituído por *crescido*, e em C04 “<acontecime>” substituído por *acontecido*, parecem confirmar esta hipótese. Esta questão está certamente relacionada com o que foi dito em vocabulário indisponível, a propósito de neologismos formais; ou seja, neste estágio de aquisição, os aprendentes não parecem dispor de um conhecimento processual muito alargado no que respeita a morfologia derivacional.

6.3.3.7. Conclusão

Da observação dos dados relativos a desvios na selecção do léxico, podemos concluir que:

- 1. nomes e adjectivos** são as categorias que apresentam menos e mais dispersas ocorrências desviantes; por outro lado, o seu número não parece estar tão relacionado com a

distância linguística como no caso das restantes categorias. O vocábulo *gente*, com uso marcado em português, é aquele que concentra o maior número de desvios.

2. os 10 **verbos** mais frequentes no *PESTRA* estão na origem de 50 % do total das ocorrências desviantes no que respeita a selecção dos verbos; os 20 verbos mais frequentes estão na origem de 55 %; ou seja, **são os verbos específicos das línguas da Europa ou específicos das línguas românicas ibéricas, portanto mais marcados numa perspectiva tipológica, aqueles que colocam mais dificuldades à aquisição.** Os problemas põem-se no estabelecimento de fronteiras entre *ter / haver, ser/estar/ficar, e ter de/dever/poder/saber* a que o PE acrescenta *ser capaz/conseguir*, difíceis de estabelecer também são as fronteiras entre os verbos *dar* e *fazer*, bem como outros verbos com características específicas em PE, como é o caso de *tomar, andar* e *vir*.

3. os casos de omissão de **preposições** são aqueles que motivaram maior número de desvios; essas omissões ocorreram sobretudo em contextos em que o falante nativo teria usado a preposição *a*, ou seja, a mais fraca do ponto de vista semântico; de facto, os resultados mostram que em quase 41 % das situações em que ela está envolvida há uma qualquer forma de desvio (omissão, inserção, ou uso desviante). Seguem-se-lhe *por* e *para*, com 35 % e com 22 %, respectivamente. A preposição ***de***, sendo a mais frequente, apresenta apenas 13,5 % de ocorrências desviantes; *em*, a segunda mais usada, não chega aos 10 % de ocorrências desviantes. Resta acrescentar que **a maior parte dos desvios em preposições está associada a verbos de alta frequência**, aos verbos de movimento (*ir, chegar, voltar*), no que respeita a preposição *a*; e, no caso da preposição *de*, à sua omissão em funções de apoio (*gostar de, precisar de*).

4. em outras categorias incluíram-se itens lexicais a que a gramática tradicional atribui a designação de advérbios, pronomes e conjunções; mas, porque ao nível do texto ou do discurso, cada um desses itens pode assumir comportamentos e valores muito diversos, abordagens mais recentes têm procurado identificar e caracterizar os seus diferentes valores sintáctico-semânticos e pragmáticos. Muitos deles estão envolvidos em operações deícticas e de quantificação, representando estas, respectivamente, 25 % e 30 % do total de desvios aqui incluídos. A estes números, acrescenta-se 15 % de ocorrências desviantes do primitivo advérbio temporal *então*, a que os aprendentes parecem atribuir as funções conclusivas dos conectores *por isso* ou *portanto*.

5. em recategorização, incluíram-se as ocorrências que resultam do uso de um item de uma determinada categoria sintáctica como se ele pertencesse a uma outra. Entre este tipo de desvios, que parecem resultar da deficiente atribuição das propriedades aspectuais aos verbos copulativos e predicativos, destacam-se os pares adjectivo / advérbio *mau/mal* e *bom/bem*. Acrescentam-se a estes alguns desvios em pares nome / adjectivo (como *natureza/natural*; *crescimento/crescido*) que parece estarem relacionados com a aquisição da morfologia derivacional.

Em conformidade com o que tinha sido dito nos capítulos 1 e 2, esperava encontrar-se um maior número de desvios em itens lexicais mais dessemantizados, mais polissémicos e nos seus usos menos prototípicos. Pôs-se ainda como hipótese que um dado item pode fazer parte do vocabulário do aprendente mas que, devido à representação que dele tem, poderia não estar disponível para um dado uso e que, nesse caso, ele seria substituído por outro. De facto, **a maior parte dos desvios não resulta de indisponibilidade vocabular, resolvida com a sua substituição por um item disponível que é generalizado.** Em vez disso, nesta fase de

aquisição, as ocorrências desviantes **resultam de modulações nas fronteiras semânticas de uma entrada e das suas relações paradigmáticas com outras a que cada uma delas está associada no léxico mental. Assim, a maior parte dos desvios ocorre em itens lexicais muito frequentes e que os aprendentes mostraram ter disponíveis para outros contextos.**

Nas secções anteriores de 6.3 *Léxico deficitário*, sob o título 6.3.1 *Desvios formais em vocabulário disponível* foram classificadas todas as ocorrências com uma configuração não coincidente com o seu modelo em PE. Sob o título 6.3.2 *Vocabulário indisponível e estratégias compensatórias* foram classificadas todas as ocorrências resultantes de empréstimos de outras línguas, que sofreram ou não processos de relexificação, bem como produtos lexicais criados pelos aprendentes a partir de material linguístico do português e ou de outras línguas. Nesta secção, sob o título 6.3.3 *Desvios na selecção do léxico* foram classificadas todas as ocorrências resultantes de selecção inadequada de um dado item para um dado contexto. **Desvios na selecção do léxico representam uma média de 27 % do total absoluto de desvios. No entanto, como pode ver na TABELA 23, a sua distribuição pelos grupos é muito diferente daquela que se verificava no caso de desvios formais e de desvios resultantes de indisponibilidade lexical.**

Tabela 24

LÉXICO DEFICITÁRIO	espanhol	francês	alemão	sueco	chinês
desvios formais em vocábulos disponíveis	53,9	41,6	33,9	30,3	20,4
desvios resultantes de indisponibilidade lexical	15,5	15	9,9	8,3	7,6
% em relação ao total absoluto de desvios	69,4	56,6	43,8	38,6	28
desvios em selecção do léxico	16,6	22,9	29,5	28,5	33,1
% em relação ao total absoluto de desvios	86	79,5	73,3	67,1	61,1

Como se pode verificar lendo a TABELA 24, **enquanto os desvios formais e os desvios resultantes de indisponibilidade lexical apresentam uma forte tendência a diminuir à medida que a distância**

linguística aumenta, no caso dos desvios resultantes da selecção inadequada de um dado vocábulo para um dado contexto, essa tendência inverte-se. Por outro lado, a mesma tabela também nos fornece informação quanto ao total de desvios que, neste trabalho, foram considerados como directamente resultantes de uma introdução incorrecta ou incompleta de uma palavra num dos sistemas de representação. Mas, representações desviantes das propriedades semânticas e sintácticas de certas entradas lexicais reflectem-se certamente no seu comportamento numa frase ou num enunciado. É disso que trata o ponto seguinte: *O léxico para além do léxico.*

6.3.4. *O léxico para além do léxico*

Em trabalhos desta natureza, é por vezes indispensável tomar decisões, ainda que o estado dos conhecimentos não ofereça fundamentação segura e adequada para elas ou possa ser de difícil aplicação. Como ressalta da bibliografia referida no Capítulo 1, este é o caso das fronteiras entre léxico, morfologia e sintaxe. Bybee 1985, como vimos, prefere até propor um *continuum* linguístico, dependente de saliência cognitiva e cultural, que vai da expressão lexical à sintáctica. É claro que, como vimos no Capítulo 2, os diferentes padrões de lexicalização que diferentes línguas adoptam constituem um sério problema para o processamento e para a aquisição de uma L2. Eles são também um obstáculo inultrapassável quando se trata de estabelecer quais os itens que constituem uma entrada lexical e qual o conteúdo de cada uma delas. No entanto, uma coisa parece certa: a representação que se tem de um determinado item lexical tem consequências nas construções em que ele está envolvido. Nesta linha de pensamento, em Leiria 1991, estudei a aquisição dos aspectos verbais expressos pelos Pretéritos Perfeito e Imperfeito. Tratava-se portanto de uma questão que envolvia relações entre léxico, morfologia e sintaxe. A investigação que aqui apresento, como disse na Introdução, é uma extensão dessa outra e foi motivada por ela. O título desta secção 6.3.4 – *O léxico para além do léxico* – pretende dar conta das consequências que as propriedades de um determinado item lexical vs. a representação desviante têm, em certos casos, na sintaxe mas muito em particular na morfossintaxe.

E porque o principal objectivo desta investigação consiste em avaliar a dimensão e as características do léxico mental de falantes de português L2 em situação de aprendizagem semi-formal, para que os seus resultados possam, de algum modo, ser aplicados ao ensino, tenho vindo a salientar o facto de certos vocábulos serem particularmente motivadores de ocorrências desviantes e de, conseqüentemente, numa abordagem lexical, lhes dever ser prestada particular atenção. Em conformidade com esse objectivo, **neste ponto, serão comentados muito brevemente certos desvios que se manifestam na sintaxe ou na morfossintaxe e que envolvem com frequência determinados vocábulos, podendo estar relacionados com propriedades do item em causa ou de outros com que ele co-ocorre.**

Visto que neste ponto serão sobretudo comentados desvios na morfossintaxe, convém relembrar aqui um facto, sobre **o ensino**, a que aludi na Introdução. Ao contrário daquilo que acontece com a maior parte dos aspectos do léxico tratados até aqui, em certos ambientes de sala de aula, é ocupado muito tempo com a realização de “exercícios de gramática”, exercícios esses que incidem, em boa parte, na morfossintaxe. É claro que esse procedimento tem conseqüências na fixação da gramática da interlíngua e ajuda certamente a prevenir o desenvolvimento de variantes pidginizantes (ver, capítulo 4, *The Basic Variety*)³¹⁴. **Mas será que essa exercitação, centrada na morfossintaxe, produz efeitos igualmente evidentes em todos os casos? Ou será que a deficiente fixação das propriedades das palavras, observadas até aqui, não se fará sentir para além do léxico?** A TABELA 25 permite algumas observações muito rápidas a propósito.

No que respeita a **ordem das palavras**, a colocação dos clíticos costuma ser um dos aspectos sintácticos que ocupa mais tempo na sala de aula (até mesmo com aprendentes que ainda estão muito longe de um estágio de aquisição em que virão a usá-los espontaneamente...). Porque a modalidade escrita lhes permite aplicar conhecimento explícito repetidamente fornecido pelo professor e porque, nesta fase, o recurso a processos anafóricos é muito limitado, o número de desvios no que respeita

³¹⁴ Certos estudos sobre o papel do ensino na aquisição e desenvolvimento linguístico devem ser realizados preferencialmente em contextos de aprendizagem em que a variável imersão esteja ausente, ou seja, em aprendizagem de língua estrangeira. Ver Capítulo 7.

a colocação de pronomes de acusativo e dativo é reduzido e, por outro lado, as omissões do pronome de flexão reflexiva são muitas. Mas, além destes, há um número relativamente pequeno de vocábulos que concentra a maior parte dos desvios em ordem de palavras. Entre eles contam-se certos adjectivos muito frequentes (*novo, pequeno, grande, diferente*) e alguns advérbios (*não, nunca, quase*). Estes dados, mais uma vez, aconselhariam uma chamada de atenção sobre sequências do tipo *uma grande festa, uma grande cidade, grandes espaços, ou uma pequena ajuda*. **Mas é de assinalar que 32 % dos 221 itens lexicais que não respeitam a ordem de palavras que um falante nativo de PE adoptaria são da responsabilidade de apenas dois vocábulos: *também* (24) e *sempre* (47), bem distribuídos por todos os grupos e por um grande número de informantes.**

Tabela 25

<i>DESVIOS SINTÁCTICOS, MORFOSSINTÁCTICOS E OUTROS</i>						
	espanhol	francês	alemão	sueco	chinês	Totais
organização sintáctica	8	11	26	48	57	150
ordem das palavras	22	28	54	68	49	221
colocação do pronome pessoal	22	25	21	11	10	89
determinação dos nomes (det+ art)	38	82	101	188	287	696
morfologia flexional: concordância verbal	15	18	23	36	50	142
concordância nominal em género	0	0	21	21	32	74
concordância nominal em número	5	13	30	27	44	119
flexão casual (clíticos)	2	5	12	8	33	60
tempo verbal	3	23	21	58	67	172
modo verbal	25	29	50	64	29	197
aspecto verbal:						
pret. perf. simples/ pret. imperfeito	3	12	66	60	57	198
pret. perf. simples/ pret. perf. composto	4	6	6	9	1	26
infinitivo	8	9	4	24	31	76
OUTROS ³¹⁵	22	26	35	51	96	230

³¹⁵ Em 'OUTROS' incluem-se ocorrências desviantes que parecem não se enquadrar em nenhum dos títulos incluídos; ou porque tinham uma expressão numérica muito reduzida e só envolviam um ou dois dos grupos (por exemplo, em FR e AL, casos relacionados com a dupla negação; opção lexical improvável, em ESP e FR, do tipo *ascensor* por *elevador*; inserção abusiva de *já*, em CH, relacionada certamente com características específicas desta língua no que respeita o aspecto verbal) ou porque, na fase actual da investigação, não foi possível interpretá-los.

total	177	287	470	673	843	2450
total absoluto de desvios	1325	1422	1770	2061	2068	8646

Ana Cristina M. Lopes 1996, no texto citado antes, sobre o item *então*, referia que, «*então* não é a única expressão que admite usos temporais e usos argumentativos: *agora*, *sempre*, e *logo* também exibem idêntica ‘flutuação’ semântica.» (M. Lopes 1996:187). E Óscar Lopes afirma: «A partícula ‘sempre’ tem um valor pragmático distinto do seu valor temporal, que se caracteriza por uma marca sintáctica e outra semântica, que são respectivamente a colocação à cabeça da frase e um valor alheio a qualquer relação tempórea.» (Óscar Lopes 1996:195). Os dados em apreço mostram que os aprendentes se restringem aos seus usos temporais, mas que, no entanto, em 36 % dos casos colocam a partícula à cabeça da frase. Contudo, é de lembrar que em PB ela também ocorre frequentemente nesta mesma posição com valor temporal ³¹⁶. Esta questão merece certamente um tratamento mais aprofundado, enquadrado no contexto da expressão da temporalidade, da coesão temporal, e da colocação das expressões de valor temporal.

A omissão ou inserção abusiva do artigo representa 8 % do total de ocorrências desviantes no *PESTRA*. É claro que este tipo de desvios resulta, do ponto de vista do aprendente, de processos complexos de aquisição da **determinação dos nomes**, que envolvem, entre outras, a aquisição da expressão da quantificação e das especificidades aspectuais (ver Soares 1997) e que estão relacionados com propriedades do nome, mas também com escolhas caso a caso da variedade linguística (ou do falante...). Desta última possibilidade é exemplo a variação no **uso ou não de artigo com nomes próprios e**, muito em especial, **com nomes próprios geográficos**, nomes de países (por exemplo: Angola, Portugal, Moçambique, Marrocos; mas, a Alemanha, o Brasil, a Guiné-Bissau, a Suécia; mas, Espanha/a Espanha, França/a França, Itália/a Itália), ou de cidades ou vilas (Lisboa, Paris, Porto de Mós, Rio Maior, Santarém; mas, o Porto, o

³¹⁶ “A mamã vai trazer azeitonas da feira. Ela sempre traz.” (telenovela *Laços de família*, episódio exibido pela SIC no dia 10.2.2001).

Cartaxo, o Rio de Janeiro; mas, Marco de Canaveses/o Marco de Canaveses). É portanto uma questão que exige muito tempo de contacto com a variedade alvo, e que, nos dados em apreço, **soma 10 % das ocorrências desviantes em determinação dos nomes** (ESP 3; FRAN 5; AL 21; SU 23; CH 16). No entanto, uma abordagem mais centrada no léxico poderá, pelo menos, chamar a atenção e treinar na sala de aula os topónimos mais relevantes para um determinado aprendente ou grupo (ver entre outras, AL: *em Alemanha* 8; SU: *em Suécia* 7, *em Dinamarca* 1, *em Finlândia* 1, *o Estocolmo* 2) ou relevantes para todos (*o Alentejo, Alfama, o Algarve, a Ásia, o Brasil, Cascais, Coimbra, Évora, a Europa, Portugal, etc..*) tentando evitar as ocorrências desviantes.

A omissão do artigo antes de possessivo representa 13 % das ocorrências desviantes incluídas neste ponto (ESP 20; FRANC 13; AL 19; SU 16; CH 25). A casualidade do número de ocorrências desviantes ser muito semelhante entre os vários grupos e de, na sua totalidade, eles serem por omissão pode dever-se ao facto de, como em italiano, mas ao contrário do que acontece em espanhol e em francês, os usos do possessivo sem artigo, quando precedem o nome, constituírem casos marcados no PE (em formas de tratamento: *Meu querido Mattias*; em exclamações: *Meu Deus!*). Embora os possessivos sejam objecto frequente de exercícios na sala de aula, é possível que nem sempre o *input* disponibilizado em tais circunstâncias seja suficientemente transparente quanto à distinção entre uso frequente e uso marcado; isto pode dever-se também à realização de exercícios do tipo – *De quem é este livro?* – *É meu*, a par de – *Que livro é aquele?* – *É o meu*, bem como ao contacto com expressões em que ele segue o nome, em sequências como *um amigo meu disse-me que* ou *tive notícias tuas*. Este resultado parece, portanto, aconselhar muito mais atenção à selecção de modelos a apresentar na sala de aula, tentando que os possessivos sejam contextualizados em frases não ambíguas. Sabe-se, por outro lado, que a determinação é uma das questões mais complexas, linguisticamente.

Até aqui temos a impressão de que todos os grupos tiveram o mesmo comportamento no que respeita o uso dos artigos. Mas, de facto, assim não é, como pode ver nos resultados fornecidos pela *TABELA 25*. Além dos casos referidos, **exceptuando o grupo espanhol** (inserção 8; omissão 3), **todos os**

outros apresentam resultados que aumentam com a distância linguística e que evidenciam dificuldades quanto à distribuição dos artigos, muito em particular, do artigo definido (FR: inserção 28, omissão 21; AL: 31, 18; SU: 67, 61; CH:73, 114). **A referência anafórica, a referência genérica e a quantificação são responsáveis pela grande número de ocorrências desviantes nos dados** dos outros grupos ³¹⁷. A estas podem ainda acrescentar-se algumas sequências lexicais, incluídas e tratadas em combinatórias (ex: *todas vezes, a espera por todas as vezes e à espera*).

Mas, ao contrário do que acontece com os nomes próprios ou com os possessivos, estas questões não costumam ser abordadas na aula nem sequer vêm adequadamente descritas em materiais para o ensino e aprendizagem de L2 ³¹⁸. E, mais uma vez, do mesmo modo que a sua explicitação é difícil também é a sua aquisição, e a prová-lo está, mais uma vez, o aumento de ocorrências desviantes em função da distância linguística.

Referi antes que a escolha do léxico como objecto de estudo veio na sequência da investigação que realizei sobre *A aquisição, por falantes de Português Europeu, língua não materna, dos aspectos verbais expressos pelos Pretéritos Perfeito e Imperfeito*. Orientaram a investigação as seguintes hipóteses gerais:

A maior ou menor dificuldade de aquisição manifestada pelo falante não nativo, no que respeita a aquisição dos aspectos verbais expressos pelos Pretéritos Perfeito e Imperfeito, está relacionada com (1) propriedades inerentes ao predicador, com (2) a língua materna de cada um, e com (3) características específicas do Português. (Leiria 1991:102).

Por limitações de tempo e do programa de tratamento de dados usado na altura, não me foi possível levar a cabo, das 6908 ocorrências de

³¹⁷ Em operações de determinação, há ainda a considerar um pequeno número de ocorrências desviantes, mas muito sistemáticas quanto aos itens lexicais envolvidos, relacionadas com a distinção entre **nomes contáveis e massivos** (*roupas*: AL 1, SU 3; *comidas*: ESP 2, FRAN 3, AL 4, SU 1). Refiro-os não pela sua expressão numérica mas porque estão directamente relacionados com esta propriedade do léxico a que, até este momento, não tinha feito referência.

³¹⁸ Tendo em consideração a escassez de estudos sistemáticos para o português sobre o artigo, pode ser de grande utilidade Bosque & Demonte (dir.) 1999, *Gramática Descriptiva de la Lengua Española*.

Pretérito Perfeito e Imperfeito existentes no *corpus* estudado ³¹⁹, uma análise em contexto que permitisse distinguir ocorrências desviantes e não desviantes. Apesar disso, e porque os dados foram analisados tendo em consideração a classe aspectual do verbo e o número de formas esperadas e de formas obtidas, foi possível concluir que, de facto, os aprendentes atribuem a um verbo o morfema que é mais relevante para a raiz desse mesmo verbo; ou seja, particularmente quando a língua materna não oferece apoio para a distinção perfectivo / imperfectivo, de que a distinção acabado / não acabado codificada pelos Pretérito Perfeito / Imperfeito é uma variante, os aprendentes optam por não contrariar valores aspectuais inerentes ao predicador. Assim, nos primeiros estádios de aquisição, parecem restringir o Imperfeito a estados e o Perfeito a situações dinâmicas; e porque a relação *télico instantâneo + morfema de Pretérito Perfeito* e *estativo + morfema de Pretérito Imperfeito* é a mais relevante, transparente e uniformemente codificada no *input* ela é mais cedo incorporada no sistema interlinguístico, tal com acontece na L1. Entretanto, e à medida que vai sendo exposto a *input* suficiente e adequado, o aprendente vai descobrindo que o Pretérito Perfeito se estende a processos acabados e a estados; e que o Imperfeito se estende também a eventos iterativos e a eventos habituais.

O programa usado nesta dissertação permitiu avaliar, no *PESTRA*, como se pode ver no *ANEXO 14*, o número de ocorrências de cada tempo verbal, por grupo, e distinguir entre formas adequadas ao contexto e formas desviantes ³²⁰ (ESP: 3,6 %; FRAN: 5,3 %; AL: 10 %; SUE: 13 %; CH: 11,7 %). Em relação às formas desviantes, permitiu ainda verificar, caso a caso, qual seria o tempo verbal adequado a um dado contexto. A *TABELA 26* resume essa informação, dando o resultado obtido em percentagens, por grupo, para cada um dos tempos verbais que registaram maior número de ocorrências.

³¹⁹ O *corpus* estudado em Leiria 1991 «é constituído por 218 documentos produzidos por 168 informantes falantes de 16 línguas maternas diferentes [árabe; chinês; norueguês, sueco, neerlandês, inglês, alemão; espanhol, francês, italiano; polaco, servo-croata, búlgaro; persa; hindi] com idades compreendidas entre os 18 e os 55 anos de idade, sendo 59 % do sexo feminino e 41% do sexo masculino. Cada documento consiste no relato de uma de três narrativas ouvidas pelos informantes aquando da prova de compreensão oral e produção escrita do exame do Curso Básico de Língua Portuguesa do Departamento de Língua e Cultura Portuguesa da Faculdade de Letras de Lisboa.» (Leiria 1991: v, vol. *Corpus*).

³²⁰ Formas verbais desviantes foram marcadas com o código X5).

Tabela 26

Percentagem de ocorrências desviantes por tempo verbal					
TEMPOS VERBAIS	espanhol	francês	alemão	sueco	chinês
Presente do Indicativo	16	22,6	12,1	18,8	23,8
Pretérito Imperfeito do Indicativo	6,4	15,3	29,6	24,7	21,3
Pretérito Perfeito Simples do Indicativo	10,6	16,1	27,9	23,4	26,7
Futuro Simples do Presente do Indicativo	1,1	2,4	0,7	0,3	1,1
Futuro do Pretérito Simples do Indicativo	3,2	0,8	1,8	1,1	1,7
Pretérito Perfeito Composto do Indicativo	5,3	4	1,8	2,6	0,3
Mais que Perfeito Composto do Indicativo	2,1	3,2	2,1	2,2	3,7
Presente do Conjuntivo	13,8	12,9	9,3	8,6	3,7
Pretérito Imperfeito do Conjuntivo	8,5	2,4	4,3	4,6	1,4
Futuro do Conjuntivo	4,3	4,8	3,6	3,8	5,1
Tempos Compostos do Conjuntivo	2,1	1,6	0,4	0,3	-
Imperativo (afirmativo)	13,8	-	0,4	0,3	-
Imperativo (negativo)	10,6	-	-	-	-
Infinitivo Simples	13,8	6,5	2,1	6	8,7
Infinitivo Flexionado	10,6	5,6	1,8	2,4	2

A TABELA 26 deve ser lida tendo em consideração que cada resultado representa a percentagem intragrupo; ou seja, por exemplo, porque os espanhóis totalizam apenas 3,6 % de formas verbais desviantes, enquanto que os suecos totalizam 13 %, aos 16 % de formas desviantes no Presente do Indicativo, no caso dos primeiros, correspondem apenas 15 formas verbais usadas indevidamente no Presente do Indicativo ou usadas em outro tempo verbal quando o contexto exigia esse tempo verbal, enquanto que aos 18,8 %, nos suecos, correspondem 69 formas nas mesmas circunstâncias. Do mesmo modo, devem ser interpretados, por exemplo, os resultados relativos ao Presente do Conjuntivo ou aos Infinitivos. De facto, os espanhóis apenas erraram em 13 % das ocorrências de Presente do Conjuntivo e os franceses em 11 %, enquanto que alemães, suecos e chineses em 27 %, 28 % e 22 %, respectivamente. Quanto aos Infinitivos, o número baixo de ocorrências desviantes, deve ser lido, em geral, como correspondendo ao escasso uso do Infinitivo Flexionado.

O objectivo da *TABELA 26* consiste em pôr em evidência os tempos verbais que parecem pôr mais problemas a cada um dos grupos. A negro, destacam-se, intragrupo, os quatro tempos verbais em que as ocorrências desviantes atingiram uma percentagem mais alta. Como é óbvio, não farei aqui um comentário detalhado dos resultados. Acrescento apenas que os resultados a negro, quando lidos intragrupo, se explicam uns aos outros. No caso dos espanhóis, por exemplo, os problemas situam-se na selecção do Presente do Indicativo *vs.* Presente do Conjuntivo, enquanto que, no caso dos franceses, eles colocam-se, sobretudo, na selecção de Presente do Indicativo *vs.* Pretérito Imperfeito. No caso dos chineses, eles parecem envolver os três tempos verbais mais frequentes do modo Indicativo, enquanto que, no caso dos dois grupos germânicos eles se concentram na oposição Pretérito Perfeito *vs.* Pretérito Imperfeito. Como disse antes, não comentarei aqui estes resultados e o seu significado no que respeita a aquisição do Tempo, Modo e Aspecto. No entanto, porque isso cabe no âmbito deste trabalho, relacionarei os desvios em tempos verbais com alguns dos verbos em que eles ocorrem.

Sendo as línguas germânicas, por um lado, sistemas que privilegiam a marcação do Tempo e não do Aspecto, e, por outro lado, no que a este diz respeito, adoptando uma perspectiva muito diferente das línguas românicas, os resultados aqui obtidos não fazem mais do que confirmar os de Leiria 1991 ³²¹. Quanto ao chinês, enquanto grupo de sistemas linguísticos em que o aspecto é prioritário, a análise aqui realizada, porque envolve também outros tempos verbais, nomeadamente o Presente do Indicativo, e porque a análise foi feita em contexto, permite ir mais longe na explicação dos desvios, em número superior ao que seria de esperar. A *TABELA 27* dá conta das ocorrências desviantes em verbos de alta frequência no *PF* e no *PESTRA* ³²².

Tabela 27

Desvios na selecção do tempo verbal adequado ao contexto em verbos de alta frequência
--

³²¹ Para informação respeitante às diferentes perspectivas aspectuais das línguas envolvidas neste estudo (românicas, germânicas e chinês) e outras (outras indo-europeias: inglês, eslavas, hindi e persa; e árabe), remeto para Leiria 1991.

³²² Foram considerados *ter de* e *dever* pelas suas relações com *poder*.

VERBO	espanhol	francês	alemão	sueco	chinês	total
ser	8	8	25	29	13	83
ter	1	0	7	7	21	36
estar	3	5	9	14	3	34
poder	1	7	2	12	8	30
fazer	2	3	5	13	7	30
ficar	2	2	5	4	9	22
gostar	0	1	7	5	5	18
querer	1	1	4	5	4	15
ir	0	0	3	3	8	14
ver	0	2	2	5	2	11
dizer	0	1	1	2	5	9
ter de	0	0	4	4	1	9
saber	0	0	2	1	5	8
haver	0	2	1	0	3	6
dever	0	0	4	1	0	5
dar	0	0	0	2	2	4
vir	0	0	1	1	0	2
(TOTAL)	18	32	82	108	96	336
% do total absoluto	38%	52%	59%	56%	55%	

Como se pode verificar, **os verbos de alta frequência são responsáveis, excepto no caso dos espanhóis, por mais de 50 % do total de desvios na selecção do tempo verbal adequado ao contexto.** Perante resultados anteriormente comentados em selecção dos verbos aspectuais, as 25 e as 28 ocorrências desviantes, respectivamente em alemães e suecos, no que respeita o verbo *ser*, bem como as 9 e as 14 dos mesmos grupos no verbo *estar*, não são de estranhar, sobretudo se tivermos em consideração que uma boa parte é relativa à oposição Pretérito Perfeito *vs.* Pretérito Imperfeito.

O que pode parecer estranho, no grupo dos chineses, é o elevado número de ocorrências desviantes que o estativo *ter* (38 % do total de 96) envolve e que se repartem pelos três tempos mais frequentes do Indicativo. No entanto, se tivermos em conta que o chinês não privilegia a marcação do Tempo e que o verbo *ter*, estativo com valor de posse, como se pode ver no quadro de Viberg, *Basic Verbs in European Languages*, (ver 6.3.3 *Desvios na selecção do léxico*, 6.3.3.3 *Verbos*) é específico de um número restrito de línguas da Europa, são mais uma vez propriedades lexicais a explicar as ocorrências desviantes, neste caso, na morfossintaxe. De facto, tal como acontece em Islandês, Finlandês ou Húngaro, em Chinês e em outras línguas asiáticas, o conceito de posse é expresso por um predicado

existencial, o que, necessariamente, também tem consequências, ao nível da aquisição, na reestruturação de todo o sistema que, no caso específico do português, envolve os verbos *ser, estar, ficar, ter e haver* ³²³.

Os dados aqui analisados confirmam mais uma vez a hipótese geral de Leiria 1991. Devo referir, no entanto, que nunca foi minha intenção empreender aqui uma análise que respondesse a todas as questões que ficaram sem resposta nessa altura. Foi antes, como disse na Introdução, testar esta hipótese alargando a investigação a outras categorias do léxico. E, de facto, tendo em conta todos os resultados analisados, parece possível terminar afirmando que (1) as propriedades das palavras, (2) a língua materna de cada um e (3) características específicas do Português controlam a aquisição e regem o desenvolvimento da interlíngua em direcção à variedade alvo.

6.4. Conclusões

Anunciei como objectivo central do estudo a avaliação, na modalidade escrita, da dimensão e características do léxico de aprendentes de Português, falantes de cinco diferentes L1 (Espanhol, Francês; Alemão, Sueco; e Chinês); e assumi a Ciência Cognitiva como sendo o enquadramento teórico adequado à investigação. Na *Parte I*, reuni informação centrada no léxico, oriunda sobretudo da Linguística, da Psicolinguística, e dos estudos de aquisição de L2 que me pareceu pertinente para esboçar esse enquadramento teórico.

Propus-me estudar um *corpus* constituído por composições escritas por falantes de cinco línguas maternas diferentes. A estes cinco *sub-corpora*, juntei um outro de controle, produzido por falantes de português língua materna, com características semelhantes quanto aos tipos de textos e às condições de produção. Usei também como material de controle a lista de vocabulário do *PF*.

Em conformidade com as hipóteses de trabalho, o estudo foi organizado em duas grandes partes. Na primeira, 6.2. *Vocabulário disponível*, avaliei a dimensão e a natureza do vocabulário de que dispõe cada um dos grupos; na segunda, 6.3. *Léxico deficitário*, classifiquei e analisei, com diferentes graus de detalhe, os aproximadamente 6 000 desvios, que se manifestavam directamente nas entradas lexicais. Foram ainda classificados os restantes desvios, aproximadamente mais 2 500, anotados no *PESTRA*. Ainda que, como salientei em devido tempo, esta

³²³ Ver Norman 1988:168; ver também Rygaloff 1973:191-210 (Chapitre VIII: La subordination nominale; *yôu* 'y avoir' et *shi* 'c'est').

tipologia e classificação estejam naturalmente sujeitas a reformulações posteriores, em função de investigação a realizar no futuro sobre estes dados, os desvios anotados foram organizados em quatro grandes grupos: 6.3.1. *Desvios formais em vocabulário disponível*, 6.3.2. *Vocabulário indisponível*, e 6.3.3. *Desvios na seleção do léxico*; e, em 6.3.4. *O léxico para além do léxico*, em que foram considerados desvios que se manifestam na sintaxe e na morfossintaxe, procurei mostrar como, em certos casos, eles são sistemáticos e atingem um número restrito de vocábulos e ou estão relacionados com as propriedades de certos itens lexicais.

Assim, em 6.2. *Vocabulário disponível*, procurei avaliar a **dimensão e a natureza do vocabulário** usado no *corpus* pelos aprendentes de PE língua não materna, confrontando-o com o *sub-corpus* de falantes nativos de PE e com o *PF*. Seguindo os critérios de lematização adoptados no *PF*, concluí então que

1. os informantes portugueses usaram aproximadamente 2000 vocábulos diferentes dos quais 50,7 % pertencem ao **PF**, enquanto que os estrangeiros usaram entre 1200 e 1500; no caso destes, a percentagem de **entradas PF** situa-se entre 63,5 %, nos espanhóis, e 74,3 %, nos chineses; quanto à percentagem de **ocorrências PF**, ela é de 87,2 % no caso dos portugueses e situa-se entre 92,6 % no *sub-corpus* francês e 95,2 % no chinês (ver *TABELA 2*); excluindo os vocábulos gramaticais homógrafos de alta frequência, não analisados no *PF*, os 100 mais frequentes vocábulos bastam para perfazer, no *PESTRA*, uma média de 46,7 % do total dessas ocorrências *PF* (ver *TABELA 3*);
2. das 2217 entradas da lista do *PF*, apenas 1454 ocorreram pelo menos uma vez; destas, apenas 371 pertencem ao conjunto de vocábulos da lista originado exclusivamente com base no inquérito de disponibilidade, o que representa menos de 3 % do total de ocorrências no *corpus* (ver *TABELAS 4 e 5*).
3. de entre as combinatórias mais frequentes destacam-se as que codificam relações temporais; pelo contrário, no que

respeita locuções conjuncionais, os materiais do *PESTRA* parecem apontar para que os aprendentes de português L2 enfrentam certas dificuldades na expansão deste tipo de sequências cristalizadas.

Estes resultados mostram que, **(1)** de facto, os diferentes grupos, quando comparados entre si, parecem dispor de um léxico mental cujo número de itens é semelhante; **(2)** a frequência dos vocábulos no *input* é um factor determinante na aquisição e que **(3)** à medida que a distância linguística vai aumentando, os aprendentes vão-se restringindo mais ao vocabulário mais frequente.

Em 6.3. *Léxico deficitário*, como disse antes, foram tratados os aproximadamente 6000 desvios que, nesta fase da investigação, foram considerados como desvios lexicais. Sob o título 6.3.1. *Desvios formais em léxico disponível*, foram classificadas todas as ocorrências com uma configuração não coincidente com o seu modelo em PE. Sob o título 6.3.2. *Vocabulário indisponível e estratégias compensatórias*, foram classificadas todas as ocorrências resultantes de empréstimos de outras línguas, que sofreram ou não processos de relexificação, bem como produtos lexicais criados pelos aprendentes a partir de material linguístico do português e ou de outras línguas. Sob o título 6.3.3. *Desvios na selecção do léxico*, foram classificadas todas as ocorrências resultantes de selecção inadequada de um dado item para um dado contexto. A *TABELA 28* reúne os resultados que foram sendo apresentados e comentados ao longo deste capítulo.

Tabela 28

LÉXICO DEFICITÁRIO	espanhol	francês	alemão	sueco	chinês	totais
DESVIOS FORMAIS						
ortografia	130	92	84	93	36	435
acentuação	372	308	248	247	80	1255
contração	16	9	2	7	1	35
formas aproximadas	122	111	127	92	205	657
flexão de nomes e adjectivos	9	10	7	1	4	31
flexão verbal	50	35	31	54	42	212

atribuição de género	8	25	54	78	26	191
concordância em género	6	0	44	52	26	128
VOCABULÁRIO INDISPONÍVEL						
empréstimos	72	51	30	25	8	186
neologismos a partir de empréstimos	64	52	13	9	1	139
neologismos formais	15	25	26	44	40	150
desvios em combinatórias						
	68	86	106	92	118	470
DESVIOS NA SELECÇÃO DO LÉXICO						
em nomes	30	36	55	55	44	220
em adjectivos	13	10	22	28	15	88
em verbos	60	116	144	172	220	712
em preposições	72	109	204	199	193	777
em outras categorias	41	51	76	97	159	423
recategorização	4	4	21	37	53	120
totais						
	1152	1130	1294	1382	1271	6229

Em 6.3.1. *Desvios formais em vocabulário disponível*, as formas desviantes foram distribuídas por quatro categorias: **(1)** aquelas que resultam de uma não conformidade com as regras de ortografia vigentes para o português europeu, **(2)** aquelas que parecem reflectir uma representação deficiente das estruturas fonético-fonológicas da variedade-alvo (formas aproximadas), **(3)** aquelas que resultam de desvios na flexão, e **(4)** aquelas que resultam de uma atribuição errada de género. Da análise, pode concluir que, no que respeita a

1. erros de ortografia, acentuação e contracções, o número de desvios diminui consideravelmente à medida que a distância linguística vai aumentando, até ser irrisório no *sub-corpus* chinês; enquanto que o grande número de erros no *sub-corpus* espanhol poderá ser explicado pelo recurso a conhecimento implícito fortemente automatizado para a produção de formas semelhantes na sua L1, em contrapartida, o baixíssimo número de erros no *sub-corpus* chinês poderá ser explicado pela transferência de estratégias de memorização visual desenvolvidas para a aquisição da escrita logográfica;

2. formas aproximadas, ou seja, formas escritas mas que reflectem a aquisição imperfeita do sistema fonológico, e apesar do comportamento de cada um dos grupos ser relativamente diferente em função da L1, no geral, os resultados apontam para uma ordem de estabilização do sistema equiparável à do português como L1; para além da distinção vozeada vs. não vozeada, é a estabilização da nasalidade e de sílabas que não seguem a estrutura canónica CV, isto é, aquelas que têm ataque ramificado ou coda, que parecem oferecer mais resistência à aquisição; por outro lado, e isso é revelado pelos dados dos chineses, o uso ou não na escrita de L1 de um sistema alfabético parece ter um forte impacto no processamento lexical e na aquisição do sistema fonológico;

3. desvios na flexão de nomes, adjectivos, e verbos são relativamente poucos em qualquer dos *sub-corpora*, acontecem sobretudo em flexões irregulares, e, de acordo com resultados obtidos para outras línguas, não reflectem os sistemas da L1;

4. desvios na atribuição de género acontecem sobretudo em nomes inanimados e reflectem, por um lado, o apoio que a L1 pode fornecer e, por outro, características da língua alvo e dos aprendentes.

No que respeita desvios formais, podemos então chegar a duas ou três conclusões gerais de naturezas diversas: **1) quanto maior é a distância linguística mais os aprendentes se apoiam em conhecimento explícito e não automatizado; 2) factores intralexicaes combinados potenciam as dificuldades de aquisição;** isso é particularmente evidente para a fonologia no caso de vocábulos em que se combina, por exemplo, nasalidade e sílabas que não seguem a estrutura canónica CV; e **3) os desvios em vocábulos muitíssimo frequentes representam uma boa parte do total.** De facto, desvios formais não só representam, no caso dos espanhóis, mais de metade do total absoluto de desvios, montante que vai

decrecendo até pouco mais de 20 % nos chineses, mas a constância dos desvios em determinadas formas de alta frequência é tal que, por vezes, quatro ou cinco formas bastam para acumular a maior parcela de desvios numa determinada categoria.

Em 6.3.2. *Vocabulário indisponível e estratégias compensatórias*, foram incluídos e analisados os itens lexicais pedidos de empréstimo a outra língua ou criados a partir de material doutras línguas ou do português. Os itens aqui incluídos foram divididos em dois grupos: **(1)** itens simples: empréstimos, neologismos a partir de empréstimos, e neologismos formais; e **(2)** combinatórias aproximadas, que resultam de processos de adição de vocábulos da língua alvo.

Da análise, pode concluir que

- 1.** quanto mais uma língua é próxima da L2 mais ela é a fonte preferida para empréstimos; quanto mais afastadas são duas línguas mais o aprendente descobre a impossibilidade de recurso a esta estratégia de substituição, recorrendo àquilo a que Poulisse (1993) chama de *substituição + estratégia*, ou seja, criação de itens novos construídos com material de línguas próximas do português, nomeadamente do espanhol, do francês e do italiano, e da língua alvo; no caso dos chineses, dada a impossibilidade de recurso à L1 e porque nenhuma outra língua parece estar muito activa, esta estratégia é executada apenas com material do português, ou seja, só produzem neologismos formais e não neologismos a partir de empréstimos;

- 2.** as combinatórias mais sujeitas a desvios são aquelas que, de um ponto de vista paradigmático, têm mais conexões com outras, pela sua proximidade formal, e em relação às quais, devido à sua frequência na língua alvo, é mais difícil usar estratégias de evitação.

Chegados a este ponto, se somarmos os desvios considerados em 6.3.1, *desvios formais em vocábulos disponíveis*, e em 6.3.2, *desvios*

resultantes de indisponibilidade lexical, ou seja, todos aqueles que afectam exclusivamente o item desviante mas não a frase, o traçado da curva em função da distância linguística é muito nítido: os espanhóis totalizam já quase 70 % do total de desvios anotados no *corpus*, enquanto que os chineses, no outro extremo, ainda não chegaram aos 30 % (ver *TABELA 14*). Mas se as relações interlexicais são determinantes e os resultados claros, exactamente o mesmo se pode dizer de factores intralexicais; de facto, **um reduzido número de itens, tanto no que respeita a ortografia e a acentuação como a fonologia ou a indisponibilidade vocabular, nomeadamente de combinatórias frequentes, é responsável por uma boa parte dos desvios, independentemente da L1 dos aprendentes.** Sendo sabido que, no processamento, os falantes prestam atenção preferencialmente ao sentido, uma abordagem centrada nos aspectos formais do léxico, que se concentre no vocabulário de alta frequência, poderá ter efeitos muito positivos na aprendizagem diminuindo drasticamente o número deste tipo de desvios.

Em 6.3.3. *Desvios na selecção do léxico*, foram incluídos 2340 dos aproximadamente 8500 desvios identificados no *PESTRA* (ver *TABELA 15*). Porque, ao contrário do que acontece com desvios formais, eles exigem uma análise muito mais atenta, minuciosa, e diversificada não foi possível realizá-la com a profundidade que merecem. De qualquer modo, distinguiram-se desvios em **(1)** nomes, **(2)** adjectivos, **(3)** verbos, **(4)** preposições, e em **(5)** outras categorias; nesta última incluíram-se, entre outros, advérbios, conectores frásicos, deícticos, e quantificadores; e em **(6)** recategorização, incluíram-se os itens lexicais usados com uma categoria sintáctica diferente da que têm em PE. Da análise dos dados, pode concluir que

1. nomes é a categoria que, relativamente ao total de ocorrências no *corpus*, concentra menor número de desvios; e em que, em contrapartida, os usos desviantes apresentam uma maior dispersão;

2. adjectivos exibem um quadro semelhante: pouquíssimas e muito dispersas ocorrências desviantes; além disso, e isto é

válido também para os nomes, o número de desvios é muito semelhante nos cinco *sub-corpora*;

3. em verbos, bem como em quase todas as outras categorias, pelo contrário, o número de desvios vai aumentando com a distância linguística; os 10 verbos mais frequentes no *PESTRA*, quase todos verbos específicos das línguas da Europa ou das línguas românicas ibéricas, são suficientes para totalizar 50 % dos usos desviantes registados nesta categoria, e 30 % dos 2340 desvios incluídos em selecção do léxico; ou seja, **quanto mais específico de um grupo de línguas ou de uma língua é um verbo mais difícil parece ser a aquisição de todas as suas propriedades semânticas**; por outro lado, fica claro, mais uma vez, que **o maior número de desvios se concentra em itens lexicais muito frequentes**; fica clara também a centralidade do verbo na aquisição e desenvolvimento de uma L2;

4. desvios em preposições confirmam este papel nuclear dos verbos; não só seguem a mesma curva e atingem valores muito semelhantes aos registados em verbos, o que significa também uns 30 % dos 2340 desvios, como são as preposições associadas a verbos mais frequentes aquelas que mais contribuem para este valor; por outro lado, e apesar de análise não ter sido muito minuciosa, parece possível afirmar que a preposição *a*, a mais débil do ponto de vista semântico, é aquela que mais dificuldades coloca à aquisição, apresentando em 40 % dos casos em que está envolvida uma qualquer forma de desvio: omissão, inserção ou uso desviante; seguem-se *lhe por* com 35 % e *para* com 22 %; a meio da tabela ficam *sobre*, *de*, *com* e *em*, com valores entre 15 e 10 %; fecham a lista *até* com 3 %, e *sem* e *desde* com 0 %;

5. em outras categorias, foram incluídos itens lexicais, tradicionalmente incluídos na categoria ADV, mas que

adquirem diferentes valores e comportamentos em função do contexto; foram aqui incluídos também deícticos, quantificadores e conectores frásicos; itens como *então* (generalizado a contextos em que o falante nativo usaria *por isso*), *aqui/ali/cá/lá*, os pares *tão/tanto*, *todo/tudo*, *todo/cada*, e *nenhum/ninguém* são responsáveis pela maior parte dos desvios;

6. em **recategorização** foram incluídos os desvios resultantes do uso de um item de uma determinada categoria sintáctica como se ele pertencesse a uma outra; entre estes destacam-se os pares adjectivo / advérbio *mau/mal* e *bom/bem*; é de acrescentar, no entanto, que a maior parte dos usos desviantes estão associados aos verbos *ser* e *estar* e parecem estar relacionados com deficiente atribuição das suas propriedades aspectuais; a recategorização, dado o traçado nítido que a curva ascendente apresenta (ESP 4; FR 4; AL 21; SU 37; CH 53), revela-se como um bom indicador da organização do léxico.

Na verdade, nesta secção, se exceptuarmos nomes e adjectivos, é claro o traçado ascendente que a curva apresenta; os resultados relativos a nomes e adjectivos confirmam que estas categorias apresentam, em geral, fronteiras semânticas muito mais definidas do que categorias mais polissémicas como os verbos ou preposições; de facto, **a maior parte dos desvios não resulta de indisponibilidade vocabular resolvida com a generalização de um item disponível, mas de modulações nas fronteiras semânticas de uma entrada e das suas conexões com outras que com ela partilham um determinado espaço semântico.** E, mais uma vez, os desvios envolvem vocábulos de alta frequência.

Finalmente, em 6.3.4. *O léxico para além do léxico*, foram comentados certos desvios que, embora se manifestassem na sintaxe ou na morfossintaxe, parecem resultar do estabelecimento deficiente das entradas lexicais envolvidas. E porque foram as propriedades aspectuais dos verbos e a investigação desenvolvida à volta deles, em Leiria 1991, a principal

motivação para este trabalho, prestou-se aqui particular atenção aos desvios na selecção do tempo verbal adequado ao contexto. Concluiu-se, mais uma vez, que, em mais de 50 % dos casos, a selecção de tempos verbais desviantes acontece em verbos de alta frequência, em particular, em verbos específicos de línguas europeias ou ibéricas.

Da análise dos resultados **intragrupos**, ou seja, tendo em consideração a percentagem que cada tipo de desvios representa nos resultados de cada grupo, ficou claro que no que respeita desvios formais em vocábulos disponíveis e desvios resultantes de indisponibilidade lexical, à medida que a distância linguística aumenta este tipo de desvios representa uma percentagem menor do total de desvios, percentagem essa que vai de 70 % no caso dos espanhóis a 28 % no caso dos chineses; isto significa que **à medida que a distância linguística entre a L1 e a L2 vai aumentando os desvios vão afectando mais as relações paradigmáticas e sintagmáticas do léxico, representando então 30 % no caso dos espanhóis e mais de 70 % no caso dos chineses.**

A leitura intergrupos das ocorrências desviantes pode organizar-se em cinco formatos, como se pode ver na *TABELA 29*: **(1)** o número de desvios mantém-se mais ou menos constante independentemente da distância linguística (—) ; **(2)** o número de desvios tem tendência a diminuir à medida que a distância linguística aumenta(+ → -); **(3)** o número de desvios tem tendência a aumentar com a distância linguística (- → +); **(4)** a curva apresenta um traçado irregular(- + -) : o número de desvios, nos grupos germânicos é mais elevado do que nos românicos e nos chineses (- + -); ou **(5)** o número de desvios é mais baixo nos germânicos do que nos outros dois grupos (+ - +). A *TABELA 29* organiza os diferentes tipos de desvios em função destes cinco formatos.

Tabela 29

LÉXICO DEFICITÁRIO: leitura intergrupos					
TIPOS DE DESVIOS	formato 1 —	formato 2 + → -	formato 3 - → +	formato 4 - + -	formato 5 + - +
formais	flexão de nomes e adjectivos	ortografia		género e concordância em género	formas aproximadas

	flexão verbal	acentuação contração			
indisponibilidade vocabular		empréstimos neologismos a partir de empréstimos	neologismos formais combinatórias aproximadas		
Seleção do item lexical	nomes		<i>verbos</i> preposições outras categorias recategorização	adjectivos	
sintaxe e morfossintaxe		colocação dos clíticos	organização sintáctica determinação dos nomes concordância verbal concordância nominal flexão casual tempo verbal OUTROS	ordem de palavras modo verbal aspecto verbal	

Nestes dados de escrita, no que diz respeito a **desvios formais**, em caso algum eles aumentam com a distância linguística (ver formato 2), tendo, pelo contrário, muito mais tendência para diminuir. Particularmente no caso das formas aproximadas (ver formato 5), o facto de estarmos a tratar dados obtidos em material escrito deve ser tido em consideração. De facto, como vimos, estes resultados dos chineses, obtidos em material escrito, mascaram um problema sério de aquisição da componente fonético-fonológica.

Do formato 2 são exemplo os desvios na ortografia e na acentuação das palavras; neste ponto, o grupo espanhol parece sofrer da proximidade entre as duas línguas; franceses, alemães e suecos apresentam resultados semelhantes, enquanto que, nos chineses, eles são muito mais baixos. A proximidade das normas ortográficas parece dificultar a aprendizagem de representações diversificadas das formas de L1 e de L2. Como vimos, no caso dos chineses, esse resultado muito baixo pode dever-se a *transfer* de estratégias adquiridas e usadas para a escrita da L1. Do mesmo modo que os resultados referentes à flexão e à atribuição de género, quando a L1 não oferece apoio transferível para a interlíngua, parecem depender também de outras características do sujeito, nomeadamente de estratégias de memorização; a prová-lo está o facto de os chineses apresentarem resultados inferiores aos germânicos. Pode acontecer, no entanto, que a

possibilidade, oferecida pela situação de escrita, de monitorizarem a sua produção tenha aqui alguns efeitos que favorecem estes resultados.

Também, por outras razões, os resultados relativos a **indisponibilidade vocabular** devem ser lidos com cautela (ver formatos 2 e 3). O facto de os espanhóis apresentarem um maior número de empréstimos não significa que o seu vocabulário seja mais pobre (ver dimensão do vocabulário); significa sim que a proximidade entre as duas línguas favorece a passagem para a interlíngua de formas muito automatizadas da L1.

Mas é na selecção do léxico disponível que a relação tipológica se faz sentir de forma consistente (ver formato 3). Isso só não se verifica no caso dos nomes (ver formato 1) e dos adjectivos (ver formato 4); no primeiro caso porque, como é sabido, nestes estádios de desenvolvimento, os nomes que usam, porque fazem parte do vocabulário básico, apresentam regra geral fronteiras semânticas relativamente definidas (mas ver *pessoa* e *gente*); no segundo caso, talvez porque os chineses restringem mais o uso a adjectivos mais frequentes do que os germânicos. Mas verifica-se de forma particularmente consistente no caso dos verbos, preposições, advérbios e quantificadores (ver formato 3). **Parece, portanto, conveniente distinguir a aquisição do significado prototípico dos itens lexicais, que depende da memória e de conhecimento explícito, da aquisição de outros aspectos semânticos; ou seja, de padrões de lexicalização característicos da L2, das relações paradigmáticas e das extensões metafóricas dos itens lexicais, das suas fronteiras semânticas e das escolhas feitas pela variedade alvo para seleccionar um determinado item para um determinado contexto.** Estes aspectos estão certamente relacionados com conhecimento implícito e com o apoio que a L1 pode fornecer.

No caso da sintaxe e morfossintaxe, também são poucos os casos que fogem à regra do aumento de desvios em função da distância linguística; a diminuição do número de desvios na colocação dos clíticos, à medida que aumenta a distância linguística (ver formato 2), deve-se ao facto de os românicos os usarem mais do que os germânicos e estes, por sua vez, mais do que os chineses. Quanto ao maior número de desvios nos germânicos do que nos chineses em ordem de palavras e modo verbal (ver

formato 4) não tenho, nesta fase da investigação, e porque encontrá-la não era seu objectivo, uma explicação para isso; o mesmo não digo do aspecto verbal que, como vimos, reflecte sistemas interiorizados para a L1.

De facto, a aquisição e uso de aspectos formais, tais como a flexão de nomes, adjectivos e verbos, já que o número de desvios se mantém estável independentemente da L1 dos aprendentes (ver formato 1), poderão apoiar-se mais em características do sujeito e em conhecimento explícito, mais ou menos automatizado, do que a componente semântica e seus reflexos ao nível da gramática; também a atribuição de género a nomes de seres inanimados e certos casos da determinação dos nomes (nomes próprios precedidos ou não de artigo) poderão depender mais de atenção, memória e de estratégias e características do sujeito do que de factores interlexicais.

O **papel da L1 e o conhecimento implícito** muito automatizado, transferível para a interlíngua, faz-se sentir tanto ao nível da aquisição como do uso. Os efeitos da proximidade entre as duas línguas ficam patentes no grande número de desvios, em ortografia e acentuação, empréstimos e neologismos a partir de empréstimos, mas também em formas aproximadas, presentes nos dados dos espanhóis; mas fica patente também na transferência das estratégias adquiridas a quando da escrita do chinês para a ortografia do português. Mas é ao nível da componente semântica dos itens lexicais, das relações paradigmáticas e sintagmáticas que estabelecem com outros, enquanto formas que lexicalizam determinados conceitos, que a relação tipológica entre a L1 e a L2 e as hierarquias de marcação da L2 se fazem sentir na interlíngua. Assim, tendo em conta que adquirir uma L2 consiste em projectar, um no outro, dois sistemas lexicais e conceptuais, **as estratégias de reconceptualização, que implicam a substituição, adição ou apagamento de chunks**

inteiros, afectam tanto mais a estrutura do enunciado quanto maior é a distância linguística entre as duas línguas. A curva ascendente dos desvios (ver formato 3), em verbos, em preposições, e em outras categorias, nomeadamente quando codificam a quantificação ou a genericidade, e também no que respeita recategorização e combinatórias, é muito clara. Mas, os resultados referentes aos tempos verbais que codificam a distinção aspectual acabado/não acabado, em especial nos dados dos germânicos e dos chineses (ver formato 4), lembram-nos que dois sistemas linguísticos mais afastados (ver português e chinês) podem, como prevê a linguística tipológica, em certos casos, apresentar mais pontos de contacto do que dois sistemas mais próximos (ver português e línguas germânicas); no entanto, eles também nos lembram, mais uma vez, que as propriedades dos itens lexicais e a marcação (ver resultados dos chineses nos verbos *ser* e *estar* e, muito em especial, em *ter*) desempenham um papel na aprendizagem. **E porque os itens lexicais mais frequentes são os mais dessemantizados, muito em especial os verbos, e, em muitos casos, os hierarquicamente mais marcados, e porque factores intralexicais combinados potenciam a dificuldade de aquisição, é neles que se concentra uma boa parte dos desvios.**

Apesar de esta abordagem centrada no léxico nos ter conduzido a uma análise dos dados que nos permitiu chegar a um número razoável de conclusões quanto **ao papel dos conhecimentos prévios e quanto ao papel das características específicas do PE na aprendizagem**, muito ficou por fazer e outras questões se põem. Para além de estudos mais detalhados sobre estes dados obtidos em documentos escritos, impõe-se uma recolha e análise de dados do oral. Essa análise permitiria avaliar, por comparação, qual a contribuição do conhecimento explícito na situação de escrita, já que, contrariamente àquilo que acontece na produção *on-line*, o tempo de reflexão proporciona a monitorização da produção. Para conclusões quanto **ao papel do ensino** seria conveniente comparar estes dados, recolhidos em situação de aprendizagem semi-formal, por um lado, com dados recolhidos em situação de aprendizagem exclusivamente informal e, por outro, com dados recolhidos exclusivamente em situação formal (língua estrangeira), ou seja, em que se verificasse a ausência de variável imersão.

Por outro lado, fica também por fazer a análise dos desvios por cada um dos documentos deste *corpus*; ou seja, observar que tipos de desvios coincidem no mesmo documento e relacioná-los com os dados não desviantes. Embora nunca esquecendo que estamos a tratar com dados da modalidade escrita, essa análise permitiria verificar quais as componentes do léxico que parecem seguir uma determinada ordem de aquisição, ensaiar escalas implicacionais, e pôr hipóteses quanto à sua relação com eventuais estádios de aquisição do PE enquanto L2.

Capítulo 7

Léxico e ensino de PE L2

7.1. Introdução

Language is so incredibly complex that there is no obvious answer to who to teach it. (Schmitt & McCarthy 1997:322)

Na *Introdução* a esta dissertação, salientei que a Linguística Aplicada deve considerar a investigação realizada em aquisição da linguagem e chamei a atenção para a urgência de os agentes de ensino de PE L2 (professores, planificadores de cursos, autores de materiais didácticos, formadores e avaliadores) deixarem de trabalhar exclusivamente com base na intuição e nas convicções a que a experiência os conduziu. Referi o facto de haver uma crise na sala de aula quando “se acaba” a morfossintaxe; quando o aluno tem quase tudo para aprender e o professor parece já não ter nada para ensinar. Esta crise resulta de, actualmente, apesar da ênfase posta nos aspectos comunicativos da linguagem, o ensino continuar a ser orientado pela sequencialização de tópicos gramaticais. Isso deve-se, certamente, ao facto de a abordagem nocional-funcional conceber a linguagem como um conjunto de actos de fala e de não oferecer um enquadramento linguístico que permita hierarquizá-los. Salientei ainda as vantagens que pode ter a investigação centrada no léxico, em particular numa paisagem de que os estudos de aquisição estão consideravelmente ausentes. E ainda que, como referi em 6.4 *Conclusões* do estudo, seja necessário recolher e analisar muitos mais dados de PE L2 (do oral, de PE enquanto língua estrangeira, de PE em contexto de aprendizagem informal, etc...), a bibliografia disponível para outras línguas e os resultados desta investigação, embora devam ser tratados com cuidado, podem sugerir alguns caminhos para o ensino.

Quero, no entanto, chamar a atenção para que o objectivo deste capítulo não é fazer propostas para actividades ou uma aplicação caso a caso das conclusões do estudo. Isso, como bem se compreende, não cabe no âmbito desta obra. Esse seria um trabalho de aplicação que teria de ter um

determinado público alvo previamente bem caracterizado. **Neste capítulo, o que pretendo é propor, para o ensino, um enquadramento coerente e fundamentado que decorre de uma abordagem lexical *lato sensu*.** A sua aplicação, uma vez seleccionado o público alvo, será tarefa dos agentes de ensino ³²⁴. **O que farei aqui será uma releitura de algumas das conclusões e das suas implicações para o ensino do PE L2.** No entanto, considero que, se acaso este trabalho pode trazer alguma contribuição para o ensino do PE L2, é sobretudo no Capítulo 6 que ela se encontra. É na análise das diferentes componentes de cada uma das palavras e na constatação de quais, de facto, põem mais problemas à aquisição que me parece residir esse contributo. Para além das transcrições de ocorrências presentes nos documentos, dos quadros com os resultados da análise dos dados, e dos comentários, também os *ANEXOS* podem fornecer informação não negligenciável na hora de reflectir sobre que palavras seleccionar, que combinatórias privilegiar, e, não menos importante, que comportamento esperar dos aprendentes perante cada uma delas.

Acrescentarei ao capítulo algumas sugestões, decorrentes dos resultados do estudo, quanto à utilização, na perspectiva de uma abordagem centrada no léxico, de materiais de apoio para professores de PE L2, nomeadamente o *Português Fundamental* e o *Nível Limiar*, e, motivada pela gritante falta de publicações nesta área, permito-me extrapolar um pouco e fazer algumas reflexões quanto à necessidade de elaboração de um corpo de textos teóricos sobre aprendizagem e ensino de PE L2, sobre o contributo que investigação em linguística teórica e aplicada pode dar, e sobre a urgência de formação de professores / investigadores nesta área.

7.2. *Uma abordagem cognitiva centrada no léxico*

A análise de materiais de PE L2 centrada no léxico permitiu-nos um tratamento bastante abrangente dos dados de aquisição. Isto sugere que, também para o ensino, uma abordagem centrada no léxico poderá ser adequada e vantajosa.

³²⁴ Modelos para actividades centradas no léxico podem encontrar-se, por exemplo, em Gairns & Redman 1986 e em Lewis 1997a e b.

Tanto a abordagem lexical como a cognitiva consideram o léxico como central na aprendizagem. A primeira sugere que o ensino se deve apoiar na observação de *corpora* de língua alvo (ver Sinclair & Renouf 1987), a segunda considera o léxico como central no processamento e na aquisição e fornece-nos enquadramento para explicarmos fenómenos relacionados com a aprendibilidade dos diferentes itens. Assim, **a abordagem lexical pode apoiar o ensino guiando a selecção do vocabulário e das suas combinatórias mais frequentes, e, como salientam os seus proponentes, se a selecção for bem feita, cobrirá a gramática relevante para cada nível de ensino** ³²⁵. **A abordagem cognitiva fornece-nos instrumentos para controlar essa selecção porque nos permite formular hipóteses quanto às probabilidades de uma dada componente ser ou não adquirida num determinado estágio de desenvolvimento, o que pressupõe, como foi feito neste trabalho, a consideração de dados de aquisição de L2. A combinação destas duas abordagens pode constituir, portanto, um enquadramento adequado e coerente para o ensino.**

Esse enquadramento aconselha a consideração, e em alguns casos a releitura, de certos pares de conceitos muito difundidos na literatura disponível para os agentes de ensino. Embora fosse possível acrescentar outros, refiro-me, por exemplo, a pares directamente relacionados com a abordagem lexical (*vocabulário e léxico, vocabulário nuclear e vocabulário periférico, vocabulário e gramática, etc.*) e a outros que têm merecido a atenção dos investigadores (*recepção e produção, competência e performance, nocional e funcional, etc...*). Será em função deles que organizarei a exposição que se segue.

7.2.1. *Alguns pares de conceitos a ter em consideração no ensino*

7.2.1.1. *Vocabulário e léxico*

Como vimos, saber uma palavra é muito mais do que a uma dada forma atribuir um significado. Saber um significado de um dado item lexical constitui apenas uma reduzida parte do seu conhecimento. Assumir um

³²⁵ A adopção desta abordagem põe uma questão para a qual chamo a atenção, mas que não discutirei aqui: o vocabulário e as suas combinatórias frequentes deve ser apresentado em “materiais autênticos” ou em materiais fabricados para o ensino? Em que circunstâncias deve ser usado um ou outro? Para que fins? (ver, em Capítulo 4, Ellis & Tanaka & Yamazaki 1994 e Ellis 1995).

conceito rico de léxico significa assumir que a aquisição de uma palavra não é monolítica, ou seja, que as várias componentes (fonológica, semântica, sintáctica, pragmática...) poderão não ser adquiridas todas ao mesmo tempo. Significa também que a aquisição de certas componentes pode ser governada por uma sequência de aquisição; ao nível da fonologia, por exemplo, serão certamente mais fáceis palavras constituídas por sílabas que respeitam o esquema CV e serão mais difíceis as que apresentam ataque ramificado ou coda, embora isso também dependa das características da sua L1 ou de outras que conheça; o mesmo é válido ao nível da componente semântica, mas serão certamente mais fáceis os sentidos mais prototípicos de um dado item lexical. Isto não significa, no entanto, que o professor deva privilegiar o critério grau de dificuldade para seleccionar um item em vez de outro. Significa sim, e antes de mais, que da representação mental que o professor tem de uma dada entrada lexical deve fazer parte um tipo de informação suplementar: consciência da natureza e do grau de dificuldade que o aprendente experimentará quando confrontado com ela.

7.2.1.2. Vocabulário nuclear e vocabulário periférico

Os investigadores continuam a discutir se é ou não possível estabelecer um vocabulário nuclear. Carter 1987 acentua que testes linguísticos não são suficientes para estabelecer um vocabulário nuclear. São necessários testes psicolinguísticos. De facto, uma das conclusões, não direi inesperada, mas surpreendente pela sua nitidez, deste estudo consiste em uma boa parte dos desvios se concentrar em um número relativamente restrito de itens de alta frequência.

Segundo Blum & Levenston 1978, os aprendentes evitam palavras que apresentem dificuldades fonético-fonológicas, semânticas ou sintácticas e preferem aquelas que podem ser generalizadas a muitos contextos. É provável que isso aconteça em certos casos. Contudo, parece-me haver uma contradição nesta afirmação. As palavras que podem ser generalizadas são, de facto, aquelas que apresentam mais dificuldades. Mas elas constituem o vocabulário tendencialmente nuclear e são, como pudemos constatar, as que concentram um maior número de desvios. O que acontece é que, em relação a estas, o aprendente não pode usar uma estratégia de evitação

porque elas são absolutamente indispensáveis. **Este resultado alerta-nos para a necessidade de não abandonarmos precocemente, ao nível da produção, o léxico básico.** E, se os resultados são tão reveladores na modalidade escrita, que permite a utilização de conhecimento explícito na monitorização, o mesmo acontecerá, certamente até de modo mais acentuado, na produção oral.

Mas **os desvios concentram-se mais em certas categorias do que em outras.** As palavras mais difíceis não são muitas, mas exigem muita atenção. Tudo se organiza à volta dos verbos básicos. Viberg 1993a põe como hipótese que, **quando os verbos básicos e as suas construções estão adquiridas, o aprendente dispõe de modelos para a produção de todas as construções.** Conectores e combinatórias frequentes, sobretudo as que codificam relações temporais, também fazem parte deste vocabulário prioritário. No entanto, os resultados parecem permitir-nos um certo alargamento das categorias nome e adjectivo. Mas, palavras novas, na medida do possível, devem vir acompanhadas de outras com que co-ocorrem frequentemente.

Embora, como salientei, seja discutível e difícil de estabelecer a fronteira entre vocabulário nuclear e vocabulário periférico, esta distinção é operacional e oferece-nos um enquadramento não só para a programação de cursos para públicos indiferenciados mas também para públicos específicos. Neste caso, o vocabulário específico, tendencialmente periférico, deve ir sendo introduzido na rede constituída pelo vocabulário nuclear (ver Capítulo 1, referências a Viberg e a domínios e sub-domínios lexicais, propostos por Martín Mingorance).

7.2.1.3. *Vocabulário e gramática*

Uma vez que, na mente de um falante, muitas das entradas lexicais são constituídas por *chunks* de diferentes tipos, tais como sequências cristalizadas, cabeças de frase, fórmulas, etc. (ver 6.2.2 *Combinatórias*), e que as gramáticas são sistemas de regras que operam sobre uma base de dados lexical (Ellis 1996), os proponentes da abordagem lexical defendem que a dicotomia vocabulário vs. gramática não é válida (Lewis 1993:vi), e que, portanto, ensinar léxico é ensinar gramática.

Sinclair e Renouf 1987 fornecem-nos a metodologia para ensinar gramática ensinando léxico: se a escolha das combinatórias, a partir de *corpora*, for feita de forma correcta, elas cobrirão a gramática básica (ver também Viberg 1993a). Por outro lado, e isto é especialmente importante no ensino de línguas como o português em que a morfossintaxe costuma ocupar a maior parte dos tempos lectivos, esta metodologia de selecção ajudará os agentes de ensino a seleccionar aquela que é, de facto, apropriado destacar em cada estágio de desenvolvimento linguístico. Este procedimento evitará, como diz Pienemann (1989:72) “storing up treasures in heaven”; ou seja, a acumulação de conhecimentos que o aprendente só estará em condições de usar muito mais tarde ou, talvez, nunca.

7.2.1.4. *Aprendizagem implícita e aprendizagem explícita*

Como vimos, no âmbito da Linguística Aplicada, as posições têm ido desde aqueles que defendem que a aprendizagem do vocabulário é completamente implícita até aqueles que defendem que vale a pena ensinar/aprender explicitamente longas listas de vocabulário. Os primeiros afirmam que, uma vez que, no processamento, a atenção está concentrada na mensagem e não na forma, o sujeito não pode controlar aquilo que está a aprender; os últimos recomendam que palavras é que devem ser ensinadas e quais os significados destas que devem ser fornecidos em primeiro lugar. Mas, mais recentemente, «the pendulum has swung from direct teaching of vocabulary (the grammar translation method) to incidental (the communicative approach) and now, laudably, back to the middle: implicit and explicit learning.» (Sökmen 1997:239). De facto, também **os resultados desta investigação apontam para que diferentes componentes das palavras exigem diferentes estratégias**. Eles apontam claramente para que o significado e os aspectos formais da morfossintaxe podem ser adquiridos com uma certa rapidez. É verdade que o facto de termos analisado aqui dados da escrita pode dar uma impressão enganadora a respeito da morfossintaxe, visto que há tempo para monitorizar, e portanto na produção escrita ela pode constituir conhecimento não automatizado.

Nick Ellis 1996 propõe que *a hipótese forte de aprendizagem explícita* é compatível com a aquisição do significado enquanto que *a hipótese fraca*

de aprendizagem implícita é suficiente para garantir a aquisição dos aspectos estatísticos e as regularidades ortográficas e fonológicas. É sabido que, em geral, o ensino (e também o falante no acto do processamento) não presta muita atenção aos aspectos formais do léxico. No entanto, os resultados desta investigação aconselham um procedimento diferente, e, mais uma vez, o ensino deve ter em consideração os dados da aquisição. Se há componentes das palavras que, não dependendo de sistemas conceptuais interiorizados com a L1, podem ser aprendidos por processos explícitos, se, além disso, e como acreditamos, o ensino pode chamar a atenção para certas regularidades (aspectos das palavras aprendidos por processos implícitos), e, assim, embora não interferindo na ordem de aquisição, acelerar o processo, então parece defensável propor algumas palavras antes de outras, certos contextos de uso antes de outros, e diferentes estratégias para a aprendizagem de diferentes componentes. É óbvio, no entanto, que a sua selecção deve ser guiada por uma série de critérios (frequência, utilidade, motivação...) e, antes de mais, o professor deve estar consciente da lei da aprendibilidade: nem tudo aquilo que se ensina se aprende.

7.2.1.5. *Estratégias de substituição e estratégias de reconceptualização*

Aprender uma língua consiste, fundamentalmente, em aprender propriedades de palavras. Assim, quanto mais próximas são duas línguas mais propriedades dos diferentes itens lexicais que as constituem serão coincidentes. Os aprendentes têm consciência deste facto. Daí que, como vimos, os espanhóis actuem sobretudo ao nível do formulador, usando estratégias de *substituição*; os franceses, conscientes de que a morfossintaxe não se pode transferir, preferam *substituição + estratégia*; enquanto que germânicos e chineses, conscientes de que muitas palavras da sua L1 têm propriedades distintas da L2, actuem mais ao nível do conceptualizador, preferindo estratégias de *reconceptualização*.

Em muitos contextos de ensino, a maior parte do tempo de que dispomos na sala de aula é ocupado com actividades que treinam a morfossintaxe, ou seja, o que se passa no formulador. Esta prática é útil para todos. No entanto, e em particular, no caso de línguas mais afastadas, ela não actua, directamente, sobre os aspectos semântico-conceptuais das

palavras e sobre as suas consequências a nível sintagmático, já que deixa de fora, por exemplo, a lexicalização de conceitos, a fusão de significados e os sentidos mais periféricos dos itens lexicais. **Pondo-se mais a tónica no léxico, na fusão de significados e nas combinatórias, procurando um equilíbrio entre estabilidade e criatividade no uso da linguagem, poder-se-á ajudar mais os alunos a pensar para falar.** Ao mesmo tempo, criar-se-á uma fonte de material para ensino que resolva o problema que constitui o professor ter a sensação que não tem nada para ensinar e o aluno a sensação de que tem tudo para aprender.

De qualquer modo, os resultados deste estudo mostram claramente que os aprendentes partem para a aprendizagem da L2 de estádios de aquisição muito diferentes. E isso tem forçosamente de ser tido em consideração no tipo de actividades seleccionadas pelo ensino. Tarefas de reconhecimento (*noticing*) serão úteis para todos os grupos, mas elas terão certamente efeitos bastante visíveis no caso dos falantes de línguas do mesmo grupo, neste caso, de **línguas românicas**. Exceptuando a componente fonético-fonológica, porque constitui conhecimento implícito muito automatizado e, por conseguinte, muito difícil de controlar na produção oral, no caso dos aprendentes deste grupo, em especial no caso dos espanhóis, as outras componentes formais, que, como vimos, são a maior fonte de desvio, ganharão bastante com actividades de reconhecimento contrastivo. Essas tarefas de reconhecimento devem ter como objectivo prioritário acentuar tudo aquilo que há de comum entre as duas línguas (ver Leiria 1998); uma vez conscientes das semelhanças, as tarefas devem assinalar as diferenças. Para este efeito, o *EuRom4, Método de ensino simultâneo de línguas românicas*, constitui um óptimo guia ³²⁶. Este tipo de actividades cobrirão uma boa parte tanto da morfologia

³²⁶ O **EuRom4** «visa uma certa forma de conhecimento ‘modesto’ das línguas». «Na perspectiva adoptada, ‘compreender’ não significa dominar totalmente tudo aquilo que se diz ou escreve nessas línguas.» «O método é o resultado de experiências levadas a cabo, durante três anos, nos quatro países envolvidos – Portugal, Espanha, Itália e França. O trabalho experimental veio confirmar a convicção que partilhávamos desde o início: quando as condições são favoráveis, aprender a compreender uma, duas ou três outras línguas românicas é tarefa fácil e rapidamente conseguida.» Porque durante o trabalho experimental os participantes «mostraram tendência para fazer comparações entre as línguas», o método fornece «algumas bases elementares de comparação entre as línguas românicas em causa, seleccionadas a partir das observações registadas.» (*EuRom 4*, 1997:3-5). A equipa portuguesa, coordenada por Maria Antónia Mota, era composta por Amália Mendes, Isabel Leiria e León Acosta.

flexional como da morfologia derivacional e também do vocabulário de alta frequência. Como vimos, os falantes de línguas próximas não se mostram muito confiantes quando se trata de transferir para a interlíngua sentidos menos prototípicos; por isso, haverá enorme vantagem em fazer notar combinatórias frequentes. Estas actividades de reconhecimento actuarão assim não só ao nível da gramática mas também do discurso.

No caso dos outros grupos, nomeadamente dos **germânicos** e dos **chineses**, como nos mostram os dados de aquisição, muitos dos desvios resultam de diferenças na lexicalização de conceitos, por vezes na fusão de significados, atingem as propriedades das palavras, reflectindo-se nas construções. Trata-se, portanto, de adquirir novos padrões de lexicalização (ver, p. ex., verbos de movimento) ou modificar perspectivas associadas à L1 (ver aspecto verbal). Porque estamos a lidar com conhecimento implícito associado à L1, será preciso muito esforço, muito tempo e muito *input* adequado. E porque nem sempre o professor poderá ser falante das línguas maternas dos seus alunos (e mesmo que o seja...), para que os conteúdos das actividades a realizar na sala de aula sejam criteriosa e adequadamente seleccionadas em função das necessidades dos aprendentes, é indispensável obter informação em textos de linguística descritiva sobre as línguas envolvidas.

7.2.1.6. *Cognatos e não cognatos*

Meara 1993 lembra-nos que parece haver quatro padrões de relações de cognaticidade entre as línguas. O mais simples, e provavelmente o mais frequente, pode ser ilustrado pelo caso do português e do chinês: as duas línguas partilham muito poucos ou quase nenhuns cognatos. Outro caso, também ele simples, é o das línguas românicas: uma boa parte do vocabulário é constituído por cognatos, ainda que, muitas vezes, nomeadamente por razões fonológicas, na prática, esse facto não seja muito evidente; isto é particularmente verdade no caso do português e do francês. O português e as línguas germânicas são exemplo de um outro tipo de relação: há um certo número de cognatos, mas que, no entanto, têm frequência muito diferente porque se restringem a certos registos ou a certos centros de interesse.

Cada um destes tipos ocasiona dificuldades de aprendizagem diferentes e exige diferentes estratégias de ensino. No entanto, apesar de existir um grande número de cognatos, em especial entre as maiores línguas da Europa, o ensino não lhes tem prestado atenção e não os tem explorado adequadamente. Isso acontece, acrescenta Meara, não só porque muitas vezes os grupos são multilíngues mas também devido aos pressupostos que têm orientado a investigação (Meara 1993:281).

Na verdade, o facto de uma palavra da L2 nos lembrar uma palavra da L1 – seja por razões ortográficas ou fonéticas – facilita a sua aprendizagem. Contudo, porque a relação de cognaticidade nem sempre é transparente (ver Kirsner *et alii* 1993), os aprendentes muitas vezes não relacionam cognatos óbvios (Garcia 1991; Hancin-Bhatt & Nagy 1994). Daí que se justifique que o ensino se ocupe deles de modo a ultrapassar essa dificuldade de identificação. Além disso, e como vimos a propósito das línguas românicas e germânicas, também é necessário consciencializar os aprendentes da possibilidade de certos vocábulos terem um uso mais restrito na L2.

Convém acrescentar, no entanto, que tem sido prestada muito mais atenção aos perigos da eventualidade de dois vocábulos formalmente semelhantes não terem o mesmo significado ou a mesma frequência de uso na L1 e na L2 do que ao tesouro que eles constituem para o processamento e para a aprendizagem os cognatos. Na verdade, falsos cognatos – os chamados **falsos amigos** – ocupam frequentemente um lugar de destaque no ensino, criando, por vezes, nos aprendentes, sobretudo de línguas próximas, como o português e o espanhol, uma sensação de grande insegurança que pode conduzir a estratégias de evitação. No entanto, falsos amigos não são aqueles que o trazem escrito na cara, mas sim os que disfarçam muito bem. Ou seja, falsos amigos perigosos não são aqueles que, uma vez encontrados e identificados como tal, nos deixam de pé atrás. É exemplo clássico o adjectivo ‘espantoso’ em português e em espanhol. Também não são particularmente perigosos vocábulos que têm uma frequência de uso muito diferente nas duas línguas, como ‘comida / comida’ e ‘refeição / refección’. Também estes, uma vez encontrados, ficam marcados com um sinal de alerta. O que não significa que, por razões que se prendem com automatização, não voltem a ser usados. Verdadeiramente

perigosos são aqueles que nos deixam criar confiança, mas que, de vez em quando, nos traem, porque a sua distribuição dentro de um mesmo campo semântico é diferente; é o caso de pares como ‘senha / seña’ e ‘bilhete / billete’. Mas particularmente traiçoeiros, como vimos pelos resultados deste estudo, são certos vocábulos de alta frequência, de que são exemplo verbos dessemantizados como ‘dar’ e ‘fazer’ ³²⁷.

Mais uma vez os resultados do estudo nos aconselham o treino de combinatórias frequentes e a observação dos dados fornecidos por *corpora* de L2. Por outro lado, alertam-nos para a urgência de pormos à disposição do público interessado dicionários de combinatórias do Português (ver Nascimento & Pereira 1995), nomeadamente dicionários escolares que tenham em consideração *corpora* de L1 e de L2 (ver Biderman 1995) ³²⁸.

7.2.2. Releitura de alguns outros pares de conceitos

7.2.2.1. Recepção e produção

Verificámos que os aprendentes, tal como os falantes nativos, para a produção de textos que envolvem um vasto leque de temáticas, usam um conjunto de vocábulos relativamente pequeno. Verificámos também que é exactamente nesse conjunto restrito que se concentra a maior parte dos desvios. Mas é certo que, particularmente se se encontram em situação de imersão, o vocabulário de que precisarão para a compreensão será muito mais vasto. É indispensável, portanto, que o ensino e os materiais que a ele se destinam respondam adequadamente a esta realidade. Ou seja, **qualquer actividade didáctica deve estabelecer claramente o seu objectivo: ou ela se destina a treinar a descodificação do discurso ou ela se destina a melhorar a produção. Se se destina a melhorar a produção, em particular nos primeiros níveis, o ensino deve ser conduzido de modo a**

³²⁷ Apesar de se falar muito de falsos amigos, a qualidade muito discutível de certos dicionários existentes no mercado revela a pouca atenção que este par de línguas tem merecido por parte dos especialistas. Veja-se, por exemplo, a 2ª edição do dicionário de Espanhol-Português da Porto Editora. De entre as palavras assinaladas como empréstimos no *sub-corpus* espanhol, confira neste dicionário, por exemplo, as entradas *contestar*, *evolucionar*, *impulsar* e *influir*. Em qualquer dos casos, elas são apresentadas como primeiro sinónimo em português.

³²⁸ O *Dicionário do Português Básico* (Vilela 1990) tem em consideração o vocabulário do PF e oferece alguns contextos frequentes. O *Dicionário da Língua Portuguesa Contemporânea* (Casteleiro dir. 2001) usou *corpora* de L1 e, neste sentido, pode ser um modelo para a selecção de contextos.

concentrar-se no vocabulário básico e na sua recombinação em sequências de alta frequência. Essas combinatórias devem ser treinadas até ser atingido um alto grau de automatização. Embora, como nos mostraram os dados de aprendentes de línguas mais afastadas do português, em particular os dos chineses, se deva fazer também uma distinção clara entre **escrito e oral**, o contacto repetido com essas sequências frequentes terá certamente efeitos também ao nível da descodificação do oral. Ajudará o ouvinte a identificar e a descodificar *chunks* de alta frequência, que funcionarão como uma rede em que se poderá apoiar, ficando a atenção disponível para o processamento de material linguístico novo e menos estável ³²⁹. O facto de os aprendentes, em particular falantes de línguas muito afastadas do português, precisarem de *input* abundante, portanto de um certo tempo, para estabelecerem uma rede mínima (uma gramática básica) que lhes permita produzir frases ou enunciados, aconselha a que, na iniciação, seja observado um período mais ou menos longo de treino da compreensão oral e escrita sem produção. A metodologia escolhida para esse treino deverá ter em consideração a relação entre a L1 e a L2 (ver 7.2.1.5. *Estratégias de substituição e de reconceptualização*).

No que respeita a aquisição de conhecimento receptivo através de material escrito, é necessário, como salientámos antes, distinguir estratégias de inferência de estratégias de retenção. A leitura de textos mais ou menos extensos, com uma percentagem controlada de vocabulário novo, deve ser incentivada e guiada. Estratégias de inferência devem ter em consideração sobretudo as deixas que o contexto e a morfologia podem disponibilizar (ver Capítulo 4 e, entre outros, Nattinger 1988). Vimos que com menos de 10.000 lemas diferentes será muito difícil atingir um nível de compreensão satisfatório de textos académicos. **Este número deve ser tido em consideração no planeamento curricular de cursos para estrangeiros, já que estes incluem, frequentemente, em níveis pouco mais do que básicos, para além das aulas de língua, disciplinas muito diversas, tais como literatura, história ou geografia.**

³²⁹ Nóia (1991:404), a propósito do ensino dos aspectos fonológicos do português a chineses, afirma: «ter-se-á de sair do nível da palavra e atingir sequências mais longas.»

7.2.2.2. *Competência e performance*

A *abordagem cognitiva* que, como vimos, absorve as mais recentes posições sobre a relação entre o léxico e a sintaxe, e que pretende integrar uma mais forte componente psicolinguística, não distingue competência de performance, mas distingue dois modos de aprendizagem e de uso. Skehan (1998:35) defende que os falantes dispõem de dois sistemas a que recorrem para a produção de discurso: um apoiado em regras (*'rule-based system'*) e outro apoiado em *chunks* prefabricados (*'exemplar-based system'*). E cita Pawley & Syder (1983:209): «What makes an expression a lexical item, what makes it part of the speech community's common dictionary, is, firstly that the meaning of the expression is not (totally) predictable from its form, second, that it behaves as a minimal unit for certain syntactic purposes, and third, that it is a social institution». E alerta para que «learners who restrict themselves to item-and-rule approaches to language will be for ever marked as non-members of the speech community they aspire to. To achieve that acceptability, to satisfy Hymes's requirement of using language which is 'done', they will have to become more lexical in their mode of communication, and correspondingly in the repertoire of language knowledge that they possess» (Skehan 1998:39). Ou seja, **uma abordagem que dê prioridade ao léxico pode resolver o problema da falta de adequação, a que conduz uma abordagem que se concentra demasiado em tópicos gramaticais, e o problema da dificuldade de gerar enunciados novos, a que conduz uma abordagem excessivamente comunicativa.**

7.2.2.3. *Fluência e correcção*

Constatámos que à medida que a distância linguística aumenta os desvios formais diminuem, enquanto que desvios relacionados com propriedades dos itens lexicais aumentam. Este facto, embora, em parte, também possa depender do estilo pessoal do aprendente, está certamente relacionado com a quantidade de conhecimento implícito que é possível transferir da L1 para a interlíngua, e com o binómio estabilidade e criatividade no uso do vocabulário, e reflecte-se no balanço entre fluência e

correção ³³⁰. Se tivermos em consideração que fluência depende do uso automatizado de *chunks* mais ou menos longos, é claro que a fluência será tanto maior quanto mais conhecimento implícito for possível transferir da L1 para a interlíngua; por outro lado, é preciso ter presente que «quanto menos um aprendente tem de aprender, para ser capaz de comunicar, mais rapidamente a sua proficiência tende a fossilizar» já que ele «não tem motivação integrativa suficiente» (Ringbom 1990:173). É natural, portanto, que fluência e correção apareçam como duas forças antagônicas. A questão está em como acomodar, ao nível do ensino (e da avaliação), o equilíbrio necessário (e, mais uma vez, tendo em conta a modalidade), em função do nível de proficiência do aluno, da relação L1 – L2, do tipo de desvio, e da perturbação que ele causa no sucesso da comunicação. **A manutenção, por parte do professor, de um equilíbrio justo e desejável depende indiscutivelmente da sua formação. Para o conseguir, o professor tem de dispor de, pelo menos, uma certa consciência do estágio de desenvolvimento do(s) aprendente(s). Para isso, a formação em aquisição de L2 é indispensável.** Ou seja, o professor tem de saber, antes de mais, se a falta de correção em causa depende de aprendizagem implícita ou de aprendizagem explícita, se é uma questão mais ligada à competência ou à performance, e se, em função do estágio de desenvolvimento, valerá ou não a pena corrigir. Por exemplo, em texto escrito, vale a pena corrigir todos os desvios de concordância. Mas valerá a pena corrigir, nos primeiros níveis, todos os desvios de idiomaticidade, sobretudo quando eles não são em sequências muito estáveis e frequentes? E, no oral, valerá a pena interromper o estudante para corrigir todos os desvios de concordância, deixando-o com a sensação de que comete muitos erros e, conseqüentemente, de que, por enquanto, seria melhor estar calado?

7.2.2.4. *Categorias nocionais e categorias funcionais*

³³⁰ «Stability and creativity are complementary and interactive factors in vocabulary use. [...] The task therefore is to create a methodology for vocabulary teaching which accommodates this equilibrium, this is a task which we are only now beginning to address» (Cowie 1987:137)

Categorias nocionais envolvem conceitos gerais como tempo, relações temporais, duração, frequência, ou quantidade; categorias funcionais referem-se a objectivos comunicativos tais como fazer ou responder a um pedido, recusar ou aceitar um convite, pedir a palavra ou interromper alguém. A abordagem comunicativa, apesar de ter enfatizado estas duas categorias de natureza semântica não se centrou no léxico. O vocabulário é introduzido para responder às necessidades de apresentação de uma noção mas, principalmente, de uma dada função (ver 4.2. *Das listas de palavras às listas de actos de fala*).

Os resultados da investigação aconselham vivamente, mais uma vez, um ensino centrado nos itens de alta frequência, já que é à volta deles que se organizam muitas das combinatórias mais frequentes. Como pudemos verificar, e porque se tratava de material escrito, a maior parte delas codificam sobretudo categorias nocionais, e os aprendentes mostraram ter dificuldade na sua aquisição. Perante os dados, e perante as tentativas que as sequências riscadas nos documentos analisados revelam, o que se sente é uma incessante busca da forma correcta, dando a entender que os aprendentes têm consciência de que o seu uso desviante é fortemente caracterizador do discurso estrangeiro. Se a expressão da categoria tempo, por exemplo, é lexicalizada diferentemente em diferentes línguas, se envolve tanto o léxico (itens simples e combinatórias) como a morfossintaxe, então, é desejável ter isso em mente no ensino. Ou seja, tratar a categoria tempo na sala de aula é muito mais do que treinar os tempos verbais.

Por outro lado, e no que respeita as categorias funcionais, mais evidentes no oral, a sequencialização no ensino dos actos de fala (sobretudo quando é possível optar entre diferentes possibilidades para uma mesma situação comunicativa) também deve ter em conta o grau de marcação da estrutura envolvida, ou seja, o factor aprendibilidade. De facto, apesar da enorme mais valia que a abordagem comunicativa trouxe para o ensino, creio que um dos principais problemas que ela coloca à aprendizagem, e que a tem posto em causa na última década, reside no facto de muitas das estruturas semantico-gramaticais em que se apoiam as categorias funcionais serem muito marcadas tanto de um ponto de vista interlinguístico como intralinguístico (ver, mais adiante 7.2.3.2. *O Nível Limiar*). Essa é, provavelmente, a razão por que alunos (e professores...) de

línguas e culturas muito afastadas das ocidentais têm manifestado tanta resistência em relação a esta abordagem (cf. Burnaby & Sun 1989).

Sinclair e Renouf (1988:160) afirmam que, **numa abordagem lexical, listas de estruturas, noções e funções serão secundárias, e resultarão da aplicação desta abordagem em vez de a controlarem.** Um livro de aluno baseado nesta abordagem será radicalmente diferente de um convencional e, acrescentam, porque a posição estruturalista ainda é muito forte, encontrará, no início, muita resistência.

Enquanto a investigação não fornecer dados suficientes para o estabelecimento de estádios de desenvolvimento que sirvam de orientação para o ensino, **uma abordagem centrada no léxico, que concilie os dados da aquisição com os dados da investigação em corpora de L1 e de L2, tendo em consideração as características e objectivos do aprendente, poderá ser mais um passo na direcção de uma doutrina para o ensino.**

7.2.3. Releitura de materiais disponíveis para o ensino de PE L2

Se, de facto, uma abordagem preferencialmente lexical parece ser a que melhor responde a um equilíbrio entre vocabulário, gramática e pragmática, do mesmo modo que certos pares de conceitos devem ser relidos à luz desta abordagem, será conveniente fazer o mesmo em relação a certos materiais de apoio disponíveis para aos agentes de ensino. Destacam-se entre estes o *Português Fundamental* e o *Nível Limiar*. O primeiro está, como vimos, na linha dos materiais que propõem que o léxico a ensinar deve ser controlado e resulta da aplicação de critérios estatísticos bastante fundamentados. O segundo responde às preocupações da abordagem comunicativa e é constituído por listas de estruturas e de actos de fala que resultaram de escolhas arbitrárias dos seus autores e não de inquéritos feitos aos falantes nativos.

7.2.3.1. O Português Fundamental

No caso do *Português Fundamental*, e apesar de ser desejável muito mais evidência como a que compõe o *corpus* aqui estudado, os resultados parecem sugerir que uma reelaboração destes materiais no sentido de distinguir o vocabulário de alta frequência, tendencialmente atemático, do

vocabulário mais ligado a certos **centros de interesse**, poderia torná-los bastante mais úteis.

A primeira e mais marcante conclusão do estudo diz respeito à **dimensão do vocabulário**. Tanto os falantes nativos como os aprendentes conseguiram escrever sobre um número relativamente grande de temas usando um vocabulário que não excedeu as 2.000 entradas, no caso dos portugueses, e que se ficou entre 1.200 e 1500, no caso dos estrangeiros. Os resultados mostram também que, à medida que a distância linguística vai aumentando, os aprendentes vão-se restringindo mais ao vocabulário mais frequente. Eles mostram ainda que das 2.217 entradas da lista do *PF*, apenas 1.454 ocorreram pelo menos uma vez; destas, apenas 371 pertencem ao conjunto de vocábulos da lista originado exclusivamente com base no inquérito de disponibilidade, o que representa menos de 3 % do total de ocorrências no *corpus*. De facto, à luz de estudos recentes, bem como dos resultados que aqui se apresentam, as 2.217 palavras propostas pelo *PF* ultrapassam largamente as 1.000 de alta frequência que cobrem 75 % de um texto e que parecem constituir o vocabulário tendencialmente atemático; além disso, segundo Nation (1993:119), com as 1000 seguintes só se ganha 6 %; em contrapartida, essas 2217 são muito menos do que as 3.000 bases (ou 5.000 palavras) que, segundo o mesmo autor, cobrem quase 90 % de qualquer texto e que muitos dizem ser indispensáveis para a compreensão; e, finalmente, estão longe de ser as 10.000 que muitos consideram como o vocabulário receptivo mínimo indispensável para ler textos académicos. **Tendo em conta que o *PF* recolheu 6.000 entradas, seria de fazer uma nova proposta de lista que distinguisse as de alta frequência e que organizasse as restantes por centros de interesse.**

Uma reelaboração dos materiais do *PF* deveria privilegiar a inclusão das concordâncias mais frequentes do vocabulário de alta frequência. Macedo 1995, em “Análise de regências fixas (*dar com, dar por,...*) a partir de listas de concordâncias”, refere um projecto em desenvolvimento no CLUL, com o apoio do programa LINGUA / SOCRATES, que usa como dados o *Corpus de Referência do Português Contemporâneo (CRPC)*, que tem como objectivo a «elaboração de material didáctico para o ensino e aprendizagem do português língua segunda», e que propõe o ensino de regências fixas «como unidade gráfica descontínua». **Projectos como este**

teriam toda a vantagem em ter em consideração os resultados de trabalhos de aquisição, nomeadamente, cruzando informação relativa à frequência de uso na L1 – distinguindo oral e escrito – com as ocorrências nos corpora de L2. Na verdade, os dados analisados nesta dissertação apontam, por exemplo, para um primeiro estágio em que os verbos referidos em Macedo 1995 (*dar* e *fazer*) são usados frequentemente em combinatórias sem preposição (ver 3.3.3. *Desvios nos verbos*); isto é válido em todos os casos, excepto para os falantes de espanhol. Parece portanto, como seria de prever, que estes atingem mais rapidamente um segundo estágio em que ocorrem regências fixas. Este é mais um exemplo de como os dados da aquisição devem ser tidos em consideração na sequencialização de materiais para o ensino. Por outro lado, chamam-nos também a atenção para a inadequação e para os limites de materiais de ensino muito generalistas.

Uma apresentação dos materiais do *PF* (e do *Corpus de Referência*, em particular para a modalidade escrita), que fornecesse as suas mais frequentes combinatórias estaria de acordo com posições actuais que desafiam claramente a distinção entre gramática e vocabulário, entre competência e performance, que sugerem uma pedagogia que concilie uma abordagem comunicativa com o ensino centrado no léxico, e poderia ser um primeiro passo para a elaboração de uma gramática para o ensino de português enquanto L2.

7.2.3.2. O Nível Limiar

«O Nível Limiar do Português integra-se no Projecto de Línguas Vivas do Conselho da Europa» e, tal como os já elaborados para outras línguas, «é um documento de apoio aos agentes de ensino do português como língua estrangeira na fase de iniciação» (*NL* pp 3).

Como referi antes, o *PF* não é exactamente discurso oral em acção. Mas também o *Nível Limiar* não se apoiou em *corpora* de oralidade obtidos em interacção verbal ³³¹. O repertório de estruturas aí incluídas (os ingleses

³³¹ «Na elaboração do **Nível Limiar** foi utilizada uma *metodologia apriorística*. Por conseguinte, não se procedeu a análises rigorosas das necessidades comunicativas dos públicos-alvo nem se fizeram inquéritos sobre o uso da língua por falantes nativos. O Nível Limiar, como os já realizados para outras línguas, foi, portanto, produzido a partir da

chamam-lhes *languages functions* e os franceses *actes de parole*) resultaram de escolhas arbitrárias dos seus autores e não de inquéritos feitos aos falantes nativos. Os autores franceses reconhecem, no entanto, que «comme pour le *français fondamental*, des enquêtes de cette nature pourraient – devraient – être entreprises.» (Coste *et alii* 1976:36).

Apesar de «o processo de ensino / aprendizagem» ser «sempre centrado no aprendente e não na matéria a ensinar» e de «esta concepção obriga[r] a que os objectivos de aprendizagem sejam definidos em termos comunicativos» (NL:3), é certamente porque não se fizeram «inquéritos sobre o uso da língua por falantes nativos» (NL:9) que tantos dos enunciados nos parecem pouco verosímeis, até mesmo na boca de falantes nativos muito proficientes, e mais preocupados em ilustrar determinadas estruturas gramaticais do que em exemplificar reais actos de fala ³³². Por outro lado, muitas das estruturas gramaticais em que se apoiam determinados actos de fala aí registados são muito marcadas tanto numa perspectiva interlinguística como intralinguística, ultrapassando, por isso, em muito, tudo aquilo que se possa entender como “fase de iniciação” de aprendizagem de L2, por mais elástico que este conceito possa ser, ou por menor que possa ser a distância entre o português e a L2 do aprendente ³³³.

Segundo os autores, o facto de se visar uma **competência comunicativa** implica que a aprendizagem da língua «enquanto sistema terá de ser feita em subordinação a este objectivo pragmático». «Esta concepção implica também que a progressão na aprendizagem seja uma progressão na comunicação e não uma progressão no conhecimento da gramática. A maior ou menor complexidade das formas linguísticas a nível do sistema é irrelevante quando os objectivos a atingir são comunicativos.» (NL:5).

intuição de falantes nativos do português dos seus autores e com base em opções acordadas com peritos do Conselho da Europa.» (NL, pp. 9).

³³² Veja-se, por exemplo: *Não é preciso que eles vão à praia* (pp150); *Diria que ele vem mais tarde* (pp153); *Tenho preferência por livros policiais* (pp162); *Exijo que me faça isso* (pp171); *Aconselhava-te a que fosses ao teatro* (pp 173). Em contrapartida, em relação a este último caso, por exemplo, não se regista *Se eu fosse a ti, ia ao teatro*.

³³³ São exemplos de estruturas muito marcadas *Parecia-me que devíamos repetir tudo* (pp 152); *Sou pela anulação da reunião* (pp153); *Permitia-me que passasse à sua frente?* (pp168); *Que dizem a ficar em casa?* (pp173); *Estou em apuros* (pp177); *Foi muita gentileza da sua parte* (pp203).

«A maior ou menor complexidade das formas linguísticas a nível do sistema» talvez possa ser «irrelevante», em certos casos, se estivermos a pensar em objectivos comunicativos para a descodificação da mensagem e para uma leitura adequada das intenções do interlocutor, mas o grau de complexidade não é certamente “irrelevante” se estivermos a pensar em produção de discurso pelo aprendente.

Apesar do destaque que a abordagem cognitiva atribui à memória, e apesar de se considerar que o aprendente pode recorrer a uma base de dados constituída por elementos soltos (*‘exemplar-based system’*), ele tem de dispor de uma gramática básica de apoio (*‘rule-based system’*) e essa segue indiscutivelmente sequências de aprendizagem impossíveis de serem alteradas pelo ensino (ver 7.2.2.2. *Competência e performance* e 7.2.2.4. *Categorias nocionais e categorias funcionais*).

Ou seja, o *Nível Limiar* do Português fornece um vasto número de opções nocionais e funcionais; muitas das quais, em especial muitas das funções, no que respeita a aprendibilidade, completamente desajustadas até mesmo para falantes de PL2 muito proficientes. Porque, de facto, uma abordagem preferencialmente lexical e cognitiva é aquela que, para o ensino, melhor responde a um equilíbrio entre vocabulário, gramática e pragmática, e que respeita estádios de aquisição, e tendo em conta que há tão poucos materiais disponíveis que apoiem os agentes de ensino, o *PF* e o *Nível Limiar* não devem ser abandonados, mas devem ser “reciclados”. Assim, do mesmo modo que seria desejável uma adequação dos materiais do *PF* ao ensino, tal como foi sugerido antes, também o *Nível Limiar* devia ter em consideração que muitos professores de português L2 não têm formação suficiente e adequada que lhes permita ter uma posição crítica perante os materiais fornecidos, e, em conformidade, disponibilizar um apêndice com informação fundamentada e actualizada, que tivesse em conta frequência, utilidade e aprendibilidade e que guiasse os seus utilizadores.

7.3. *Formação de professores e investigação sobre o PE L2*

Language teaching sometimes claims to be a profession. If it is, its practitioners cannot simply rely on recipes and techniques; they need an explicit theoretical basis for their classroom procedures. In

my view, too few language teachers exhibit the kind of intellectual curiosity and readiness to change which is normally associated with professional status. Linguistics and methodology are both comparatively new disciplines and major developments have occurred in recent years. It is disappointing that so few teachers are anxious to inform themselves about such changes, and incorporate the insights into their teaching. (Lewis 1993:ix)

Há já quase trinta anos, quando os estudos de aquisição de L2 começaram a despertar o interesse dos investigadores e dos agentes de ensino na Europa e na América, e quando, ao mesmo tempo, Portugal e o Português despertavam o interesse do mundo, aqui, felizmente, estávamos todos tão ocupados com a paz, com o pão, com a habitação e com a saúde, que a motivação para aprender e ensinar a nossa língua dispensava a investigação (ver Leiria 1989). À medida que a vida foi tomando o seu curso de normalidade cinzenta, que outros povos e outras línguas foram ocupando a atenção do mundo, e que, entre nós, se foi falando cada vez mais de “Lusofonia”, a motivação para aprender e ensinar a nossa língua foi deixando de dispensar a investigação.

Faerch afirma que «It is usually taken for granted that for teaching to be conducted at an appropriately high level presupposes that it is dependent on ongoing involvement in research. The opposite position, that research could be dependent on close involvement in teaching or equivalent activity, is not so often discussed.» (Faerch 1982:61-2; citado em Phillipson 1991:4-5). O que esta afirmação nos diz é que o ensino e a investigação têm de andar de mãos dadas e que cada professor é um potencial investigador. A experiência de ensino dos professores de língua constitui um tesouro que deve ser rentabilizado pela investigação. No entanto, para que a experiência, muitas vezes construída sobre a intuição e sobre convicções infundadas, não se transforme num mal maior, os professores de língua devem apoiar o ensino em investigação. Contrariamente àquilo que acontecia há alguns anos, há hoje um corpo de trabalhos em aquisição de outras línguas e em linguística descritiva do português que pode ser posto ao serviço da investigação e da formação de professores de PL2.

Ouve-se dizer com frequência que faltam materiais para o ensino do PE L2. É verdade. Mas não é menos verdade que essa falta decorre do amadorismo com que esta área tem sido encarada. Só haverá materiais em

número e qualidade suficientes quando houver agentes de ensino com formação adequada. Este objectivo pode ser atingido, antes de mais, com um forte empenho na formação de professores, formação essa que os prepare para uma observação crítica dos materiais disponíveis, o que corresponde, pelo menos, a introduzi-los a procedimentos próximos dos da investigação, a nível metodológico.

A aprendizagem e o ensino do português enquanto L2 acontece em três espaços com características distintas: **(1)** em Portugal e no Brasil, onde ele é L1 para a maior parte dos falantes e onde, portanto, os aprendentes estrangeiros se encontram em contexto de imersão linguística, combinado ou não com aprendizagem formal, **(2)** em países ou territórios que foram colónias de Portugal, onde ele é L2 para muitos falantes e L1 para alguns, e ensinado por professores com as mesmas características; e **(3)** em outros países do mundo, onde ele é aprendido como LE, e ensinado tanto por professores falantes não nativos como por falantes nativos de PB como de PE, nomeadamente por leitores do Instituto Camões.

Qualquer um destes contextos de aprendizagem tem potencialidades para investigação com características e objectivos próprios ³³⁴. E só combinando a investigação realizada nos diversos espaços se pode ter uma visão do que significa aprender português. Em relação a línguas em que a investigação já é bastante abundante, como é o caso do Alemão, do Francês, do Inglês ou do Sueco, costumam os investigadores lamentar que a área da LE tenha prestado muito mais atenção ao ensino do que à investigação; e que, caso apoiem o ensino em investigação, o que nem sempre acontece, sejam sobretudo devedores de investigação feita em contexto de aprendizagem informal, ou em contexto combinado de informal e formal. São exemplo de aprendizagem exclusivamente informal vários projectos desenvolvidos na Alemanha e de aprendizagem combinada os desenvolvidos na Suécia, onde a frequência de cursos de língua é obrigatória para os imigrantes. No que respeita o PE, a questão está em que a investigação em contexto de aprendizagem informal ou semi-formal, ou

³³⁴ Os volumes III e IV do PPOM (Gonçalves & Stroud 1998 e 2000) são um bom exemplo de aplicação de resultados de investigação ao ensino.

seja a que seria realizada em Portugal, é praticamente inexistente, não estando, portanto, disponível para os professores de PE LE.

No entanto, há determinados estudos que devem ser desenvolvidos em contexto de aprendizagem exclusivamente formal, ou seja LE. A ausência da variável imersão permite estudos sobre o papel que o ensino pode ter na aquisição e desenvolvimento linguístico impossíveis de levar a cabo em contexto de aprendizagem informal ou semi-formal. É possível constituir grupos de sujeitos com características semelhantes (sendo a L1, neste caso, geralmente a mesma, pode controlar-se mais facilmente o nível de conhecimentos na língua em estudo, assim como o conhecimento de outras L2) e, por exemplo, treinar cada um dos grupos recorrendo a métodos de ensino muito diversos, ou ensinar a um grupo um determinado aspecto da gramática e a outro não. Por exemplo, em que medida é que uma regra fornecida na sala de aula, abusivamente generalizada («depois de se usa-se conjuntivo»), actuando como filtro sobre uma estrutura que parece adquirida, é responsável por esta substituição?:

A03 *No autocarro sentei-me ao lado dum rapaz muito giro. [...] quando vi o rapaz estava a rir também. [Eu] perguntou-lhe se <falava> falasse inglês.*

Trabalhos de investigação que comparem produções de aprendentes em contexto de aprendizagem informal (por exemplo, alemães a viverem em Portugal que nunca tenham frequentado aulas) e aprendentes de PLE (alemães que estudem numa universidade alemã) poderão fornecer-nos informação sobre este e outros aspectos relacionados com o papel do ensino na aprendizagem.

Quero com isto dizer que qualquer professor, independentemente do seu contexto de ensino, pode desenvolver investigação. A questão está em ter formação que o desperte para a necessidade e para as vantagens da sua realização e que o guie na selecção do objecto adequado ao seu contexto de ensino. Por outro lado, o português é raramente a segunda língua que os estrangeiros aprendem. Ela é quase sempre a terceira ou até mesmo a

quarta. Convém, portanto, que o professor, além de ter conhecimentos de linguística descritiva, que lhe permita contrastar o português com a L1 dos seus alunos, os tenha também de outras línguas que eles conheçam. Dou um exemplo. Ao contrário do que acontece com a maior parte dos documentos que constituem o *sub-corpus* alemão analisado nesta dissertação, o documento A09 quase não tem desvios no que respeita a oposição aspectual codificada pelos Pretéritos Perfeito e Imperfeito. Isso deve-se certamente ao facto de, como pude verificar na ficha do aluno, ele saber russo que, como se sabe, apoia o seu sistema verbal na distinção aspectual Perfectivo / Imperfectivo (ver Leiria 1991).

Se os estudos de linguística descritiva e de aquisição da linguagem podem ajudar os professores a prever e a explicar o papel dos conhecimentos prévios na aquisição de L2, também os estudos sobre a variação do português são indispensáveis. Actualmente, os estudantes, futuros professores de português, chegam à universidade (e será que muitos não saem da universidade?) com preconceitos que reflectem os da classe dominante, e que ficaram patentes, em 1986, aquando da polémica sobre o Acordo Ortográfico. Políticos, escritores, jornalistas e até poetas e sociólogos confundiram escrita com língua e acharam-se os donos desta ³³⁵. Falaram e continuam a falar de Lusofonia e de defesa da língua, contabilizaram os milhões de falantes, concluíram que o português é a quinta (a quarta? a sexta? a nona?) língua mais falada do mundo, mas esqueceram-se de dizer, entre outras coisas, que Portugal tem menos habitantes do que S. Paulo e que, quem o diz são os historiadores da língua (cf. Teyssier 1980), em muitos aspectos o PE está hoje muito mais afastado do português do século XVI do que o PB. Durante a polémica, querendo dar “uma achega construtiva”, alguém propôs «o recrutamento de qualificados professores com espírito de missão, que estejam dispostos a ser colocados no Brasil [para darem] aulas intensivas especializadas em zonas mais complexas e susceptíveis da língua (pronomes, participio presente, concordâncias, etc.)». ³³⁶

³³⁵ Em Castro et *alii* (1986:219-275) transcrevem-se extractos de textos de jornal que ilustram esta afirmação.

³³⁶ Luísa Costa Gomes e Mário de Carvalho, *Jornal de Letras* de 09.06.1986. Ver Castro, Duarte & Leiria 1986:239.

Para evitar posições preconceituosas e desinformadas como esta que, em nome da defesa da língua, podem levar os professores de português no estrangeiro a hostilizar os seus colegas brasileiros, criando situações incompreensíveis para os alunos, e que, em última análise, podem ter como efeito a diminuição do interesse pelo estudo da língua portuguesa, é indispensável também uma formação mínima em sociolinguística. Essa formação é ainda mais imprescindível nas nossas relações com os países africanos em que o português é língua oficial ou com Timor.

Actualmente, a linguística, se ainda não conseguiu responder de forma decisiva à questão “como se aprende uma língua?”, desenvolveu disciplinas que, combinadas, podem fornecer aos professores uma melhor compreensão deste fenómeno: a gramática descritiva, a sociolinguística, a dialectologia, a pragmática, a história da língua, mas também a psicolinguística, a neurolinguística e, sem dúvida, a aquisição de L1 e de L2. Uma formação que contemple estes saberes fornecerá ao professor atitudes mais saudáveis e informadas em relação à língua que ensinam. Contudo, trabalhos de investigação, porque muito direccionados, específicos e teóricos, são, muitas vezes, árvores que escondem a floresta. Por isso, para além de estudos sobre aquisição do PE L2, e para os incentivar, seria desejável que os especialistas portugueses em cada uma das áreas relevantes se empenhassem na redacção de textos de divulgação que despertassem o interesse dos professores e que os incentivassem a ir mais além.

7.4. Conclusão

Os resultados do estudo que constitui o objecto central desta dissertação mostram claramente as vantagens de um ensino centrado no léxico que tenha em consideração os dados da aquisição e são aplicáveis (com alguns cuidados e adaptações) a qualquer situação de ensino / aprendizagem de PE L2. Isso deve-se, pelo menos em parte, ao facto de não se ter adoptado uma metodologia contrastiva, mas de a perspectiva de análise ter posto a tónica nas especificidades do PE. Por isso, e pela variedade de línguas maternas que constituem os conhecimentos prévios dos aprendentes envolvidos, eles disponibilizam informação que pode apoiar os agentes de ensino, tanto de falantes das línguas consideradas como de outras da mesma família. Eles permitem também uma reflexão a autores de materiais didácticos na hora de seleccionarem e de sequencializarem os conteúdos, a avaliadores na hora de julgarem o peso que devem atribuir àquilo que está correcto e àquilo que está incorrecto. Eles podem também servir de guia a um lexicógrafo que decida realizar um dicionário básico para estrangeiros, ou, quem sabe, a um gramático que decida empreender a tarefa de fazer uma gramática para estrangeiros.

O português tem uma história de ensino como L2 não direi tão pujante mas, no mínimo, tão extensa como outras grandes línguas, nomeadamente como o inglês. Como referi na Introdução a esta dissertação, o objectivo prévio deste estudo é o de contribuir modestamente para começar a criar um corpo de estudos que disponibilize informação sobre aquisição do PE L2 aplicável ao ensino, porque é minha convicção que a melhor defesa que nós, professores, podemos fazer da nossa língua é aprendermos como ela se aprende, para aprendermos a ensiná-la.

REFERÊNCIAS BIBLIOGRÁFICAS

ACOSTA, L & I LEIRIA

1997 O papel dos conhecimentos prévios na aquisição de uma língua não materna, *Polifonia (Revista do Grupo Universitário de Investigação em Línguas Vivas-UNIL)* Lisboa, Edições Colibri, nº1:57-80.

AITCHINSON, J & D LEWIS

1995 How to handle wimps: incorporating new lexical items as an adult, *Folia Linguistica* XXIX, 7-20.

AITCHINSON, J

1992 *Language and Mind*, Penguin English.

1994 *Words in the Mind. An Introduction to the Mental Lexicon*, Oxford, Blackwell.
1987.

1996 Taming the wilderness: words in the mental lexicon, G Anderman & M Rogers (eds.) *Words, Words; Words. The Translator and the Language Learner*, Clevedon, Multilingual Matters 15-26.

ALLAN, K

1992 Semantics: An overview, W Bright (ed.), *International Encyclopedia of Linguistics*, Oxford, OUP, vol. 3: 394-399.

ALLWRIGHT, J

1991 The characterization of teaching and learning environments: problems and perspectives, K de Bot & R Ginsberg & C Kramersch (eds.), *Foreign Language Research in Cross-Cultural Perspective*, Amsterdam, J Benjamins, 161-174.

ALTMAN, R

1997 Oral production of vocabulary: a case study, J Coady & T Huckin (eds.) *Second Language Vocabulary*, Cambridge, CUP, 69-97.

AMRHEIN, P

1992 The comprehension of quasi-performative verbs in verbal commitments: New evidence for componential theories of lexical meaning, *Journal of Memory and Language* 31:756-784.

ANDERMAN, G & M ROGERS (eds.)

1996 *Words, Words, Words. The Translator and the Language Learner*, Clevedon, Multilingual Matters.

ANDERMAN, G & M ROGERS

1996 The translator and the language learner - linguistics revisited, G Alderman & M Rogers (eds.) *Words, Words, Words. The Translator and the Language Learner*, Clevedon, Multilingual Matters 1-14.

ANDERMAN, G

1996 The word is my oyster: the language learner and the translator, G Alderman & M Rogers (eds.) *Words, Words, Words. The Translator and the Language Learner*, Clevedon, Multilingual Matters 41-55.

ANDERSEN, R W

1979 Expanding Schumann's pidginization hypothesis, *Language Learning* 29:105-119.

1983 Transfer to somewhere, Gass & Selinker (eds.) *Language Transfer in Language Learning*, Rowley, MA: Newbury House, 83-188.

1984 *Second Language. A Cross-linguistic Perspective*, Rowley, Newbury House.

1990 Models, processes, principles and strategies: Second language acquisition inside and outside the classroom, B VanPatten & J F Lee (eds.) *Second Language Acquisition and Foreign Language Learning*, Philadelphia, Clevedon, 45-71.

ANDERSON, J M

1971 *The Grammar of Case: Towards a Localistic Theory*. Cambridge, CUP.

1983a *The Architecture of Cognition*, Cambridge, Mass., Harvard University Press.

1983b A spreading activation theory of memory, *Journal of Verbal Learning and Verbal Behavior* 22: 261-295.

1995 Varieties of Case Grammar, in E F K Koerner & R E Asher (eds.) *Concise History of the Language Sciences. From the Sumerians to the Cognitivists*, Oxford, Pergamon, 352-364.

ANDRADE, E & VIANA, M C

1993 Sinérese, diérese e estrutura silábica, *Actas do IX Encontro da APL*, Lisboa, APL.

ARD, J & S GRASS

1987 Lexical constraints on syntactic acquisition, *Studies in Second Language Acquisition* 9: 233-252.

- ARNAUD, P & S SAVIGNON
 1997 Rare words, complex lexical units and the advanced learner, J Coady & T Huckin (eds.) *Second Language Vocabulary Acquisition*, Cambridge, CUP, 157-173.
- ARONOFF, M
 1976 *Word Formation in Generative Grammar*, Cambridge, Mass, The MIT Press.
- AU, T KIT-FONG
 1983 Chinese and English counterfactuals: the Sapir-Whorf hypothesis revisited, *Cognition* 15, 155-187
- AXELSSON, M
 1994 *Noun phrase development in Swedish as a second language. A study of adult learners acquiring definiteness and the semantics and morphology of adjectives*, Centre for Research on Bilingualism, Stockholm University.
- BAILEY, N & C MADDEN & S KRASHEN
 1974 Is there a 'Natural Sequence' in Adult second language learning? *Language Learning* 24:235-243.
- BARALO, M
 1997 La organización del lexicón en lengua extranjera, *Revista de Filología Románica* 14, v. 1:59-71, Madrid, Universidade Complutense.
- BARBOSA, A G
 1999 *Para uma história do português colonial: aspectos linguísticos em cartas de comércio* (2 vol.) Dissertação de doutoramento apresentada à Universidade Federal do Rio de Janeiro.
- BATES, E & B MACWHINNEY
 1982 Functionalist approaches to grammar, E Wanner & L R Gleitman (eds.) *Language and Acquisition. The State of the Art*, Cambridge, CUP, 173-218.
 1989 Functionalism and the Competition Model, B MacWhinney & E Bates (eds.) *The crosslinguistic study of sentence processing*, New York, CUP, 3-73.
- BATORÉO, H J
 1996 *Contribuição para a caracterização da interface expressão linguística - cognição espacial no Português Europeu: abordagem psicolinguística da expressão do espaço em narrativas provocadas*. 2 volumes. Dissertação de Doutoramento apresentada à FLUL.
- BAUER, L & P NATION

1993 Word families, *International Journal of Lexicography*, 6:259-279.

BECHARA, E

1999 *Moderna Gramática Portuguesa*, Rio de Janeiro, Editora Lucerna, 37ª edição revista e ampliada.

BENSON, M

1990 Collocations and general-purpose dictionaries, *International Journal of Lexicography*, 3:23-35.

BEVER, T

1970 The cognitive basis for linguistic structures, J Hayes (ed.) *Cognition and the development of language*. New York: Wiley.

BIALYSTOK, E & M SHARWOOD SMITH

1985 Interlanguage is not a state of mind: an evolution of the construct for second language acquisition, *Applied Linguistics* 6:101-117.

BIALYSTOK, E

1978 A theoretical model of second language learning, *Language Learning* 28:69-83.

1981 The role of linguistic knowledge in second language use. *Studies in Second Language Acquisition* 4:31-45.

1982 On the relationship between knowing and using forms. *Applied Linguistics* 3:181-206.

1990 *Communication Strategies*, Oxford, Basil Blackwell.

1994 Representation and ways of knowing: Three issues in second language acquisition, N Ellis (ed.) *Implicit and Explicit Learning of Languages*, Londres, Academic Press, 549-569.

BICKERTON, D

1975 *Dynamics of a Creole System*. Cambridge:CUP.

1984 The language bioprogram hypothesis, *The Behavioral and Brain Sciences* 7:173-221.

BIDERMAN, M T C

1994 Vocabulário fundamental: cultura e sociedade, *Actas do Congresso Internacional sobre o Português*, vol. III, Lisboa, APL, 215-237.

1995 Desenho e análise de um *corpus* de português contemporâneo, *Actas do XI Encontro da APL*, vol. I, Lisboa, APL, 129-141.

BIERWISCH, M & R SCHEREUDER

1992 From concepts to lexical items, *Cognition* 41:23-60.

- BLANCHE-BENVENISTE, C
 1993 Les unités: langue écrite, langue orale, C Pontecorvo & C Blanche-Benveniste (eds.), *Proceedings of the Workshop on Orality versus Literacy: Concepts, Methods and Data*. Siena, Italy, 24-26 September 1992. Estrasburgo, European Science Foundation, 133-194.
- BLEY-VROMAN, R
 1989 What is the logical problem of foreign language learning? S Gass & J Schachter (eds.) *Linguistic Perspectives on Second Language Learning*, Cambridge, CUP, 44-68.
- BLOOMFIELD, L
 1933 *Language* New York : Holt, Rinehart & Winston
- BLUM, S & E A LEVENSTON
 1978 Universals of Lexical Simplification, *Language Learning* 28/2: 399-415.
- BLUMENTHAL, A
 1970 *Language and Psychology: Historical Perspectives in Psycholinguistics*. New York, John Wiley & Sons.
- BOGURAEV, B & J PUSTEJOVSKY
 1996 (eds.) *Corpus Processing for Lexical Acquisition*, Cambridge, Mass, The MIT Press.
 1996 Issues in Text-based lexicon acquisition, B Boguraev & J Pustejovsky (eds.), 3-17.
- BOLINGER, D L
 1963 The uniqueness of the word, *Lingua* 12:113-136.
- BONGAERTS, T & B PLANKEN & E SCHILS
 1995 Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis, D Singleton & Z Lengyel (eds.) *The Age Factor in Second Language Acquisition*, Clevedon, Multilingual Matters, 30-50.
- BOWERMAN, M
 1989 Learning a semantic system: What role do cognitive predispositions play? M L Rice & R L Schiefelbusch, (eds.), *The Teachability of Language*, Baltimore , P H Brookes, 133-169.
- BRESNAN, J & R M KAPLAN
 1982 Lexical-Functional Grammar: A Formal System for Grammatical Representation, J Bresnan (eds.), *The Mental*

Representation of Grammatical Relations, Cambridge, Mass.: MIT Press, cap 4:173-281.

BRITO, A M

1998 A relação semântica lexical - sintaxe na gramática generativa: um breve balanço a propósito da natureza aspectual e da estrutura argumental de alguns tipos de verbos, *Revista da Faculdade de Letras - Línguas e Literaturas*, Porto, XV, 377-420.

1999 *Os estudos de sintaxe generativa em Portugal nos últimos trinta anos*, APL Braga.

BROEDER, P & G EXTRA & R VAN HOUT, & K VOIONMAA

1993 Word formation processes in talking about entities, C Perdue (ed.) *Adult Language Acquisition: Cross-linguistic Perspectives*. Vol 2, *The Results*, Cambridge, CUP 41-72.

BROEDER, P & K PLUNKETT

1994 Connectionism and second language acquisition, N Ellis (ed.) *Implicit and Explicit Learning of Languages*, Londres, Academic Press, 421-453.

BROWN, T S & F L PERRY JR.

1991 A comparison of three learning strategies for ESL vocabulary acquisition, *Tesol Quarterly* 25: 655-670.

BROWN, R

1973 *A first language. The early stages*. Cambridge, Harvard University Press.

BURNABY, B & Y SUN

1989 Chinese teachers' views of western language teaching: context informs paradigms, *Tesol Quarterly* 23:219-239.

BURT, M K & H C DULAY

1975 (ed.) *New Directions in Second Language Learning teaching and Bilingual Education*, Washington, TESOL.

BUTLER, C & R MAIRAL & J MARTÍN ARISTA & F J RUIZ DE MENDOZA

1999 (eds.) *Nuevas perspectivas en Gramática Funcional*, Barcelona Editorial Ariel S A.

BYBEE, J L

1985 *Morphology: A study of the relation between meaning and form*, Amsterdam, Benjamins.

- CABRAL, L
 2005 *Complementos verbais preposicionados do Português em Angola*,
 Dissertação de Mestrado apresentada à Faculdade de Letras da
 Universidade de Lisboa.
- CAGLIARI, L C
 1999 Breve história das letras e dos números, G Massini-Cagliari & L
 C Cagliari, *Diante das Letras. A escrita na alfabetização*.
 Campinas, Mercado de Letras, 163-185 (1993)
- CANO AGUILAR, R
 1999 Los complementos de régimen verbal, I Bosque & V Demonte
 (dir.) *Gramática Descriptiva de la Lengua Española*, Madrid,
 Real Academia Española / Espasa Calpe, vol. 2:1807-1854.
- CARROLL, S E
 1995 The hidden dangers of computer modelling: remarks on Sokolik
 and Smith's connectionist learning model of French gender,
Second Language Research 11:193-205.
- CARTER, R
 1987 *Vocabulary: applied linguistic perspectives*, London, Allen &
 Unwin.
- CARTER, R & M MCCARTHY
 1988a Word lists and learning words: some foundations, R
 Carter & M McCarthy (eds.) *Vocabulary and Language
 Teaching*, Londres, Longman, 1-17.
 1988b Lexis and structure, R Carter & M McCarthy (eds.)
Vocabulary and Language Teaching, Londres, Longman, 18-38.
 1988c New directions in vocabulary studies, R Carter & M
 McCarthy (eds.) *Vocabulary and Language Teaching*, Londres,
 Longman, 60-61.
- CASTELEIRO, J M (dir.)
 2001 *Dicionário da Língua Portuguesa Contemporânea*, Lisboa,
 Academia das Ciências de Lisboa / Editorial Verbo.
- CASTELEIRO, J M & A MEIRA & J PASCOAL
 1988 *Nível Limiar. Para o ensino/aprendizagem do português como
 língua segunda / língua estrangeira*, Strasbourg, Conselho da
 Europa; Lisboa, Instituto de Cultura e Língua Portuguesa.
- CASTILHO, A T DE
 1995 A Gramática do Português Culto Falado no Brasil, *Actas do XI
 Encontro da Associação Portuguesa de Linguística*, vol. III, 73-
 103.

- CASTRO, I & I DUARTE & I LEIRIA
 1987 *A Demanda da Ortografia Portuguesa. Comentário do Acordo Ortográfico de 1986 e subsídios para a compreensão da questão que se lhe seguiu*, Lisboa, Edições João Sá da Costa.
- CASTRO, S L
 2000 A linguagem escrita e o seu uso: uma perspectiva cognitiva. M R Delgado-Martins & G Ramalho & A Costa (org.), *Literacia e Sociedade. Contribuições Plurais*, Lisboa, Caminho, 131-147.
- CASTRO-CALDAS, A & A REIS
 2000 Neuropsicologia do analfabetismo. Considerações a propósito de um projecto de desenvolvimento, M R Delgado-Martins & G Ramalho & A Costa (org.), *Literacia e Sociedade. Contribuições Plurais*, Lisboa, Caminho, 155-183.
- CHAMOT, A & J O'MALLEY
 1994 Language learner and learning strategies, N Ellis (ed.) *Implicit and Explicit Learning of Languages*, Londres, Academic Press, 371-392.
- CHEN, P
 1993 Modern written Chinese in development, *Language in Society* 22:505-537.
- CHOMSKY, N
 1957 *Syntactic Structures* The Hague: Mouton.
 1959 Recensão a B F Skinner, Verbal Behavior; *Language* 35:26-57.
 1965 *Aspects of the theory of syntax*, Cambridge, Mass: MIT Press.
 1970 Remarks on Nominalizations, R A Jacobs, & P Rosenbaum (orgs) *Readings in English Transformational Grammar*, Ginn & Company, Waltham, Mass, 184-221.
 1981 *Lectures on Government and Binding*, Dordrecht: Foris.
 1982 *Some Concepts and Consequences of the Theory of Government and Binding*, Cambridge, Mass., MIT Press.
 1986a *Barriers*, Cambridge, Mass, Mit Press.
 1986b *Knowledge of language: its nature, origin and use*. New York Preager; tradução port.: *O conhecimento da língua. Sua natureza, origem e uso*, pref. I S Duarte, IX-XVI, Caminho, Lisboa, 1994.
 1989 *Some notes on economy of derivation and representation*, MIT Working Papers in Linguistics 10, 43-74.
 1995 *The Minimalist Program*, MIT Press, Cambridge, Mass.
- CHOMSKY, N & M HALLE
 1968 *The sound pattern of English*, New York: Harper & Row.

- CLAHSEN, H
 1985 Profiling second language development: a procedure for assessing L2 proficiency, K Hyltenstam & M Pienemann, *Modelling and Assessing Second Language Acquisition*, Clevedon: Multilingual Matters, 283-332.
- CLAHSEN, H & P MUYSKEN
 1986 The availability of Universal Grammar to adult and child learners - a study of the acquisition of German word order, *Second Language Research* 2:93-119.
- CLARK, E
 1978 Discovering what words can do, *Papers from the Parasession on the Lexicon*, CLS 14, 34-57.
 1993 *Lexicon in Acquisition*, Cambridge, CUP.
- COHEN, A
 1989 Attrition in the productive lexicon of two Portuguese third language speakers, *Studies in Second Language Acquisition* 11:135-149.
 1990 Strategies in second language learning: insights from research, R Phillipson, & E Kellerman & L Selinker & M Sharwood-Smith & M Swain (eds.) *Foreign / Second Language Pedagogy Research*, Clevedon, Multilingual Matters, 107-119.
- COMRIE, B
 1976 *Aspect*, Cambridge, CUP
 1981 *Language universals and linguistic typology: syntax and morphology*, 2nd ed. 1989, Oxford, Blackwell.
- COOK, V
 1988 *Chomsky's Universal Grammar: An Introduction*. Oxford, Blackwell.
 1991 *Second Language Learning and Language Teaching*, London, Edward Arnold (2.Acquisition of Vocabulary, 37-44)
 1993 *Linguistics and Second Language Acquisition*, London, The Macmillan Press.
 1994 The metaphor of access to universal grammar in L2 learning, N Ellis (ed.) *Implicit and Explicit Learning of Languages*, London, Acad. Press, 477-502.
- CORBETT, G
 1991 *Gender*, Cambridge, Cambridge University Press.
- CORDER, S P

- 1967 The significance of learners' errors, J C Richards (ed) *Error Analysis. Perspectives on Second Language Acquisition*, New York, Longman, 1974: 19:27.
- 1971 Idiosyncratic dialects and error analysis, J C Richards (ed.) *Error Analysis. Perspectives on Second language Acquisition*, New York, Longman, 1974: 158-171.
- 1983 A role for the mother tongue, S Gass & L Selinker (eds.) *Language Transfer in Language Learning*, Amsterdam, J Benjamins, 1992: 17-31.

COSERIU, E

- 1977 *Principios de Semántica Estructural*, Madrid, Gredos.
- 1981 *Lecciones de Lingüística General*, Madrid, Gredos.

COSTE, D & J COURTILLON & V FERENCZI & M MARTINS-BALTAR & E PAPO & E ROUET

- 1976 *Un Niveau-Seuil*. Conseil de l'Europe. Projet Langues Vivantes. Hatier.

COULMAS, F

- 1999 *The Blackwell Encyclopedia of Writing Systems*, Oxford, Blackwell (1996).

COWIE, A P

- 1987 Stable and creative aspects of vocabulary use, R Carter & M McCarthy (eds.) *Vocabulary and Language Teaching*, Londres: Longman, 126-139.

CROFT, W

- 1990 *Typology and Universals*, Cambridge, CUP.

CROW, J T & J R QUIGLEY

- 1985 A semantic field approach to passive vocabulary acquisition for reading comprehension, *Tesol Quarterly* 19: 497-513.

CUNHA, C & L F L CINTRA

- 1984 *Nova Gramática do Português Contemporâneo*, Lisboa, Edições João Sá da Costa.

DAHL, O

- 1985 *Tense and Aspect Systems*, Oxford Blackwell.

D'ANDRADE R

- 1989 Cultural Cognition, M I Posner (ed.) *Foundations of Cognitive Science*, Cambridge, Mass; MIT Press, 795-830.

DE BOT, K

1992 A bilingual production model: Levelt's 'speaking' model adapted, *Applied Linguistics* 13:1-24.

DE BOT, K & R SCHEREUDER

1993 Word production and the bilingual lexicon, Schreuder & Weltens (eds.) *The Bilingual Lexicon*, Benjamins, Amsterdam, 191-214.

DE BOT, K & A COX & S RALSTON & A SCHAUFELI & B WELTENS

1995 Lexical processing in bilinguals, *Second Language Research* 11:1-19.

DE GROOT, A

1993 Word-type effects in bilingual processing tasks: support for a mixed representational system, R Schreuder & B Weltens (eds.) *The Bilingual Lexicon*, Benjamins, Amsterdam, 27-51.

DE GROOT, A & G L J NAS

1991 Lexical representation of cognates and noncognates in compound bilinguals, *Journal of Memory and Language* 30: 90-123.

DECAMP, D

1971 The study of pidgin and creole languages, Hymes (ed.) *Pidginization and Creolization of Languages*, Cambridge, CUP, 13-39.

DELL, G

1986 A spreading-activation theory of retrieval in sentence production. *Cognition* 93: 321-383.

DICKERSON, L J

1975 The learner's interlanguage as a system of variable rules, *TESOL Quarterly* 9:215-231.

DICKERSON, W B

1976 The psycholinguistic unity of language learning and language change, *Language Learning* 26:215-231.

DIETRICH, R

1990 Nouns and verbs in the learner's lexicon, H W Dechert (ed.) *Current Trends in European Second Language Acquisition Research*, Clevedon, Multilingual Matters, 13-22.

DIK, S

1968 *Coordination: Its implications for the theory of general linguistics*, Amsterdam: North-Holland

- DIK, S C
 1978 *Functional Grammar*. Amsterdam North-Holand
 1989 *The Theory of Functional Grammar*, vol I, *The Structure of the Clause*. Dordrecht: Foris.
- DIXON, R M W
 1977 Where have all the adjectives gone? *Studies in Language* 1:19-80.
- DUARTE, I
 2000 Predicados complexos e ensino do Português como língua não-materna, Com. apresent. no 1º Encontro de Linguística do ano 2000, Maputo, INDE.
- DULAY, H C & M K BURT
 1974 Natural Sequences in Child Second Language Acquisition, *Language Learning* 24:37-53.
- ECKMAN, F R
 1977 *Markedness and the contrastive analysis hypothesis*, *Language Learning* 27:315-330.
- ELLIS, N
 1994a Introduction. Implicit and explicit language learning – An overview, N Ellis (ed.) *Implicit and Explicit Learning of Languages*, London, Academic Press, 1-31.
 1994b Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation, N Ellis (ed.) *Implicit and Explicit Learning of Languages*, London, Academic Press, 211-281.
 1994 (ed.) *Implicit and Explicit Learning of Languages*, London, Academic Press.
 1995a Emergentism, Connectionism and Language Learning, *Language Learning* 48: 631-664.
 1995b Vocabulary acquisition: psychological perspectives, *The Language Teacher* 19, 2:12-16.
 1996 Sequencing in SLA; phonological memory, chunking and points of order, *Studies in Second Language Acquisition* 18: 91-126.
 1997 Vocabulary acquisition: word structure, collocation, word class, and meaning. N Schmitt & M McCarthy (eds.) *Vocabulary. Description, Acquisition and Pedagogy*, Cambridge, CUP, 122-139.
 1999 Cognitive Approaches to SLA, *Annual Review of Applied Linguistics* 19, CUP, 22-79.
- ELLIS, N & A BEATON
 1993 Psycholinguistic determinants of foreign language vocabulary learning, *Language Learning* 43:559-617.

ELLIS, N & N LAPORTE

1997 Contexts of acquisition: effects of formal instruction and naturalistic exposure on second language acquisition, A de Groot & J Kroll (eds.) *Tutorials in Bilingualism. Psycholinguistic Perspectives*, New Jersey, L Erlbaum Associates Pub, 53-83.

ELLIS, R

1984 *Classroom Second Language Development*, Oxford: Pergamon.

1994 A theory of instructed second language acquisition, N Ellis (ed.) *Implicit and Explicit Learning of Languages*, Londres, Academic Press, 79-114.

1995 Modified oral input and the acquisition of word meaning, *Applied Linguistics* 16:409-441.

ELLIS, R & Y TANAKA & A YAMAZAKI

1995 Classroom interaction, comprehension and the acquisition of the L2 word meanings, B Harley (ed.) *Lexical Issues in Language Learning*, Language Learning/Benjamins., Amsterdão, 187-229.

ERTESCHIK-SHIR, N & T R RAPOPORT

1995 A theory of verbal projection, G Matos, M Miguel, I Duarte, I Faria (orgs) *Interfaces in Linguistics Theory*, APL, Colibri, 127-148.

ERVIN, S M & C E OSGORD

1954 Second language learning and bilingualism, *Journal of Abnormal Social Psychology Supplement*, 139-146.

ERVIN-TRIPP, S M

1974 Is second language learning like the first?, *TESOL Quarterly* 8:111-127.

EUBANK, L

1996 Negation in early German-English interlanguage: more valueless features in the L2 initial state, *Second Language Research* 12:73-106.

Eurom4

1997 *Método de ensino simultâneo das línguas românicas. Método por la enseñanza simultánea de las lenguas románicas. Metodo di insegnamento simultaneo delle lingue romanze. Méthode d'enseignement simultanée des langues romanes*, Firenze, La Nuova Italia Editrice.

FAERCH, C & G KASPER

1983 Plans and strategies in foreign language communication, C Faerch & G Kasper (eds.) *Strategies in Interlanguage Communication*, London, Longman, 20-60.

FARIA, I H

1999a Expressões idiomáticas, metáforas, emoções, sentidos figurados e sujeitos experienciadores. I H Faria (org) *Homenagem a Lindley Cintra*. Lisboa, Cosmos-FLUL, 377-402.

1999b De causa a causatividade - percursos na representação verbal da causalidade. Workshop de Gramática e Cognição, Tavira, Setembro de 1999; ms.

(em preparação)

Pontos de idiomaticidade: um estudo experimental com expressões idiomáticas do Português Europeu, ms.

FARIA, I H & H BATORÉO

1994 Corpus de aquisição do Português Europeu: Base de dados CHILDES, *RILP* n°11, *A criança e a linguagem*, 137-145.

FIKKERT, P

1994 *On the acquisition of prosodic structure*, Dordrecht: HIL.

FILLMORE, C J

1968 The case for case, in E Bach & R Harms (orgs) *Universals in Linguistic Theory*, New York, Holt, Rinehart & Winston.

1976 Frame semantics and the nature of language. *Annals of the New York Academy of Science* 280:20-31.

1985 Frames and the semantics of understanding, *Quaderni di Semantica* VI:222-254.

FIRTH, J R

1957 The Technique of Semantics, *Papers in Linguistics*. Oxford, OUP (1935)

FISCHER, U

1994 Learning words from context and dictionaries: an experimental comparison, *Applied Psycholinguistics* 15: 551-574.

FLYNN, S & W O'NEIL

1988 (eds.) *Linguistic Theory in Second Language Acquisition*, Dordrecht: Kluwer Academic.

FLYNN, S

1987 *A Parameter Setting Model of L2 Acquisition: Experimental Studies in Anaphora*, Dordrecht, Reidel.

FODOR, J

1983 *The modularity of mind: an essay on faculty psychology*, Cambridge, MA, MIT Press.

FOLEY, W & R VAN VALIN

1980 Role and Reference Grammar, E Moravcsik & J Wirth (eds.) *Syntax and Semantics 13:Current Approaches to Syntax*, New York, Academic Press, 329-351.

FOLEY, W & R VAN VALIN

1984 *Functional Syntax and Universal Grammar*, Cambridge, CUP.

FORSTER, K I

1992 Basic issues in lexical processing, W Marslen-Wilson (ed.) *Lexical Representation and Process*, Cambridge Mass., MIT Press, 75-107.

FREITAS, M J

1997 *Aquisição da estrutura silábica do Português Europeu*, Dissertação de doutoramento apresentada à Universidade de Lisboa.

FRIES, C C

1945 *Teaching and Learning English as a Foreign Language*. Ann Arbor, University of Michigan Press.

FROTA, S

1992 Prosódia do advérbio na frase. Interacção e convergência, *Actas do VIII Encontro da Associação Portuguesa de Linguística* :204-224, Lisboa, APL.

GAIRNS, R & S REDMAN

1993 *Working with words. A guide for teaching and Learning Vocabulary*, Cambridge, CUP. 1986.

GALLOWAY, L M

1983 Études cliniques et expérimentales sur la répartition hémisphérique du traitement cérébral du langage chez les bilingues: modèles théoriques, *Langages* 72: 79-123.

GARCIA, G E

1991 Factors influencing the English reading test performance of Spanish-speaking Hispanic students. *Reading Research Quarterly* 26:371-92.

GASS, S

1979 Language transfer and universal grammatical relations, *Language Learning* 29:327-344.

- 1987 Introduction to the use and acquisition of second language lexicon, *Studies in Second Language Acquisition* 9:129-132.
- 1988a Integrating research areas: A framework for second language studies, *Applied Linguistics* 9:198-217.
- 1988b Second language vocabulary acquisition, *Annual Review of Applied Linguistics* 9:92-106.
- 1993 Second language acquisition: past, present and future, *Second Language Research* 9:99-117.

GEERAERTS, D

- 1989 Prospects and problems of prototype theory, *Linguistics* 27:587-612.
- 1997 *Diachronic Prototype Semantics. A Contribution to Historical Lexicology*. Oxford, Clarendon Press.

GERGANOV, E & K RANGELOVA

- 1993 Effects of some motivational factors on foreign language vocabulary acquisition, B Kettermann & W Wieden (eds.) *Current Issues in European Second Language Acquisition Research*, Tübingen, Gunter Narr, 243-251.

GERGELY, G & C PLÉH

- 1994 Lexical processing in an agglutinative language and the organization of the lexicon, *Folia Linguistica* XXVIII/1-2:176-204.

GERRIG, R J

- 1986 Process and products of lexical access, *Language and Cognitive Processes*, 1:187 195.

GIBBS, R W

- 1996 What's cognitive about cognitive linguistics? G H Casad (ed), *Cognitive Linguistics in the Redwoods. The Expansion of a New Paradigm in Linguistics*, Berlin, Mouton de Gruyter, 27-53.

GIVÓN, T

- 1979 *On Understanding Grammar*, New-York: Academic Press.
- 1982 Logic vs. pragmatics, with natural language as a referee: towards an empirically viable epistemology, *Journal of Pragmatics* 6: 81-133.
- 1984 *Syntax: A functional-typological Introduction*, vol 1. Amsterdam, Benjamins.
- 1985 Function, Structure, and Language Acquisition, D I Slobin (ed.) *The Crosslinguistic Study of Language Acquisition*, Hillsdale, Lawrence Erlbaum, 1005-1027.

GOLDBERG, A E

1995 *Constructions. A Construction Grammar Approach to Argument Structure*, Chicago, The University Chicago Press.

GOMES DE MORAIS, A & A TEBEROSKY

1994 Erros e transgressões infantis na ortografia do português, *Discursos. Estudos de Língua e Cultura Portuguesa* 8:15-52, Lisboa, Universidade Aberta.

GONÇALVES, A

1992 *Para uma sintaxe dos verbos auxiliares em Português Europeu*, Dissertação de mestrado apresentada à Faculdade de Letras da Universidade de Lisboa.

GONÇALVES, P

1989 A variação do português dentro do Português, *RILP* 1: 15-20, Lisboa, Associação das Universidades de Língua Portuguesa.

1994 Uma hipótese sobre estratégias de aprendizagem do léxico do português L2 em Moçambique, *Actas do Congresso Internacional sobre o Português*, v.II:471-490, Lisboa, APL – Colibri.

1997 Tipologia de 'erros' do Português Oral de Maputo: um primeiro diagnóstico, C Stroud & P Gonçalves (org) *Panorama do Português Oral de Maputo, A construção de um banco de 'erros'* (vol II), Maputo, INDE, 37-70.

GONÇALVES, P & C STROUD (org.)

1998 *Panorama do Português Oral de Maputo. Estruturas Gramaticais do Português: problemas e exercícios*. (vol III), Maputo, INDE.

2000 *Panorama do Português Oral de Maputo. Vocabulário básico do Português (espaço, tempo e quantidade). Contextos e práticas pedagógicas*, (vol IV), Maputo, INDE.

GOUGENHEIM, G R & R MICHÉA & P RIVENC & A SAUVAGEOT

1956 *L'élaboration du français élémentaire*, Paris.

1964 *L'élaboration du Français Fondamental* (1er degré), Paris, Didier.

GOUIN, F

1892 *L'art d'enseigner et d'étudier des langues*. Paris.

GOULDEN, R & P NATION & J READ

1990 How large can a receptive vocabulary be? *Applied Linguistics* 11: 341-363.

GRABE, W & F L STOLLER

1997 Reading and vocabulary development in a second language: A case study, J Coady & T Huckin (eds.) *Second Language Vocabulary Acquisition*, Cambridge, CUP, 98-122.

GREEN, D W

1993 Towards a model of L2 comprehension and production, R Schreuder & B Weltens (eds.) *The Bilingual Lexicon*, Benjamins, Amsterdam, 249-277.

GREEN, D W & P MEARA

1987 The effects of script on visual search, *Second Language Research* 3:102-117.

GREENBERG, J H

1966 Some universals of grammar, with particular reference to the order of meaningful elements. J H Greenberg (ed.) *Universals of Language*, Cambridge, Mass: MIT Press, 73-113.

GRICE, H P

1975 Logic and Conversation, P Cole & J L Morgan (eds.), *Syntax and Semantics* 3, New York, Academic Press, 41-58.

GRIMSHAW, J

1990 *Argument Structure*, Cambridge Mass: MIT Press.

GROSJEAN, F

1989 Neurolinguistics, Beware! The bilingual is not two monolinguals in one person, *Brain and Language* 36:3-15.

GRUBER, J S

1965 *Studies in Lexical Relations*. Ph.D. diss., MIT.

HAASTRUP, K

1985 Lexical inferencing - a study of procedures in reception, *Scandinavian Working Papers in Bilingualism* 5:63-86.

1990 Developing learners' procedural knowledge in comprehension, R Phillipson, & E Kellerman & L Selinker & M Sharwood-Smith & M Swain (eds.) *Foreign / Second Language Pedagogy Research*, Clevedon, Multilingual Matters, 120-133.

HAKUTA, K

1976 A case study of a Japanese child learning English as a second language, *Language Learning* 26: 321-351.

HALLIDAY, M A K

- 1985 *An Introduction to Functional Grammar*. London,.Arnold.
- HANCIN-BHATT, B & W NAGY
 1994 Lexical transfer and second language morphological development, *Applied Psycholinguistics* 15: 289-310.
- HASUIKE, R & O TZENG & D HUNG
 1986 Script effects and cerebral lateralization: The case of chinese characters,J Vaid (ed.) *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*, Hillsdale, L Erlbaum Assoc., 275-288.
- HATCH, E
 1978 Acquisition of syntax in a second language, J Richards (ed.) *Understanding Second & Foreign Language Learning*, Rowley, MA: Newbury House, 34-70.
- HATCH, E & C BROWN
 1995 *Vocabulary, Semantics and Language Education*, Cambridge,CUP.
- HAZENBERG, S & J H HULSTIJN
 1996 Defining a minimal receptive second-language vocabulary for non-native university students: an empirical investigation, *Applied Linguistics* 17:145-163.
- HOCKETT, C E
 1954 Chinese versus English: An exploration of the Whorfian thesis. H Hoijer (ed.) *Language in Culture*, Chicago, University of Chicago Press.
- HUCKIN, T & M HAYNES & J COADY
 1993 (eds.) *Second Language Reading and Vocabulary Learning*, Norwood, NJ, Ablex.
- HUDSON, W
 1989 Semantic theory and L2 lexical development, S Gass & J Schachter (eds.) *Linguistic Perspectives on Second Language Acquisition*, Cambridge, CUP, 222-238.
- HUDSON, T
 1995 Nothing does not equal zero?, Problems with applying developmental sequence findings to assessment and pedagogy, *Studies in Second Language Acquisition* 15:495-503.
- HULSTIJN, J

1997 Mnemonic methods in foreign language vocabulary learning: theoretical considerations and pedagogical implications, J Coady & T Huckin (eds.) *Second Language Vocabulary Acquisition*, Cambridge, CUP, 203-224.

HUMBOLDT, W VON

1836 *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Königliche Akademie der Wissenschaften. Tradução: *On language: The diversity of human language structure and its influence on the mental development of mankind*. Cambridge: CUP, 1988.

HUMMEL, K M

1986 Memory for bilingual prose, J Vaid (ed.) *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*, LEA; Hillsdale, 47-64.

1993 Bilingual memory research: From storage to processing issues, *Applied Psycholinguistics* 14: 267-284.

HYAMS, N

1986 *Language Acquisition and the Theory of Parameters*, Dordrecht, Reidel Pub. Comp.

HYLTENSTAM, K

1977 Implicational patterns in interlanguage syntax variation, *Language Learning* 27:383-411.

1984 The use of typological markedness conditions as predictors in second language acquisition: the case of pronominal copies in relative clauses, R Andersen (ed) *Second Language. A Cross-linguistic Perspective*, Rowley, Newbury H P, 39-58.

1985 L2 learners' variable output and language teaching, Hyltenstam & M Pienemann (eds.) *Modelling and Assessing Second Language Acquisition*, Clevedon, Multilingual Matters, 113-136.

1986 Markedness, language universals; language typology, and second language acquisition, C Pfaff (ed.) *First and Second Language Acquisition Processes*, Cambridge, Newbury House Publishers, 55-78.

HYMES, D

1972 On communicative competence. J B Pride & J Holmes (eds.) *Sociolinguistics*, New York, Penguin, 269-293.

IJAZ, H

1986 Linguistic and cognitive determinants of lexical acquisition in a second language, *Language Learning* 36:401-451.

- ILARI, R
 1996 A categoria advérbio na gramática do português falado, *Actas do Congresso Internacional sobre o Português*, vol. I, Lisboa, APL - Edições Colibri, 107-139.
- IRIARTE SANROMÁN, A
 1996 Co-ocorrência léxica no dicionário de Espanhol - Português, *Actas do XII Encontro da APL*, Braga / Guimarães, APL, 149-156.
- JACKENDOFF, R
 1972 *Semantic Interpretation in Generative Grammar*, Cambridge, Mass, MIT Press.
 1983 *Semantics and Cognition*, Cambridge, MA:MIT Press.
 1990 *Semantic Structures*, Cambridge, Mass., MIT Press.
 1991 Parts and boundaries, *Cognition* 41: 9-45.
- JACOBS, B
 1988 Neurobiological differentiation of primary and secondary language acquisition, *Studies in Second Language Acquisition* 10:303-337.
- JACOBS, B & J SCHUMANN
 1992 Language acquisition and neurosciences: Towards a more integrative perspective, *Applied Linguistics* 13:282-301.
- JAGTMAN, M & T BONGAERTS
 1994 COMOLA: A computer system for the analysis of interlanguage data, *Second Language Research* 10:49-83.
- JAKOBSON, R
 1941 *Kindersprache, Aphasie, und allgemeine Lautgesetze*, Uppsala, Almqvist & Wiksell. (Tradução: *Child language, aphasia, and phonological universals*, The Hague, Mouton, 1968.
- JAMES, C
 1998 *Errors in Language Learning and Use. Exploring Error Analysis*, London, Longman.
- JOHNSON-LAIRD, P N
 1983 *Mental Models. Towards a Cognitive Science of Language, Inference and Consciousness*, Cambridge, CUP.
- JUFFS, A
 1996 *Learnability and the Lexicon. Theories and Second Language Acquisition Research*, Amsterdam, Benjamins. Comp.

KATZ, J J

1964 Analyticity and contradiction in natural language. J A Fodor & J J Katz (eds.) *The Structure of Language: Readings in the Philosophy of Language*, Englewood Cliffs, N J: Prentice-Hall.

1972 *Semantic Theory*, New York, Harper & Row.

1975 Logic and Language: an examination of recent criticisms of intensionalism. K Gunderson (ed.) *Language, Mind, and Knowledge, Studies in the Philosophy of Science*, Vol 7, Minneapolis: Univ of Minnesota Press.

1981 *Language and Other Abstract Objects*. Totowa, N J: Rowman & Littlefield.

KATZ, J J & J A FODOR

1963 Structure of a semantic theory, *Language* 39:170-210.

KAY, P

1997 *Words and the Grammar of Context*, Stanford, California, CSLT.

KEENAN, E L & B COMRIE

1977 Noun phrase accessibility and universal grammar, *Linguistic Inquiry* 8: 63-99.

KELLERMAN, E

1990 Compensatory strategies in second language research: a critique, a revision, and some (non) implications for the classroom, R Phillipson, & E Kellerman & L Selinker & M Sharwood-Smith & M Swain (eds.) *Foreign / Second Language Pedagogy Research*, Clevedon, Multilingual Matters, 142-161.

KEMPSON, R

1977 *Semantic Theory*, London : CUP.

KENNEDY, G

1998 *An Introduction to Corpus Linguistics*, London, Longman.

KESS, J F

1992 *Psycholinguistics: Psychology, Linguistics and the Study of Natural Language*, Amsterdam, Benjamins Publishers Comp.

KIRSNER, K & E LALOR & K HIRD

1993 The bilingual lexicon. Exercise, meaning and morphology, R Schreuder & B Weltens (eds.) *The Bilingual Lexicon*, Benjamins, Amsterdam, 249-277.

KLEIN, W & C PERDUE

1997 The basic variety (or: Couldn't natural languages be much simpler?), *Second Language Research* 13:301-347.

KODA, K

1996 L2 word recognition research: A critical review, *The Modern Language Journal* 80:450-460.

1997 Orthographic knowledge in L2 lexical processing: A cross-linguistic perspective, J Coady & T Huckin (eds.) *Second Language Vocabulary Acquisition*, Cambridge, CUP, 35-52.

KRASHEN, S

1981 *Principles and Practice in Second Language Acquisition*, New York: Prentice Hall.

1988 *Second Language Acquisition and Second Language Learning*, New York: Prentice Hall (1976-1980)

1994 The input hypothesis and its rivals, N Ellis(ed.) *Implicit and Explicit Learning of Languages*, London: Academic Press, 45-77.

KUMPF, L

1984 Temporal systems and universality in interlanguage: a case study Eckman & Nelson (eds.) *Universals of Second Language Acquisition*, Rowley, Newbury House, 132-143.

LABOV, W

1966 *The Social Stratification of English in New York City*. Center for Applied Linguistics, Washington D.C.

1969 Contraction, deletion, and inherent variability of the English copula, *Language* 45: 715-762.

1978 Denotational structure, D Farkas & W M Jacobsen & K W Todrys (eds) *Papers from the Parasession on the Lexicon*, Chicago, Chicago Ling. Society.

LADO, R

1957 *Linguistic Across Cultures*, Ann Arbor, Mich.

LAHIRI, A & W MARSLER-WILSON

1991 The mental representation of lexical form: A phonological approach to the recognition lexicon, *Cognition* 38: 243-294.

LAKOFF, G

1971 On generative semantics, Steinberg, D & L Jakobovits (orgs) *Semantics, An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*, Cambridge, CUP, 232-296.

1987 *Women, fire and dangerous things: What categories reveal about the mind*. Chicago, University of Chicago Press.

LAKOFF, G & M JOHNSON

- 1980 *Metaphors we live by*, Chicago, University of Chicago Press.
- LAMBERT, W E & J HAVELKA & C CROSBY
 1958 The influence of language-acquisition contexts on bilingualism, *Journal of Abnormal and Social Psychology* 102:981-983.
- LANGACKER, R W
 1987 *Foundations of Cognitive Grammar*, vol.I, *Theoretical Prerequisites*, Stanford, California, Stanford University Press.
 1988 A usage-based model, B Rudzka-Ostyn (ed.) *Topics in Cognitive Linguistics*, Amsterdam, J Benjamins, 127-161.
 1991 *Foundations of Cognitive Grammar*. Vol. 2: *Descriptive Application*. Stanford, Cal: Stanford University Press.
- LARSEN-FREEMAN, D & M LONG
 1991 *An Introduction to Second Language Acquisition Research*, London: Longman.
- LAUDANNA A & W BADECKER & A CARAMAZZA
 1992 Processing inflectional and derivational morphology, *Journal of Memory and Language*, 31:333-348.
- LAUFER, B
 1990a 'Sequence' and 'Order' in the development of L2 lexis: Some evidence from lexical confusions, *Applied Linguistics* 11:281-296.
 1990b Why are some words more difficult than others? Some intralexical factors that affect the learning of words, *International Review of Applied Linguistics* 28:293-305.
 1997a The lexical plight in second language reading: words you don't know, words you think you know, and words you can't guess, J Coady & T Huckin (eds.) *Second Language Vocabulary Acquisition*, Cambridge, CUP, 20-34.
 1997b What is in a word that makes it hard or easy: some intralexical factors that affect the learning of words, N Schmitt & M McCarthy (eds.) *Vocabulary. Description, Acquisition and Pedagogy*. Cambridge: CUP. 140-155.
- LAUFER, B & P NATION
 1995 Vocabulary size and use: Lexical richness in L2 written production, *Applied Linguistics* 16:307-322.
- LAWSON, M J & D HOGBEN
 1996 The vocabulary learning strategies of foreign language students, *Language Learning* 46:101-135.
- LEIRIA, I

- 1989 O ensino de Português a estrangeiros na Faculdade de Letras de Lisboa, *Seara Nova* 21:19-20. Reimpresso em *RILP – Revista Internacional de Língua Portuguesa* 3: 50-53, 1990.
- 1991 *A aquisição por falantes de Português Europeu, língua não materna, dos aspectos verbais expressos pelos Pretéritos Perfeito e Imperfeito, Tese de Mestrado, FLUL.*
- 1998a Professores de Português para o ano 2000, *Palavras* (Revista da Associação de Professores de Português), Lisboa, APP, 12: 77-84.
- 1998b Falemos antes de 'verdadeiros amigos' P F Pinto & N Júdice (eds), *Para acabar de vez com Tordesilhas*, Lisboa, Edições Colibri & APP & S. I. de Português Língua Estrangeira, 11-30.
- 1999 Português língua segunda e língua estrangeira: investigação e ensino, apresentado ao 1º Congresso de Português língua não materna, Lisboa 21-23 de Outubro 1999; *Idiomático* (Revista do Centro Virtual Camões) .

LENGYEL, Z

- 1995 Some critical remarks on the phonological component, D Singleton & Z Lengyel (eds.) *The Age Factor in Second Language Acquisition*, Clevedon, Multilingual Matter, 124-134.

LENNON, P

- 1991 Error: some problems of definition, identification and distinction, *Applied Linguistics* 12:180-196.

LEVELT, W J M

- 1989 *Speaking: From Intention to Articulation*, Cambridge, CUP.

LEVELT *et alii*

- 1991 The time course of lexical access in speech production: a study of picture naming, *Psychological Review* 98: 122-142.

LEVENSTON, E

- 1979 Second language acquisition: Issues and problems, *Interlanguage Studies Bull.* 4: 147-160.

LEVIN, B & S PINKER

- 1991 Introduction to the special issue of Cognition on Lexical and Conceptual Semantics, *Cognition* 41:1-7.

LEWIS, M

- 1993 *The Lexical Approach: The State of ELT and a Way Forward*. Hove: Language Teaching Publications. 1996.

1997a *Implementing the Lexical Approach*, Hove, Language Teaching Publications

- 1997b Pedagogical implications of the lexical approach, Coady & Huckin (eds.) *Second Language Vocabulary Acquisition*, Cambridge, CUP, Applied Linguistics, 255-270.

- LI, XIAOLONG
 1988 Effects of contextual cues in inferring and remembering meanings of new words, *Applied Linguistics* 9:402-413.
- LICERAS, J M
 1997 El saber sí ocupa lugar: La ciencia cognitiva y la adquisición del español como lengua segunda, *Revista de Filología Románica* 14, I: 299-318, Madrid, Univ. Complutense.
- LIGHTBOWN, P
 1985 Great expectations: Second language acquisition research and classroom teaching, *Applied Linguistics* 6:173-189.
- LINNARUD, M
 1975 Lexis in free production: An analysis of the lexical texture of swedish students' written work, *Swedish-English Contrastive Studies*, Report n°6, Lund University.
- LITTLE, D
 1994 Words and their properties: arguments for a lexical approach to pedagogical grammar. T Odlin (ed.) *Perspectives on Pedagogical Grammar*, Cambridge, CUP.
- LONG, M H
 1981 Input, interaction, and second-language acquisition, *New York Academy of Sciences Annals*. New York, 379: 259-278.
- LOPES, A C M
 1996 Semântica lexical e interpretação textual, *Actas do Congresso Internacional sobre o Português*, v.II:445-470, APL - Colibri.
 1996 *Então* - elementos para uma análise semântica e pragmática, *Actas do XII Encontro da APL*, vol.III:177-190.
- LOPES, C
 2003 *Inserção de a gente no quadro pronominal do português*, Madrid/Leipzig, Vervuert Verlag.
- LOPES, O
 1996 Partículas de agulhagem discursiva n' 'O Crime do Padre Amaro', *Actas do Congresso Internacional sobre o Português*, vol. I, Lisboa, APL - Edições Colibri, 189-202.
- LORD, R
 1974 Learning vocabulary, *International Review of Applied Linguistics* 12:239-247.

- LUKATELA, G & C CARELLO & M T TURVEY
 1987 Lexical representation of regular and irregular inflected nouns, *Language and Cognitive Processes*, 2:1-17.
- LUPPESCU, S & R DAY
 1995 Reading, dictionaries and vocabulary learning, B Harley (ed.) *Lexical Issues in Language*, Amsterdam, Language Learning/ J. Benjamins Pub., 229-251.
- LYONS, J
 1959 *Linguistic Semantics: An Introduction*. Cambridge, CUP
 1968 *Introduction to Theoretical Linguistics*, Cambridge, CUP.
- MACEDO, E
 1995 Análise de regências fixas (dar com, dar por, ...) a partir de listas de concordâncias, *Actas do XI Encontro da APL*, v III, Lisboa, APL, 373-382.
- MACWHINNEY, B
 1986 The Competition Model, McClelland & Rumelhart (eds.) *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*. Cambridge, Mass. MIT Press., 249-308.
- MACWHINNEY, B & C SNOW
 1985 The child language data exchange system, *Journal of Child Language* 12:271-296.
- MAIA, C A
 1986 *História do galego-português. Estudo linguístico da Galiza e do noroeste de Portugal desde o século XIII ao século XVI*, Coimbra, INIC.
- MAIRAL USÓN, R
 1999 El componente lexicón en la Gramática Funcional, C Butler & R Mairal & J Martín Arista & F J Ruiz de Mendoza (eds.) *Nuevas perspectivas en Gramática Funcional*, Barcelona, Editorial Ariel, 41-98.
- MARANTZ, A
 1995 The minimalist program, G Webelhuth (ed.), *Government and Binding Theory and the Minimalist Program*, Oxford, Blackwell, 349-382.
- MARQUILHAS, R
 1996 *A Faculdade das Letras. Leitura e escrita em Portugal no século XVII*, Dissertação de doutoramento apresentada à Universidade

de Lisboa (ed.: Lisboa, Imprensa Nacional-Casa da Moeda, 2000).

MARSLEN-WILSON, W & A NIX & G GASKELL

1995 Phonological variation in lexical access: Abstractness, inference and English assimilation, *Language and Cognitive Processes* 10: 285-308.

MARSLEN-WILSON, W & L TYLER

1980 The temporal structure of spoken language understanding, *Cognition* 8: 1-71

MARSLEN-WILSON, W & C BROWN & L K TYLER

1988 Lexical representations in spoken language comprehension, *Language and Cognitive Processes*, 3:1-16.

MARTÍN ARISTA, J

1999a La Gramática de Dik y las teorías funcionales del lenguaje, C Butler & R Mairal & J Martín Arista & F J Ruiz de Mendoza (eds.) *Nuevas perspectivas en Gramática Funcional*, Barcelona Editorial Ariel S A 13-39.

1999b Expresión y motivación en Gramática Funcional, C Butler & R Mairal & J Martín Arista & F J Ruiz de Mendoza (eds.) *Nuevas perspectivas en Gramática Funcional*, Barcelona Editorial Ariel S A 173-217.

MARTÍN MINGORANCE, L

1984 Lexical fields and stepwise lexical decomposition in a contrastive English-Spanish verb valency dictionary, R Hartmann (ed.), *LEX'eter' 83: Proceedings of the International Conference on Lexicography*, 226-236, Tübingen: Niemeyer.

1987 Pragmatics features in the lexicon of a Functional Grammar, *International Pragmatics Conference*, 17-22.

1995 Lexical logic and structural semantics: methodological underpinnings in the structuring of a lexical database for natural language processing, U Hoinkes (ed.) *Panorama der Lexikalischen Semantik*, 461-474, Tübingen: Gunter Narr.

MARTINET, A

1995 Functional Grammar, E F K Koerner & R E Asher (eds.) *Concise History of the Language Sciences. From the Sumerians to the Cognitivists*. Oxford, Pergamon, 290-295.

MARTÍNEZ ALMOYNA, J

1990 *Dicionário de Espanhol - Português*, Porto, Porto Editora, 2ª edição.

- MARTOHARDJONO, G & S FLYNN
 1995 Is there an age factor for universal grammar? D Singleton & Z Lengyel (eds.) *The Age Factor in Second Language Acquisition*, Clevedon, Multilingual Matters, 135-153.
- MATEUS, M H
 1994 *Syllable structure in Portuguese*. Universidade de Lisboa (ms.).
- MATEUS, M H & A BRITO & I DUARTE & I FARIA
 1983 *Gramática da Língua Portuguesa*, Lisboa, Caminho, 1989.
- MCCARTHY, M
 1984 A new look at vocabulary in EFL, *Applied Linguistics* 5:12-22.
- MCCARTHY, M & R CARTER
 1997 Written and spoken vocabulary, N Schmitt & M McCarthy (eds.), *Vocabulary. Description, Acquisition and Pedagogy*, Cambridge, CUP, 20-39.
- MCENERY, T & A WILSON
 1996 *Corpus Linguistics*, Edimburgo, Edinburgh Univ. Press.
- MCINTOSH, A
 1961 Patterns and ranges, *Language* 37:325-337.
- MCLAUGHLIN, B
 1978 The Monitor Model: some methodological considerations, *Language Learning* 28:309-332.
 1990 Restructuring, *Applied Linguistics* 11:113-128.
- MCLAUGHLIN, B & T ROSSMAN & B MCLEOD
 1983 Second language learning: an information-processing perspective. *Language Learning* 33:135-158.
- MEARA, P
 1980 Vocabulary acquisition: a neglected aspect of language learning, *Language Teaching and Linguistics, Abstracts* 13:221-246.
 1983 Word associations in a foreign language, *Nottingham Linguistics Circular* 11, 2:29-38.
 1993 The bilingual lexicon and the teaching of vocabulary, R Schreuder & B Weltens (eds.) *The Bilingual Lexicon*, Benjamins, Amsterdam, 279-297.
 1995 Editorial: single-subject studies of lexical acquisition, *Second Language Research* 11: n° 2:i-iii.
 1996 The classical research in L2 vocabulary acquisition, G Anderman & M Rogers (eds.) *Words, Words, Words. The*

- Translator and the Language Learner*, Clevedon, Multilingual Matters, 27-40.
- 1997 Towards a new approach to modelling vocabulary acquisition, N Schmitt & M McCarthy *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge, CUP, 109-121.
- MEARA, P & S INGLE
- 1986 The formal representation of words in an L2 Speaker's lexicon, *Second Language Research* 2:160-171.
- MEDEIROS DE CARVALHO, N
- 1991 Língua falada x Língua escrita. Abordagem do léxico, *Actas do VII Encontro da APL*, 49-63.
- MEISEL, J & H CLAHSN & M PIENEMANN
- 1981 On determining development stages in natural second language acquisition, *Studies in Second Language Acquisition* 3: 109-135.
- MELKA, F
- 1997 Receptive vs. productive aspects of vocabulary, N Schmitt & M McCarthy (eds.), *Vocabulary. Description, Acquisition and Pedagogy*, Cambridge, CUP, 84-102.
- MEYER, A S
- 1996 Lexical access in phrase and sentence production: results from picture - word interference experiments, *Journal of Memory and Language* 35:477-496.
- MILLER, G A
- 1956 The magical number seven, plus or minus two: some limits on our capacity of processing information. *Psychological Review* 63:81-97.
- MILLER, G & P JOHNSON-LAIRD
- 1976 *Language and Perception*, Cambridge Mass, Harvard Univ. Press.
- MOENS, M & M STEEDMAN
- 1988 Temporal Ontology and temporal reference, *Computational Linguistics*, 14, n° 2, 15-18.
- MONDRIA, J-A & M W -DE BOER
- 1991 The effects of contextual richness on the guessability and the retention of words in a foreign language, *Applied Linguistics* 12:249-267.
- MONTAGUE, R

1973 The Proper treatment of quantifiers in ordinary English, J Hintikka & J Moravcsik & P Suppes (eds.) *Approaches to Natural Language*, Dordrecht: Reidel. Reimpr. em R Thomason (ed.), *Formal Philosophy: Selected papers of Richard Montague*. New Haven: Yale UP, 1974.

MOON, R

1997 Vocabulary connections: multi-words items in English, N Schmitt & M McCarthy (eds.) *Vocabulary. Description, Acquisition and Pedagogy*. Cambridge: CUP. 40-63.

MOREIRA DA SILVA, J

1995 Considerações sobre movimento de ADV, *Diacrítica* 10, Braga, Universidade do Minho, 31-46.

MORENO, A & A TUZINE

1997 Distribuição social de variáveis linguísticas no Português Oral de Maputo, C Stroud & P Gonçalves (org) *Panorama do Português Oral de Maputo, A construção de um banco de 'erros'* (vol II), Maputo, INDE, 71-91.

MORTON, J

1979 Word Recognition, J Morton & J C Marshall (eds.) *Psycholinguistics Series*, vol.2 *Structures and Processes*, Londres, Elek.

MUYESKEN, P

1988 Are Creoles a special type of languages? F J Newmeyer (ed.) *Linguistics: The Cambridge Survey*, Cambridge, CUP, vol.II: 285-301.

NA, L & I S P NATION

1985 Factors affecting guessing vocabulary in context, *RELC Journal* 16:33-42. *Language Learning and Bilingualism*, Abo Akademi, 39-71.

NAGY, W

1997 On the role of context in first and second language vocabulary learning, N Schmitt & M McCarthy (eds.) *Vocabulary, Description, Acquisition and Pedagogy*, Cambridge, CUP, 64-83.

NASCIMENTO, M F B & J B GONÇALVES

1995 *Corpus de Referência do Português Contemporâneo (CRPC)*, *Actas do XI Encontro da APL*, vol. I, Lisboa, APL, 143-149.

NASCIMENTO, M F B & L A S PEREIRA

1995 Dicionários de combinatórias do português: associações lexicais frequentes observadas num *corpus* de português contemporâneo, *Actas do XI Encontro da APL*, vol. II, Lisboa, APL, 43-54.

NASCIMENTO, M F & P MARRAFA & L PEREIRA & R RIBEIRO & R VELOSO & L WITTMANN

1997 *Le-parole* – do corpus à modelização da informação lexical num sistema-multifunção, *Actas do XIII Encontro da Associação Portuguesa de Linguística*, 115-134.

PAWLEY, A & F SYDER

1983 Two puzzles for linguistic theory: nativelike selection and nativelike fluency, J C Richards & R Schmidt (eds.) *Language and Communication*, London, Longman.

NATION, P

1993a Vocabulary size, growth, and use, R Schreuder & B Weltens (eds.) *The Bilingual Lexicon*, Benjamins, Amsterdam, 115-134.

1993b Using dictionaries to estimate vocabulary size: essential, but rarely followed, procedures, *Language Testing* 10:27-40.

NATION, P & J NEWTON

1997 Teaching vocabulary, J Coady & T Huckin (eds.) *Second Language Vocabulary Acquisition*, Cambridge, CUP, 238-254.

NATTINGER, J R

1980 A lexical phrase grammar for ESL. *TESOL Quarterly* 14:337-344.

NATTINGER, J & J DECARRICO

1997 *Lexical Phrases in Language Teaching*. Oxford OUP (1992).

NELSON, G

1997 A study of the top 100 wordforms in ICE-GB text categories, *International Journal of Lexicography* 10:112-134.

NEMSER, W

1971 Approximative systems of foreign language learners, *International Review of Applied Linguistics* 9:115-123. Reimpresso em J Richards (ed.) 1974, *Error Analysis*, London: Longman, 55-63.

NESI, H & P MEARA

- 1994 Patterns of misinterpretation in the productive use of EFL dictionary definitions, *System* 22:1-15.
- NETO, P M
1995 Combinatórias lexicais num *corpus* linguístico especializado, *Actas do XI Encontro da APL*, I, Lisboa, APL, 359-366.
- NEWELL, A
1990 *Unified Theories of Cognition*, Cambridge, MA, Harvard University Press.
- NÓIA, M
1991 Para uma boa intercompreensão oral, *Actas do Seminário Internacional Português como Língua Estrangeira* (9-12 Maio 1991, Macau, Direcção dos Serviços de Educação, 395-409.
- NORMAN, J
1988 *Chinese*, Cambridge, CUP.
- NUNAN, D
1985 Content familiarity and the perception of textual relationship in second language reading, *RELC Journal* 16:43-51.
1991 Methods in second language classroom-oriented research. A critical review. *Studies in Second Language Acquisition* 13:249-274.
- O'DELL, F
1997 Incorporating vocabulary into the syllabus, N Schmitt & M McCarthy (eds.) *Vocabulary. Description, Acquisition and Pedagogy*. Cambridge: CUP. 258-278.
- ODLIN, T
1989 *Language Transfer. Cross-Linguistic Influence in Language Learning*, Cambridge, CUP, (5. Semantics: 71-84.
- OGDEN, C & I RICHARDS
1923 *The Meaning of Meaning*, London.
- OLDFIELD, R C
1966 Things, words and the brain. *Quarterly Journal of experimental Psychology*.18: 340-353.
- OLSHTAIN, E

1987 The acquisition of new word formation processes in second language acquisition, *Studies in Second Language Acquisition* 9:221-132.

OXFORD, R & R SCARCELLA

1994 Second language vocabulary learning among adults: State of the art in vocabulary instruction, *System* 22:231-243.

PALMBERG, R

1985 How much english vocabulary do swedish-speaking primary-school pupils know before starting to learn English at school?, H Ringbom, *Foreign Language Learning and Bilingualism*, Abo Akademi, 89-97.

1987 Patterns of vocabulary development in foreign-language learners, *Studies in Second Language Acquisition* 9:201-220.

PALMER, H E

1921 *The oral method of teaching language*, Cambridge, Heffer.

PARADIS, M

1985 On the representation of two languages in one brain, *Language Sciences*, 61:1-39.

1986 Foreword a J Vaid (ed.) *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*, Hillsdale, New Jersey, L Erlbaum, xi-xii.

1994 Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA, N Ellis (ed.) *Implicit and Explicit Learning of Languages*, Londres, Academic Press, 393-419.

1997 The cognitive neuropsychology of bilingualism, A de Groot & J Kroll (eds.) *Tutorials in Bilingualism. Psycholinguistic Perspectives*, New Jersey, L Erlbaum, 331-354.

1998 The other side of language: Pragmatic competence, M Paradis (ed.) *Pragmatics in Neurogenic Communication Disorders*, Oxford, Pergamon, 1-10.

PARIBAKHT, P S & M WESCHE

1993 The relation between reading comprehension and second language development in a comprehension based ESL program, *TESL Canada Journal* 11:9-29.

1997 Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition, J Coady & T Huckin (eds.) *Second Language Vocabulary Acquisition*, Cambridge, CUP, 174-200.

- PARK, S & T ARBUCKLE
 1977 Ideograms versus alphabets: Effects of script on memory in 'biscrptural' korean subjects, *Journal of Experimental Psychology: Human Learning and Memory* 3:631-642.
- PAYRATÓ, L
 1998 *Profesión, Linguista. Panorama de la Lingüística Aplicada*, Barcelona, Ed. Ariel.
- PERDUE, C (ed.)
 1993 *Adult Language Acquisition: Cross-linguistic Perspectives. 2, The Results*, Cambridge, CUP.
- PEREIRA, L A
 1994 Para um dicionário de combinatórias do português, *Actas do Congresso Internacional sobre o Português*, v III, Lisboa, APL - Colibri, 197-206.
- PERES, J A (ed.)
 1993 *Semântica, Discursos 4*, Lisboa, Universidade Aberta.
- PETERS, A M
 1985 Language segmentation: Operating principles for the perception and analysis of language, D Slobin (ed.) 1985a *The Crosslonguistic Study of Language Acquisition*, Hillsdale, L Erlbaum, 1029-1067.
- PHILLIPSON, R & E KELLERMAN & L SELINKER & M SHARWOOD-SMITH & M SWAIN (eds)
 1991 *Foreign / Second Language Pedagogy Research*, Clevedon, Multilingual Matters.
- PIENEMANN, M
 1985 Learnability and syllabus construction, K Hyltenstam & M Pienemann (eds.) *Modelling and Assessing Second Language Acquisition*, Clevedon, Multilingual Matters, 23-75.
 1986 Psychological constraints on the teachability of languages, C Pfaff (ed.) *First and Second Language Acquisition Processes*, Cambridge, Newbury House Publishers, 143-168.
 1989 Is language teachable? Psycholinguistic experiments and hypothesis, *Applied Linguistics* 10: 52-79.
 1992 COALA - A Computational system for interlanguage analysis, *Second Language Research* 8:59-92.
- PIENEMANN, M & M JOHNSTON & G BRINDLEY

- 1988 Constructing an acquisition -based procedure for second language assessment, *Studies in Second Language Acquisition* 10:217-243.
- PIENEMANN, M & M JOHNSTON & J MEISEL
 1994 The Multidimensional Model, Linguistic Profiling, and related issues. A Reply to Hudson, *Studies in Second Language Acquisition* 15:495-503.
- Português Fundamental
 1984, *Vocabulário e Gramática*, tomo I, *Vocabulário*, Lisboa, INIC-CLUL.
 1987, *Métodos e Documentos*, tomo 1e 2, Lisboa, INIC-CLUL.
- POULISSE, N
 1990 *The Use of Compensatory Strategies by Dutch Learners of English*, Berlin, Mouton De Gruyter.
 1993 A theoretical account of lexical communication strategies, R Schreuder & B Weltens (eds.) *The Bilingual Lexicon*, Benjamins, Amsterdam, 157-189.
 1997 Language production in bilinguals, A de Groot & J Krool (eds.) *Tutorials in Bilingualism. Psycholinguistic Perspectives*. New Jersey, L Erlbaum, 201-224.
- POULISSE, N & T BONGAERTS
 1994 First language use in second language production, *Applied Linguistics*, 15: 36-57.
- PRENDERGAST, T
 1846 *The mastery of languages, or the art of speaking foreign tongues idiomatically*, London, R Bentley.
- PRESTON, D R
 1989 *Sociolinguistics and Second Language Acquisition*, Oxford; Blackwell.
- PRINCE, P
 1996 Second language vocabulary learning: The role of context versus translations as a function of proficiency, *The Modern Language Journal* 80:478-493.
- PUSTEJOVSKY, J
 1991 The syntax of event structure, *Cognition* 41:47-81.
- RAMAT, P & D RICCA
 1994 Prototypical adverbs: on the scalarity/ radiality of the notion of adverb, *Rivista di Linguistica* 6:289-326.

READ, J

- 1993 The development of a new measure of L2 vocabulary knowledge, *Language Testing* 10:355-371.
1997 Vocabulary and testing, N Schmitt & M McCarthy (eds.) *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge, CUP, 303-320.

REY, A

- 1991 A propos de la sémantique lexicale, *Actes du Colloque de Lexicologie Française, Travaux de Linguistique* 23:179-193.

RICHARDS, J C

- 1974 (ed.) *Error Analysis. Perspectives on Second Language Acquisition*, New York, Longman.
1976 The role of vocabulary teaching, *TESOL Quarterly* 10:77-89.

RINGBOM, H

- 1985 The influence of Swedish on the English of Finnish learners, H Ringbom (ed.) *Foreign Language Learning and Bilingualism*, Abo Akademi, 39-71.
1990 Crosslinguistic lexical influence and foreign language learning, Phillipson, R & E Kellerman & L Selinker & M Sharwood-Smith & M Swain (eds.) *Foreign / Second Language Pedagogy Research*, Clevedon, Multilingual Matters, 172-182.

RIO-TORTO, M G

- 1994 Processos e paradigmas de formação de palavras em português, *Actas do Congresso Internacional sobre o Português*, v III, Lisboa, APL, 275-298.
1996 Sincronia, diacronia e análise genolexical, *Diacrítica* 11, Braga, Universidade do Minho, 227-244.

RIVENC, P

- 1979 Le français fondamental vingt-cinq ans après, *Le Français dans le monde* 148:15-22.
1987 Vocabulário frequente e vocabulário disponível, *Português Fundamental, Métodos e Documentos*, Lisboa, INIC, CLUL, tomo 2: 3-26.

RIVERS, W M

- 1983 *Communicating naturally in a second language. Theory and practice in language teaching*. New York, CUP.

ROGERS, M

- 1996 Beyond the dictionary: the translator, the L2 learner and the computer, G Anderman & M Rogers (eds.) *Words, Words*,

Words. The Translator and the Language Learner, Clevedon, Multilingual Matters, 69-95.

ROMAINE, S

1988 *Pidgin and Creole Languages*, London: Longman.

ROSCH, E

1973a Natural Categories, *Cognitive Psychology*, 328-350.

1973b On the internal structure of perceptual and semantic categories, T E Moore (ed.) *Cognitive Development and the Acquisition of Language*, New York, Academic Press, 11-144.

1975 Cognitive Representations of Semantic Categories, *Journal of Experimental Psychology*, 104:192-233.

RUBENSTEIN, H & L GARFIELD & J A MILLIKAN

1970 Homographic entries in the internal lexicon, *Journal of Verbal Learning and Verbal Behavior* 9:487-494.

RUMELHART, D E & J L MCCLELLAND

1986 On learning the past tenses of English verbs, J L McClelland & D E Rumelhart (eds.) *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*. Cambridge, Mass. MIT Press., 216-271.

RYAN, A

1997 Learning the orthographical form of L2 vocabulary - a receptive and productive process, N Schmitt & M McCarthy (eds.), *Vocabulary. Description, Acquisition and Pedagogy*, Cambridge, CUP, 181-198.

RYAN, A & P MEARA

1991 The case of the invisible vowels, *Reading in a Foreign Language* 7:531-540.

RYGALOFF, A

1973 *Grammaire élémentaire du chinois*, Paris, Presses Universitaires de France.

SALMONS, J

1993 The structure of the lexicon: evidence from German gender assignment, *Studies in Language* 17:411-435.

SANDRA, D

1994 The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective, *Language and Cognitive Processes* 9:227-269.

- SASANUMA, S & O FUGIMURA
 1972 An analysis of writing errors in Japanese aphasic patients: kanji and kana words, *Cortex* 8:265-282.
- SAUVEUR, L
 1874 *An introduction to the teaching of living languages without grammar or dictionary*, Boston, Schoenhof & Moeller.
- SCALISE, S
 1994 *Morfologia*, Bologna, Il Mulino.
- SCARAMUCCI, M R
 1997 A competência lexical de alunos universitários aprendendo a ler em inglês como língua estrangeira, *DELTA* 13, São Paulo, Unicamp, 215-246.
- SCHANK, R & R ABELSON
 1977 *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum.
- SCHMITT, N
 1997 Vocabulary learning strategies, N Schmitt & M McCarthy (eds.) *Vocabulary, Description, Acquisition and Pedagogy*, CUP, 199-227.
- SCHMITT, N & M MCCARTHY
 1997 Editors' comments, N Schmitt & M McCarthy (eds.) *Vocabulary, Description, Acquisition and Pedagogy*, Cambridge, CUP, 228-236 e 321-326.
- SCHMITT, N & P MEARA
 1997 Researching vocabulary through a word knowledge framework. Word associations and verbal suffixes, *Studies in Second Language Acquisition* 19:17-36.
- SCHREUDER, R & B WELTENS
 1993 The Bilingual Lexicon: An Overview, in Schreuder & Weltens, *The Bilingual Lexicon*, Amsterdam; J Benjamins, 1-10.
- SCHUMANN, J H
 1978a *The Pidginization Process. A Model of Second language Acquisition*, Rowley, MA: Newbury House.
 1978b Social and psychological factors in second language acquisition, J Richards (ed.), *Understanding Second and Foreign Language Learning*, Rowley, MA: Newbury House, 163-178.

- 1978c The relationship of pidginization, creolization and decreolization to second language acquisition, *Language Learning* 28:369-379.
- SCHWARTZ, B D & R A SPROUSE
1996 L2 cognitive states and the full transfer/ full access model, *Second Language Research* 12:40-72.
- SCIULLO, A & E WILLIAMS
1987 *On the Definition of Word*, Cambridge, MIT Press.
- SEGURA DA CRUZ, M L
1984 O vocabulário do Português Fundamental, *PF Métodos e Documentos* tomo 2, Lisboa, INIC, CLUL, 461-477.
- SEIDENBERG, M S & J L MCCLELLAND
1989 A Distributed, developmental model of word recognition and naming, *Psychological Review* 96:523-568.
- SELINKER, L
1972 Interlanguage, *International Review of Applied Linguistics* 10:209-231. Reimp. em J Richards (ed.) 1974, *Error Analysis*, London, Longman, 31-54.
- SERVICE, E
1992 Phonology, working memory, and foreign-language learning, *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 45: 21-50.
- SERVICE, E & V KOHONEN
1995 Is the relation between phonological memory and foreign language learning accounted for the vocabulary acquisition? *Applied Psycholinguistics* 16:155-172.
- SHARWOOD-SMITH
1988 Applied linguistics and the psychology of instruction: A case for transfusion?, W Rutherford & M Sharwood-Smith (eds.) *Grammar and Second Language Teaching: a Book of Readings*, Rowley, MA, Newbury House, 206-223.
- SILVA, A SOARES DA
1995a A gramática cognitiva. Apresentação e uma breve aplicação, *Diacrítica* 10:83-116.
1995b Fundamentos teóricos e contributos pedagógicos da Gramática Cognitiva, *Actas do XI Encontro da APL*, v III, Lisboa, APL, 401-412.

- 1995c Sobre a estrutura da variação lexical. Elementos da Lexicologia Cognitiva, *Actas do XI Encontro da APL*, v III, Lisboa, APL, 413-424.
- 1996a A mudança semântica como reorganização de protótipos. O verbo deixar, *Actas do XII Encontro da APL*, vol I:317-327.
- 1996b Dos conceitos lexicais aos conceitos gramaticais. Aspectos da gramaticalização, *Diacrítica* 11, Braga, Universidade do Minho, 113-138.
- 1999 *A Semântica de Deixar. Uma contribuição para a abordagem cognitiva em Semântica Lexical*. Lisboa, Fundação Calouste Gulbenkian e Ministério da Ciência e Tecnologia.

SILVA, J

- 1994 Sobre modais (+ MOV) cristalizados, *Diacrítica* 9, Braga, Universidade do Minho, 343-356.

SILVA, N MARTINS DA

- 1991 B. F. Skinner: Psicologia e Utopia, *Revista Portuguesa de Psicologia*, 23: 85-100.

SILVA-JOQUIM, C

- 1988 Referências espaciais – algumas considerações, *Homenagem a Lindley Cintra*, Encontro Regional da Associação Portuguesa de Linguística, Lisboa (a publicar).

SINCLAIR, J

- 1991 *Corpus, Concordance, Collocation*, Oxford, OUP.

SINCLAIR, J & A RENOUF

- 1988 A lexical syllabus for language learning, R Carter & M McCarthy (eds.) *Vocabulary and Language Teaching*, London, Longman, 140-160.

SINGLETON, D

- 1999 *Exploring the Second Language Mental Lexicon*, Cambridge, Cambridge Applied Linguistics.

SKEHAN, P

- 1998 *A Cognitive Approach to Language Learning*, Oxford OUP.

SKINNER, B F

- 1957 *Verbal Behavior*, New York, Appleton Century Crofts (Trad. port. 1978, São Paulo: Cultrix).

SLOBIN, D I

- 1973 Cognitive prerequisites for the development of grammar, in C A Ferguson & D I Slobin (eds.) *Studies of Child Language Development*, N. Y., Holt, Rinehart & Winston, 175-208.
- 1977 Language change in childhood and in history, J MacNamara (ed.) *Language Learning and Thought*, New York, Academic Press, 185-214.
- 1985a Introduction: Why study acquisition crosslinguistically?, D I Slobin (ed.) *The Crosslinguistic Study of Language Acquisition*, Hillsdale, Lawrence Erlbaum, 3-24.
- 1985b Crosslinguistic evidence for the language-making capacity, D I Slobin (ed.) *The Crosslinguistic Study of Language Acquisition*, Hillsdale, Lawrence Erlbaum, 1157-1256.
- 1990 Learning to think of speaking: native language, cognition and rhetorical style”, Comunicação apr. na *International Pragmatics Conference*, Barcelona.

SOARES, N V

- 1997 *Os sintagmas nominais simples. Para uma sintaxe comparada de Português, Francês e Inglês*, Tese de Mestrado apresentada à Faculdade de Letras da Universidade de Lisboa.

SÖDERMAN, T

- 1989 Word associations of foreign language learners and native speakers - a shift in response type and its relevance for a theory of lexical development, *Scandinavian Working Papers in Bilingualism* 8, Universidade de Estocolmo, 114-121.

SOKOLIC, M & M SMITH

- 1992 Assignment of gender to french nouns in primary and secondary language. A connectionist model, *Second Language Research* 8:39-58.

SÖKMEN, A

- 1997 Current trends in teaching second language vocabulary, N Schmitt & M McCarthy (eds.) *Vocabulary. Description, Acquisition and Pedagogy*, Cambridge, CUP, 237-257.

SPOLSKY, B

- 1988 Bridging the gap: a general theory of second language learning, *Tesol Quarterly* 22:377-396.
- 1989 *Conditions for Second Language Learning*, Oxford, OUP.

STAROSTA, S

- 1973 The faces of case, *Language Sciences* 25:1-14.

- STEINBERG, D D
1982 *Psycholinguistics. Language, Mind and World*, London, Longman.
- STEINBERG, D & L JAKABOVITS
1971 (orgs) *Semantics, An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*, Cambridge, CUP.
- STERN, H H
1987 *Fundamental Concepts of Language Teaching*, Oxford, OUP (1983)
- SUAREZ, A & P MEARA
1989 The effects of irregular orthography on the processing of words in a foreign language, *Reading in a Foreign Language* 6:349-356.
- SWAN, M
1997 The influence of the mother tongue on second language vocabulary acquisition and use, N Schmitt & M McCarthy (eds.) *Vocabulary. Description, Acquisition and Pedagogy*, Cambridge: CUP, 156-180.
- SWEET, H
1964 *The practical study of languages: A guide for teachers and learners*. London, OUP (1899).
- TALMY, L
1983 How language structures space, H Pick & L Acredolo (eds) *Spatial Orientation: theory, research and application*, New York, Plenum Press, 225-282.
1985 Lexicalization patterns: semantic structure in lexical forms, T Shopen (ed.) *Language Typology and syntactic description*. Vol III: *Grammatical categories and the lexicon*, Cambridge CUP, 57-149.
- TARONE, E
1983 On the variability of interlanguage systems, *Applied Linguistics* 4:142-163.
- TENNY, C
1994 *Aspectual roles and the syntax-semantics interface*, Dordrecht, Kluwer Academic Publishers.
- TEIXEIRA, J

1995 Vir, voltar, chegar, *Diacrítica* 10, Braga, Universidade do Minho, 131-136.

TEYSSIER, P

1980 *História da Língua Portuguesa*, tradução de Celso Cunha, Lisboa, Sá da Costa, 1982.

THOMPSON, S

1992 Functional Grammar, W Bright (ed.) *International Encyclopedia of Linguistics*, Oxford OUP vol. 2: 37-40.

THORNDIKE, E L

1924 The vocabularies of school pupils, J Carelton Bell (ed.) *Contributions to Education*, New York: World Book Co., 69-76.

TIMBERLAKE, A

1982 Invariance and the syntax of russian aspect, P J Hopper (ed.) *Tense-Aspect. Between Semantics and Pragmatics*, Amsterdam, J Benjamins, 305-331.

TOWELL, R & R HAWKINS

1994 *Approaches to Second Language Acquisition*, Clevedon, Multilingual Matters.

TZENG, O & D HUNG & W WANG

1977 Speech recoding in reading chinese characters, *Journal of Experimental Psychology: Human Learning and Memory* 3:621-630.

VAINIKKA, A & M YOUNG-SCHOLTEN

1996 Gradual development of L2 phrase structure, *Second Language Research* 12:7-39.

VAN ELS, T & T BONGAERTS & G EXTRA & C VON OS & A-M DIETEN

1984 *Applied Linguistics and the Learning and Teaching of Foreign Languages*, London, Edward Arnold (1977)

VANDELOISE, C (ed.)

1991 *Sémantique Cognitive*, *Communications* 53, Paris, Seuil.

VANPATTEN, B

1990 Theory and research in second language acquisition and foreign language learning: on producers and consumers, B VanPatten & J F Lee (eds.) *Second Language Acquisition and Foreign Language Learning*, Philadelphia, Clevedon, 17-26.

VIBERG, A

1984 The verbs of perception: a Typological study, *Linguistics* 21: 123-162.

1986 The study of crosslinguistic lexicology, O Dahl (ed.) *Papers from the Ninth Scandinavian Conference of Linguistics*, Universidade de Estocolmo, 313-326.

1993a Crosslinguistic perspectives on lexical organization and lexical progression, K Hyltenstam & A Viberg, *Progression and Regression in Language*, Cambridge, CUP, 340-385.

1993b The acquisition and development of Swedish as a first and a second language: The cause of clause combining and sentential connectors, B Kettermann & W Wieden (eds.) *Current Issues in European Second Language Acquisition Research*, Tübingen, G. Narr, 293-306.

1993c Vocabularies, I Ahlgren & K Hyltenstam (eds.) *Bilingualism in Deaf Education*, Hamburgo, Signum, 169-199.

VILLALVA, A

1994 *Estruturas morfológicas. Unidades e hierarquias nas palavras do português*, Dissertação de Doutorado apresentada à FLUL.

VILELA, M

1990 *Dicionário do Português Básico*, Porto, Edições Asa.

VOIONMAA, K

1993 *On the Semantics of Adult Verb Acquisition*, Dep. of Linguistics, Universidade de Gotemburgo.

WATERMAN S A

1996 Distinguished Usage, Boguraev B & J Pustejovsky (eds.) *Corpus Processing for Lexical Acquisition*, Cambridge, Mass, The MIT Press, 143-172.

WEINREICH, U

1974 *Languages in Contact. Findings and Problems*. Mouton, The Hague (1953)

WEST, M

1930 Speaking-vocabulary in a foreign language. *Modern Language Journal* 14:509-521.

1953 *A general service list of English words*. London, Longman, Green and Comp.

WHITE, L

- 1985 The prodrop parameter in adult second language acquisition, *Language Learning* 35: 47-62.
- 1987 Against comprehensible input: The input hypothesis and the development of L2 competence, *Applied Linguistics* 8:95-100.
- 1989 *Universal Grammar and Second Language Acquisition*, Amsterdam: J Benjamins.

WIDDOWSON, H G

1978 *Teaching Language as Communication*, Oxford; Oxford University Press.

1989 Knowledge of language and ability for use, *Applied Linguistics* 10:128-137.

WIERZBICKA, A

1988 *The Semantics of Grammar*, Amsterdam, J Benjamins.

WILKINS, D

1972 *Linguistics in Language Teaching*, London, Edward Arnold

1976 *Notional Syllabuses*, Oxford, OUP.

WILLIS, D

1990 *The Lexical Syllabus*, London, Collins-COBUILD.

WONG, S C

1988 What we do and don't know about chinese learners of English: A critical review of selected research, *RELC Journal* 19:1-20.

XAVIER, M F & M H MATEUS (orgs.)

1990 *Dicionário de Termos Linguísticos*, Lisboa, Associação Portuguesa de Linguística, Instituto de Linguística Teórica e Computacional, Edições Cosmos, 2 vols.

YANG, L

1997 Tracking the acquisition of L2 vocabulary: the Keki Language experiment, J Coady & T Huckin (eds.) *Second Language Vocabulary Acquisition*, Cambridge, CUP, 125-156.

YULE, G & E TARONE

1990 The other side of page: integrating the study of communication strategies and negotiated input in SLA, R Phillipson, & E Kellerman & L Selinker & M Sharwood-Smith & M Swain (eds.) *Foreign / Second Language Pedagogy Research*, Clevedon, Multilingual Matters, 162-171.

ZECHMEISTER, E B & C D'ANNA & J W HALL & C H PAUS & J A SMITH

1993 Metacognitive and other knowledge about the mental lexicon: Do we know how many words we know?, *Applied Linguistics* 14:188-206.

ZHOU, X & W MARSLER-WILSON

1994 Words, morphemes and syllables in the chinese mental lexicon, *Language and Cognitive Processes* 9: 393-422.

ZIMMERMANN, R

1997 Historical trends in second language vocabulary instruction, J Coady & T Huckin (eds.) *Second Language Vocabulary Acquisition*, Cambridge, CUP, 5-19.

ZOBL, H

1980 Developmental and transfer errors: their common bases and (possibly) differential effects on learning, *TESOL Quarterly* 14:469-483.

ZOBL, H & J LICERAS

1994 Functional categories and acquisition orders, *Language Learning* 44:159-180.