

ESTUDOS DE LINGUÍSTICA

VOLUME I

ANA R. LUÍS
COORD.



IMPRESA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

(Página deixada propositadamente em branco)



I N V E S T I G A Ç Ã O



Coordenação editorial

Imprensa da Universidade de Coimbra

Email: imprensauc@ci.uc.pt

URL: http://www.uc.pt/imprensa_uc

Vendas online: <http://www.livrariadaimprensa.com>

Concepção gráfica

António Barros

Capa

Vitrais da Sé Catedral de Vila Real, 2003, Portugal

De João Vieira (1934-2009), em forma de evocação e

homenagem da Universidade de Coimbra à Obra do Artista

Infografia

www.artipol.net

Execução gráfica

www.artipol.net

ISBN

978-989-26-0139-7

ISBN Digital

978-989-26-0231-8

DOI

<http://dx.doi.org/10.14195/978-989-26-0231-8>

Depósito legal

339286/12

Obra publicada com o apoio de:

FCT Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR Portugal

© Dezembro 2011, Imprensa da Universidade de Coimbra

ESTUDOS DE LINGUÍSTICA

VOLUME I

ANA R. LUÍS
COORD.

IMPRESA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

(Página deixada propositadamente em branco)

SUMÁRIO

Nota Prévia	7
-------------------	---

SECÇÃO 1: CRIOLÍSTICA

Coimbra University as an international research center in Contact Linguistics	11
John Holm	

Contributos para uma bibliografia anotada dos crioulos de base portuguesa e do Português vernáculo em África	25
Sandra Madeira	

SECÇÃO 2: LÉXICO

Aspectos da aprendizagem de colocações em Português L2.....	37
Ângela Costa	

Entrar de cabeça/sauter à pieds joints. Análise contrastiva de somatismos em Português e em Francês.....	49
Elise Malho	

SECÇÃO 3: MORFOLOGIA E SINTAXE

Variação da produtividade de sufixos agentivos em idade escolar.....	71
Ana Lúcia Almeida	

Padrões na aprendizagem da estrutura de concordância em Português Língua Estrangeira	89
Tânia Ferreira	

SECÇÃO 4: SEMÂNTICA

Oposições aspetuais em Português como Língua Segunda –

O caso dos falantes de Russo em contexto de imersão 101

Catarina Silva

O ensino do ‘Present Perfect’ a alunos falantes nativos

do Português Europeu 113

Filomena Ferreira

6

SECÇÃO 5: PRAGMÁTICA

A argumentação oral formal em contexto de sala de aula 133

Carla Marques

Refutação e coerência pragmática 149

Sara Sousa

SECÇÃO 6: PROCESSAMENTO AUTOMÁTICO DA LÍNGUA PORTUGUESA

Onto.PT: construção automática de uma ontologia lexical

para o Português 161

Hugo Oliveira e Paulo Gomes

Integração linguística em sistemas de conversão de grafema

para fone(ma) 181

Sara Candeias e Fernando Perdigão

NOTA PRÉVIA

O volume temático que agora se publica contém textos selecionados do 1º Encontro de Pós-Graduação em Linguística, realizado na Faculdade de Letras da Universidade de Coimbra, de 10 a 11 de Dezembro de 2010, sob a organização do Centro de Estudos de Linguística Geral e Aplicada. O volume inclui uma conferência plenária, proferida pelo Professor John Holm, e ainda onze textos, na sua maioria, versões desenvolvidas das comunicações apresentadas.

Todos os textos publicados foram submetidos a uma avaliação anónima prévia, tendo cada um sido avaliado por dois membros da Comissão Científica do Encontro de Pós-Graduação. Os autores receberam os pareceres anónimos e tiveram oportunidade de rever os seus textos. Em cumprimento da política editorial da Imprensa da Universidade de Coimbra, o volume foi posteriormente submetido a arbitragem científica anónima externa.

Agradecemos a todos os autores que se disponibilizaram para publicar os seus textos neste volume. Agradecemos ainda aos membros da Comissão Científica que colaboraram na avaliação das diversas contribuições escritas, bem como ao revisor anónimo.

Gostaríamos igualmente de deixar aqui os nossos agradecimentos a Isabel Lopes, Joana Vieira Santos e Paulo Nunes da Silva, pela colaboração, e às entidades que apoiaram a publicação do presente volume: a Fundação para a Ciência e a Tecnologia e a Imprensa da Universidade de Coimbra.

Coimbra, 3 de dezembro de 2011

Ana R. Luís

(Página deixada propositadamente em branco)

**SECÇÃO 1:
CRIOLÍSTICA**

(Página deixada propositadamente em branco)

John Holm

Faculdade de Letras da Universidade de Coimbra/CELGA

COIMBRA UNIVERSITY AS AN INTERNATIONAL RESEARCH CENTER IN CONTACT LINGUISTICS

1. Introduction

I would like to talk to you this morning about how Coimbra University has become a research center in contact linguistics with an international reputation¹. Because I am myself a contact linguist here at Coimbra University, this puts me in the awkward position of having to sing my own praises. It is not conventional to do this, and I'm sure you can imagine how reluctant I am. However, language being a social activity makes linguistics a social science, so sometimes linguists must grant themselves permission to suspend conventions when they need to examine certain social phenomena. For example, lexicographers who want to include taboo words in their studies may ignore those conventions that would prohibit the very pronunciation of these words, conventions that would force them to say something like "the F-word" instead of the word itself. That's fine unless it's unclear which "F-word" they have in mind, in which case they need to get to the point and say what they mean.

So I'm asking your permission to say directly what I mean, not least because the rise of research in contact linguistics to a position of international prominence at Coimbra University reveals some very interesting facts not only about the different individuals whose combined efforts made this happen, but also facts about how academia and the diffusion of knowledge work.

¹ Conferência de Abertura do Primeiro Encontro de Pós-Graduação em Linguística do Centro de Estudos de Linguística Geral e Aplicada, Faculdade de Letras da Universidade de Coimbra, 10 de dezembro 2010.

2. Coimbra creolists and JPCL

12 But first I think I should defend my claim that Coimbra University has attained a position of international prominence in research in contact linguistics. You only need to refer to the tables of contents of the last three issues of the leading journal in the field, the Journal of Pidgin and Creole Languages, published by John Benjamins twice a year in Amsterdam and Philadelphia. I don't think there could be any argument that the JPCL is not intended for an international readership, if only because it, like the Journal of Portuguese Linguistics, is published in English in a country where English is not the national language (although I would be among the first to recognize the usefulness of Portuguese in the international diffusion of knowledge). It should also be noted that the JPCL is a peer-reviewed scholarly journal, meaning that all the articles it publishes have to be reviewed and recommended for acceptance by at least three scholars whose expertise in their field of specialization within contact linguistics is fairly widely recognized. This is not a particularly pleasant process, and thanking anonymous reviewers for their helpful suggestions is most often done through gritted teeth. Still, it is the only way to prevent a scholarly journal from becoming the private fiefdom of an editor who can bestow or withhold publishing favors to his personal friends. Thus in its evaluation of the quality of the work published by the research institutes that it supports financially, the Portuguese Ministry of Education's Fundação para a Ciência e a Tecnologia is always interested in whether or not the journal in which a work is published has an international readership and is peer-reviewed.

The tables of contents to which I referred you indicate how many of the authors whose works appeared in the last three issues of JPCL belong to CELGA's Research Team in Creole Linguistics: more are from Coimbra than from any other single university. Of course I'm only talking about three issues of the journal, but they're the last three issues, and this means we're doing very well indeed in relation to our colleagues (and competitors) around the world.

To give credit where it's due, I should say a word about the articles and their authors. The 2009 article (JPCL, volume 24, number 2) was written by

Incanha Intumbo and me and is entitled “Quantifying superstrate and substrate influence.” It’s the main article of that issue, some 57 pages in length. It represents a new way of answering the perennial question of which languages were more influential in shaping the grammar of creoles: the European languages which supplied their vocabulary, or the African languages used by the earliest creole speakers? Intumbo is from Guinea-Bissau and has two first languages: Creole and Balanta (one of the country’s 22 African languages). We took some 100 grammatical structures that are often different in creole languages and their European lexical-source languages and compared their equivalents in Guinea-Bissau Creole, Balanta and Portuguese. We found that 11% of the structures are the same in Creole and Portuguese (but not Balanta); nearly 30% are the same in Creole and Balanta (but not Portuguese); but almost 33% are the same in Creole, Balanta and Portuguese. However, if we consider these three categories together—superstrate and substrate influence combined with convergence—they account for an overwhelming 74% of the Creole’s structures, resolving the perennially unanswerable question of where the grammar of creole languages comes from.

13

The next article (JPCL, volume 25, number 1, 2010) was written by Hugo Cardoso, a Coimbra student who studied with me for his licenciatura and became interested in creole linguistics, which he later pursued as a doctoral student at the University of Amsterdam. They sent him off to India to see if there were still any speakers of creole Portuguese in Diu. He found 125 of them and wrote an excellent dissertation about their speech, part of which formed the basis of this article, “The African slave population of Portuguese India: Demographics and impact on Indo-Portuguese.” He and his doctoral advisor, Umberto Ansaldo, have done much to revive scholarly interest in the creoles of Asia, particularly those based on Portuguese. First, they used the occasion of Cardoso’s 2009 doctoral defense, which brought together specialists quite literally from around the globe, to hold a workshop on the Portuguese-based creoles of Asia in Amsterdam, then they edited these papers for a special edition of the *Journal of Portuguese Linguistics* (volume 8, number 2). After Cardoso accepted a post-doc at the University of Macau and Ansaldo a professorship at the University of Hong Kong, the latter took on a guest editorship of this special issue of JPCL dedicated to pidgins and

creoles in Asian contexts (25:1). We feel very lucky that Hugo Cardoso has again chosen to make Coimbra his intellectual home by opting to join CELGA, our linguistic research institute.

14

The latest issue of JPCL in which Coimbra creolists cast a long shadow is volume 25, number 2, also published in 2010. The main article, 54 pages in length, is by Bart Jacobs, a scholar who has just joined the CELGA team as a researcher in contact linguistics. It is entitled “Upper Guinea Creole: Evidence in favor of a Santiago birth.” I realized how important this article is when I heard a leading linguist refer to it at a conference in Paris last summer even before it had been published. Jacobs has taken on and apparently resolved another perennial debate about the origin of two very similar creoles based on Portuguese, that of Cape Verde and that of Guinea-Bissau, which along with the creole of Casamance in Senegal are called the Upper Guinea Creoles. The similarity of their vocabularies and grammars allows speakers of these varieties to understand each other, unlike speakers of Gulf of Guinea creoles like Santomense. It has long been thought that the Upper Guinea creoles must have a common linguistic ancestor, but the question was whether this proto-creole first emerged in what is today Cape Verde or in Guinea-Bissau. By using newly uncovered linguistic data and reanalyzing established sociohistorical facts, Jacobs argues convincingly that the port of Cacheu in what is now Guinea-Bissau was founded and settled in the late 16th century by Cape Verdeans coming from the island of Santiago who brought their language with them. Jacobs came to work with me at the University of Coimbra from the Netherlands under the Erasmus program, where he had been working on his master’s thesis on Papiamentu, a Caribbean creole based on both Portuguese and Spanish. When he completed that, he won a full scholarship to do his Ph.D. at the University of Munich on a co-tutela with me at the University of Coimbra, which allows him to be a part of our linguistic research institute. You’ll note that he has published a review in the same issue of JPCL on a book on the African element in Cape Verdean.

I was also lucky enough to get a book that I co-edited reviewed in the same issue of JPCL: *Contact Languages: Critical concepts in language studies*, co-edited by Susanne Michaelis of the Max Planck Institute in Leipzig, Germany, which has been cooperating with teams from Coimbra’s CELGA

and many other research institutes in a massive project to produce the Atlas of Pidgin and Creole Structures, or APiCS. I say we were “lucky” to get our book reviewed not because our names are obscure in the field but rather because this set of five volumes has 2,300 pages and costs \$1,390. At that price the publisher was not eager to give away free copies to any reviewers, and getting them to supply even this single review copy was something of a triumph. In fact, the reviewer says that the set of books “...covers over two centuries of writing and theoretical thinking on language contact....and as such is a goldmine for students and researchers alike....The only real drawback of this anthology is its cost” (Mather 2010: 390). However, Routledge know their market, and the set has been selling well to university libraries in the Far East, which for historical reasons often did not acquire the books and journals from which most of the articles in this anthology are reprinted. In fact, one of the factors that motivated me to take on a project of this size—selecting the articles, introducing them to make clear the intellectual context in which they were written and the reasons they are of lasting importance—was the hope that it could provide a virtual library in contact linguistics for the many new universities and research institutes that are emerging in the former colonies where pidgin and creole languages are spoken today. Having taught at the College of the Bahamas for two years, I know that these are exactly the institutions in the most need but also those least likely to have the necessary funds for expensive academic books. My goal is to find a way match such institutions with stipends from non-governmental organizations such as UNESCO, and the publisher has agreed to supply the set to such institutions at cost.

15

3. Creole linguistics in Portugal

I trust that this look at the last three tables of contents of the Journal of Pidgin and Creole Languages will convince you that the University of Coimbra is one of the bright spots on the world map as seen by contact linguists. But how did this happen? Some twenty-five years ago when I was writing *Pidgins and Creoles* (Holm 1988-89), contact linguistics in Portugal

did not seem to be going anywhere. There had been a very promising start a century earlier when the publications of Coimbra's Adolfo Coelho had come to the attention of the German father of creole linguistics, Hugo Schuchardt, working at the University of Graz in Austria. Coelho had been collecting samples of Creole Portuguese spoken in colonies from Guinea to Macau, and Schuchardt appreciated their value as data for building linguistic theory. Silvio Sousa (2007), who traced the development of Portuguese linguistics during this period, notes that Schuchardt's interest inspired a number of speakers of not only Portuguese but also of Creole Portuguese to collect and publish more data on a number of varieties from Capeverdean (Brito 1887) to Indo-Portuguese (Dalgado 1900-1901). Despite the work on Portuguese-based creoles by linguists as distinguished as Leite de Vasconcellos (1901) and others, the interest in them that blossomed in the late 19th and early 20th century gradually waned, although they continued to receive the occasional attention of Portuguese dialectologists and missionaries. By the second half of the 20th century, the few Portuguese linguists who worked on these languages such as Graciete Nogueira Batalha and Jorge Morais Barbosa were largely outnumbered by non-Portuguese linguists such as France's Alain Kihm, who deplored the lack of scholarly attention the Portuguese-based creoles were receiving in his 1980 dissertation on the Creole of Guinea Bissau, concluding that "Les absents ont tort" (Kihm 1980:3), i.e. that those not working on the Portuguese-based creoles were simply wrong not to be doing so.

4. Becoming a creolist

I came to the study of the Portuguese-based creoles through the backdoor: I was interested in Black and Creole varieties based on the European languages I knew, which included Spanish. I was born in 1943 in the American Midwest, which was less linguistically isolated than it might have seemed because of immigration: my mother had been born in a German-speaking community in Russia (so she spoke German with her family after they immigrated to Michigan) and my father's family had immigrated from

Denmark (so he spoke Danish with his family) but my parents spoke English with one another and that became my mother tongue. My father had been a sailor as a young man and learned the value of languages in traveling. They both managed to get a university education in the 1930s despite the economic depression and they were both on the political left. I remember that we all started learning Spanish in the early 1950s because my parents were seriously considering emigrating to Uruguay: Americans on the political left were losing their jobs because of a fear of Communism led by Senator Joseph McCarthy. Soon, however, McCarthy was censured by the Senate, my father kept his job and we stayed in Michigan. Later my family drove to Mexico on a vacation to practice our Spanish and I had a wonderful time: it made me understand that even though it's a lot of work to learn another language, it's really worth the effort.

17

In the 1950s there was segregation or separation of the white and black races even in northern states like Michigan. My parents belonged to a political organization, the Fellowship of Reconciliation, which trained whites and blacks to cooperate to undermine the laws that kept public places like restaurants segregated. Not everyone in Jackson, Michigan believed that these anti-segregation tactics were a good thing, but eventually they succeeded in ending legal segregation. I went to an integrated elementary school and had black friends, but by the time we were in high school and started dating, teenage social groups had segregated themselves and young people realized that there were two different societies and we only belonged to one of them.

I studied German in high school and hoped to study in West Germany during my junior (or third) year of university, but I was also interested in other languages and cultures and decided to study Spanish and Russian in high school as well. When I graduated in 1961 I got a summer job as a cook at a tourist resort at the Grand Canyon out West in Arizona. I managed to get fired within three weeks. I was too embarrassed to go back home to Michigan in disgrace, so I decided to hitch-hike to California and then Mexico to try to get a job as an interpreter. A friend at the Canyon agreed to forward the letters I would continue writing to my parents that didn't mention this change of plans. With the \$85 I had earned at the Grand

Canyon I managed to get to Mexico but I didn't manage to get a job interpreting (Mexicans talk too fast) so I went on to Central America.

18 I ran out of money in Managua, Nicaragua and got a job as a gardener for some North American missionaries who eventually put me on an airplane back to the United States in time to start university. One of the most amazing things I discovered there was that black people from Nicaragua's Caribbean coast spoke what they called "Pirate English"—a Creole English that I could hardly understand.

At the University of Michigan in the 1960s a lot of students chose to ignore the social barriers between America's blacks and whites, supporting the desegregation of schools in the American South. The next summer I worked as a camp counselor in a Jewish summer camp outside New York City, which made me more aware of Jewish culture and ended my plans to study in Germany. I decided to study in Paris during my junior year instead, but I didn't have much time to learn French so (leaning on my knowledge of Spanish) I took the first and third semester courses and skipped the second and fourth, and hoped I could get by. I managed to get a job as a counselor at a summer camp in southern France in 1963, and by the end of the summer my French was fluent but bad.

I had a wonderful semester rather than year at the Sorbonne. I got too greedy for more languages and took an evening course in Italian in Paris and then left in March to travel on the cheap (hitch-hiking and youth hostels), spending one month in Vienna and then another in Italy. My father, an electrical engineer, was working in Pakistan's new capital, Islamabad, and I found there was a special bus for students going from London to New Delhi for only 100 pounds that I could take to visit my parents. They had a more realistic knowledge of traveling by land in that part of the world and hastened to assure me that there was no need for me to come, and they encouraged me to continue traveling in Europe instead. I went on to Greece, Turkey and the USSR, and finally visited my cousins in Denmark before returning to the United States for my last year of university. I had majored in English mainly because I liked to read, but as graduation approached in 1965, I realized I didn't actually know how to earn money. Some South American friends suggested that I apply for a job teaching

English as a foreign language at the University of the Andes in Bogotá, Colombia. I got the job and enjoyed it and finally became fluent in Spanish. I decided to go back to the United States and get a master's degree at New York City's Columbia University, where I eventually switched from English literature to Teaching English as a Foreign Language. This got me a job teaching in the German-speaking part of Switzerland, first in Zurich and later in the Alpine canton of Obwalden for a total of three years.

19

When I returned to Michigan I found a job teaching at the Detroit Institute of Technology, a small downtown college that had suddenly found itself without students after the 1967 race riots triggered an exodus of its white population to the suburbs. Its interim solution was to accept hundreds of students from India's Gujarat state who had not had English-medium instruction: I was to teach it to them. They also began recruiting students from among Detroit's black population, so I also taught standard English to students who spoke what is now called African American English. I became very interested in why its grammar was so different from other English dialects. Sociolinguists had begun to publish on American Black English, linking it historically to Caribbean varieties of English and African languages. I received a grant to pursue my study of African American English at the State University of New York in Cortland and eventually at the University of London. There I wrote my doctoral dissertation on the Creole English of Nicaragua's Caribbean coast, the variety which I had first encountered in 1961 and which I returned to Nicaragua to study in 1976. I used this creole to link certain grammatical features in American Black English to their counterparts in African languages.

I also became very interested in the problems that speakers of Black and Creole English had in learning the standard variety, which led me to a teaching job at the College of the Bahamas in Nassau. I realized that many grammatical problems were linked to differing subcategorizational rules in specific words, so I began to collect data on these words systematically on cards. When I ordered these alphabetically I realized I had the beginning of a dictionary that no one had ever written. I finally published the Dictionary of Bahamian English with Alison Shilling in 1982; we recently turned over the copyright to the College of the Bahamas so they could make it available

on line at www.dloc.com, and I'm happy to report that it's getting over one thousand hits a month.

20 In 1980 I accepted a job teaching English and linguistics at Hunter College of the City University of New York. There I taught students who were native speakers of other languages as well as those whose first language was Black or Creole English. I edited a book on Central American English (Holm 1983) and co-edited another on Caribbean English in general (Holm and Görlach 1986). Eventually I was promoted to full professor and began teaching in the Ph.D. Program in Linguistics at the City University of New York's Graduate Center, where I had a number of talented doctoral students from a number of different countries who wanted to write their dissertations on the many creole and semi-creole languages that immigrants had brought to New York City.

For my first sabbatical year (1986-87) I got a Fulbright scholarship to teach a course in creole linguistics at the University of London and to finish writing a two-volume survey called *Pidgins and Creoles* for Cambridge University Press (Holm 1988-89), which became something of a standard work. I had tried to use my knowledge of Spanish to read the works in Portuguese on the creoles based on that language, but I realized I wasn't understanding enough so I began to study Portuguese formally (New York City offers courses on a surprising number of languages, and I was also able to study Dutch, Haitian Creole, and Yoruba, a Nigerian language). A fieldtrip to Bahia in Brazil in 1983 got me interested in Afro-Brazilian varieties of Portuguese, and then in the summer of 1988 I came to Coimbra University to take a more advanced course in Portuguese for foreigners.

Meanwhile creole linguistics had taken off as an academic discipline in the United States as well as northern European countries such as France, Holland and Germany. I attended the first international colloquium on Portuguese-based creoles held in Portugal (University of Lisbon, 1991). I received another Fulbright scholarship to spend my second sabbatical year (1993-94) at the University of Coimbra, where I taught another course on creole linguistics and worked on two long-term book projects: one on the partial restructuring of languages (later published as Holm 2004) and another on comparative creole syntax (Holm and Patrick 2007).

5. Coimbra creolists

In 1998 I accepted the offer of a permanent position at the University of Coimbra as a *professor catedrático convidado* (I didn't understand all the implications of being *convidado* until it was too late, but later I was able to get rid of this part of my title by becoming *agregado*). Again I was lucky enough to start off with some intellectual stars among my students, particularly Liliana Inverno and Dominika Swolkien. The creation of the master's program in Descriptive Linguistics later revealed the talents of two creole speakers who are talented linguists, Carlos Fontes and Incanha Intumbo (for their accomplishments see our website, <http://www.uc.pt/creolistics>) as well as Hugo Cardoso and Bart Jacobs, whose work I discussed above. Christina Märzhäuser, on a co-tutela from the University of Munich, wrote a fascinating doctoral dissertation on Capeverdean lyrics in Lisbon rap music and Sandra Madeira is about to publish her master's thesis, Towards an annotated bibliography of restructured Portuguese in Africa. And we have some new stars in the making...

21

Of course this didn't happen over night. Coimbra's creolists have benefitted not only from the organization of appropriate graduate courses at this university but also from the growing interest in contact linguistics in Portugal. The creolists at Lisbon's universities (both the Clássica and the Nova) supported the creation of a learned society for the study of Portuguese-and Spanish-based creoles at the 1999 *Workshop sobre Crioulos de Base Lexical Portuguesa* to provide an accessible forum for creolists and their post-graduate students to present and debate their research. This led to the first meeting of the *Associação: Crioulos de Base Lexical Portuguesa e Espanhola* here at Coimbra University in June, 2001 in a joint meeting with the Society for Pidgin and Creole Linguistics (SPCL, based in the USA). So far we have met not only here in Coimbra (2001, 2006) but also the Universidade Clássica de Lisboa (2002), the University of A Coruña (2003), Curaçao (2004), the University of Orléans (2005), the University of Amsterdam (2007), the Universidade Nova de Lisboa (2008), the University of Cologne in Germany (2009), the Centre National de Recherche Scientifique in Paris (2010) and next July we will meet at the Universidade do Porto (2011). We try to have joint meetings with

societies like the SPCL , which met with us in Coimbra and Amsterdam as well as Curaçao, where we were also joined by the Society for Caribbean Linguistics, based in the Anglophone Caribbean. We also cooperate with the *Comité International des Études Créoles* (CIEC) in France and the *Associação Brasileira de Estudos Crioulos e Similares* (ABECS), which has invited the ACBLPE to meet with them in São Paulo in 2012. Our members regularly publish in their journals and we now have our own on-line journal, the *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* (RCBLPE). It is based at the University of Macau, while our website is based at the University of Orléans. The ACBLPE has four working languages (Portuguese, Spanish, French and English) in which members can present their work.

Coimbra creolists have also benefitted from the support of Coimbra's linguistic research center, CELGA, whose three main approaches to the study of Portuguese are based on diachrony, synchrony and contact. I have had the honor of coordinating the third group of researchers, and I am deeply proud of what my collaborators have accomplished thus far. I will retire in a couple of years, but it is my hope that the University of Coimbra won't give up its position as an outstanding international research center in contact linguistics.

6. References

- Brito, António de Paula (1887). *Dialectos crioulos-portugueses. Apontamentos para a gramática do crioulo que se fala na ilha de Santiago de Cabo Verde* [with introduction]. Republished in Jorge Morais Barbosa (ed.) (1967). *Estudos linguísticos Crioulos*. Lisboa: Academia Internacional de Cultura Portuguesa, pp. i-xx, 329-404. Translated from the Portuguese and Capeverdean by D. Swolkien, S. Gonçalves and J. Holm in J. Holm and S. Michaelis (eds.) (2009). *Contact Languages: Critical Concepts in Linguistics*. 5 volumes. London, New York: Routledge. vol. 1, 110-130.
- Dalgado, Sebastião Rodolfo (1900-1901). *Dialecto Indo-Português de Goa*. *Revista Lusitana* VI, 63-84.
- Holm, John (ed.) (1983). *Central American English*. Heidelberg: Julius Groos Verlag.
- Holm, John (1988-89). *Pidgins and Creoles*. Cambridge: Cambridge University Press, 2 vol.
- Holm, John (2004). *Languages in Contact: the partial restructuring of vernaculars*. Cambridge: Cambridge University Press.
- Holm, John and Alison Shilling (1982). *Dictionary of Babamian English*. Cold Spring NY: Lexik House.
- Holm, John and Manfred Görlach (eds.) (1986). *Focus on the Caribbean*. Amsterdam: John Benjamins.

- Holm, John and Peter L. Patrick (eds.) (2007). *Comparative Creole Syntax: Parallel outlines of 18 creole grammars*. London: Battlebridge.
- Kihm, Alain (1980). *Aspects d'une syntaxe historique: études sur le créole portugais de Guiné-Bissau*. Thèse de doctorat de 3e cycle. Université de Paris III, Sorbonne Nouvelle.
- Mather, Patrick-André (2010). Review of J. Holm and S. Michaelis (eds.) (2009). *Contact Languages: Critical concepts in language studies*. London and New York: Routledge, 5 vols. *Journal of Pidgin and Creole Languages* 25:2, 390-397.
- Sousa, Sílvio Moreira de (2007). *A Teoria Crioula de Adolfo Coelho segundo a correspondência com Hugo Schuchardt e Leite de Vasconcelos*. Master's thesis, University of Vienna.
- Vasconcellos, José Leite de (1901). *Esquisse d'une dialectologie portugaise*. Doctoral dissertation, U. of Paris. Paris/Lisbon: Aillaud & Cie.

(Página deixada propositadamente em branco)

CONTRIBUTOS PARA UMA BIBLIOGRAFIA ANOTADA DOS CRIoulos DE BASE PORTUGUESA E DO PORTUGUÊS VERNÁCULO EM ÁFRICA

1. Introdução

As bibliografias são um instrumento de pesquisa imprescindível. A necessidade de existir uma bibliografia completa na área das línguas crioulas e dos pidgins tornou-se cada vez mais evidente, levando à publicação, há mais de trinta anos, da bibliografia de Reinecke *et al.* (1975)², que, apesar de todo este tempo, continua a ser a referência de maior autoridade para obras e artigos publicados até à data. Infelizmente, tornou-se igualmente necessário atualizar urgentemente todas as suas secções devido ao rápido desenvolvimento da área desde então³.

Assim, a falta de referências mais recentes – particularmente, no caso dos crioulos de base portuguesa⁴ - levou Tomás (1992)⁵ a publicar a primeira bibliografia dos crioulos de base portuguesa na Ásia, o que evidenciou

² A data limite de publicação das obras referenciadas foi o final de 1971, o que representa um período de perto de 40 anos.

³ Na verdade, um ano depois, Baker e Stein apresentaram ‘A supplementary bibliography of French-based Indian Ocean creoles within the framework of *A Bibliography of Pidgin and Creole Languages*’ no *Journal of Creole Studies* 1(2), pp. 237-280 (1976) que pretendia atualizar e corrigir a bibliografia de Reinecke *et al.* (1975) na área dos crioulos de base francesa. Cerca de dez anos mais tarde, Valdman, Chaudenson e Hazaël-Massieux (1983) publicaram a primeira *Bibliographie des Etudes Créoles* cobrindo assim todas as variedades linguísticas reestruturadas de base francesa no mundo. A *Bibliographie des Etudes Créoles* de Hazaël-Massieux (1991) constitui uma versão aperfeiçoada, revista e atualizada e que se encontra digitalizada pelo Indiana University Creole Institute.

⁴ A coletânea de Morais-Barbosa (1967) foi, sem dúvida, uma das maiores e mais importantes referências nesta área. Inclui um panorama dos estudos crioulos portugueses na altura.

⁵ A data limite de publicação das obras referenciadas foi dezembro de 1989 (há cerca de 20 anos), salientando-se igualmente aqui a necessidade em atualizar esta obra.

mais ainda a necessidade atual de constituir uma bibliografia atualizada, anotada e razoavelmente completa dos crioulos de base portuguesa e do português vernáculo em África.

2. Descrição da bibliografia

26

2.1. Objetivo e organização

O objetivo da minha tese de mestrado foi precisamente o de disponibilizar tal instrumento para a comunidade científica, isto é, um inventário dos principais trabalhos publicados sobre os crioulos de base portuguesa em África – em particular, sobre os crioulos da Alta Guiné (o crioulo português (CP) da Guiné-Bissau ou *Kriyol*, o CP de Casamansa e as variedades cabo-verdianas de Sotavento e de Barlavento); os crioulos do Golfo da Guiné (nomeadamente, as variedades de São Tomé – o *santomense* ou *forro*, o angolano ou *Ngola* e o português dos Tonga - e o CP da ilha do Príncipe ou *principense*, bem como o crioulo português da ilha de Ano Bom ou *Fa d'Ambô*); e finalmente sobre o português parcialmente reestruturado (nomeadamente, o português vernáculo de Angola e o de Moçambique).

Procurou-se assim referenciar artigos, monografias, dissertações, teses e outros trabalhos académicos e científicos publicados sobre e em cada língua com anotações desde a sua primeira referência até hoje, incluindo descrições e discussões da língua em questão, transcrições de textos orais e textos escritos na língua que podem eventualmente servir para uma análise linguística. Foram igualmente introduzidos alguns materiais não publicados (principalmente, teses de mestrado e de doutoramento) pela sua pertinência e o seu interesse, bem como algum material ainda no prelo que merece desde já ser mencionado e algumas referências disponíveis on-line publicadas em revistas estáveis na Internet (*Creolica*, *Journal of Portuguese Linguistics*, *Journal of Pidgins and Creoles*, *Papia*, *Revista dos Crioulos de Base Lexical Portuguesa e Espanhola*, etc.).

2.2. Conteúdos da bibliografia

Os conteúdos desta bibliografia estão divididos em quatro partes. A primeira parte (“Portuguese-based creoles in Africa (General)”) inclui referências gerais cobrindo as variedades de base portuguesa em África. Trata-se de bibliografias, trabalhos coletivos, obras que abordam a expansão e os fenómenos de criouliização da língua portuguesa, bem como a influência recíproca do português e de outras línguas, publicações sobre várias variedades, obituários ou biografias de estudiosos e investigadores da área.

27

A segunda parte sobre os crioulos da Alta Guiné (“Upper Guinea Creoles”) está dividida em duas secções: Guiné-Bissau / Casamansa⁶ e ilhas de Cabo Verde. Quanto a esta subsecção, uma divisão dialetal do crioulo é tradicionalmente associada à divisão geográfica das ilhas em dois grupos, as ilhas de Sotavento a sul (Santiago, Fogo, Maio e Brava) e as de Barlavento a norte (Santo Antão, São Vicente, Santa Luzia, São Nicolau, Boa Vista e Sal). Contudo, as referências ao CP de Cabo Verde entraram todas numa única secção mas, sempre que possível, especificou-se na anotação qual o dialeto estudado⁷.

A terceira parte (“Gulf of Guinea Creoles”) incide sobre os crioulos do Golfo da Guiné e inclui cinco variedades de CP distribuídas por três áreas geográficas: São Tomé e Príncipe (aqui tratados separadamente – a ilha de São Tomé e a ilha do Príncipe – ao contrário da bibliografia de Reinecke *et al.* (eds.) (1975)) e a ilha de Ano Bom, que faz parte da República da Guiné Equatorial. Existem três variedades reestruturadas em São Tomé: o santomense, o angular e o português dos Tonga. Esta última subsecção foi incluída na bibliografia apesar de os poucos estudos existentes sugerirem que esta variedade tem uma génese e um desenvolvimento distintos.

⁶ Visto alguns autores (Rougé 2004) considerarem que as diferenças linguísticas não são muito significativas, decidiu-se tratar ambas as línguas na mesma secção (ao contrário do que acontece na bibliografia de Reinecke *et al.* (1975) em que a secção 18 (p.89-90) é dedicada à Guiné e a secção 19 (p.91) a Ziguinchor (Senegal)).

⁷ Pereira (2006) apresentou uma comunicação no VI Encontro da ACBLPE de 28 a 30 de junho de 2006, na Faculdade de Letras da Universidade de Coimbra, intitulada “Argumentos históricos e linguísticos contra a oposição entre crioulo de Barlavento e de Sotavento, em Cabo Verde”, em que argumenta contra esta distinção.

Por fim, a quarta parte (“Partially restructured Portuguese”) referencia material sobre as variedades parcialmente reestruturadas do português em Angola e em Moçambique. Estas variedades não crioualizadas do português vernáculo, que conservam parte da morfossintaxe da língua-fonte mas com um número significativo de traços de substrato e da interlíngua (Holm 2004), podem ser comparadas com as variedades do português não-padrão que se desenvolveram no Brasil, excetuando o facto de que, em África, ainda estão a ser influenciadas pelas suas línguas de substrato.

3. Alguns problemas encontrados na elaboração da bibliografia

3.1. Nomes para as diferentes secções e subsecções

Inicialmente, considerou-se que as secções deviam referir-se ao nome das variedades do CP ou do português vernáculo em questão, por exemplo, na secção sobre São Tomé e Príncipe, existiriam quatro subsecções: Santomense, Angolar, Português dos Tonga e Principense. Contudo, revelou-se ser difícil escolher um nome em particular devido às discussões à volta do politica, sociologica e linguisticamente correto. Por exemplo, em inglês, o termo CP (“Creole Portuguese”, isto é, Português crioulo) é muito útil para identificar o superstrato num levantamento linguístico, mas pode ser interpretado como referindo-se a uma variedade da língua portuguesa, o que implicaria uma visão neo-colonialista. Então, por que razão não utilizar o termo “Portuguese Creole” (ou seja, crioulo português), como sugerido por Philippe Maurer (comunicação pessoal)? Ou o nome local utilizados pelos seus falantes?

O quadro seguinte sugere quão variadas podem ser as escolhas no caso das variedades crioulas de base portuguesa em São Tomé e Príncipe.

Português	Inglês	Nome local
crioulo de São Tomé crioulo são-tomense / santomense	São Tomé CP (Holm, 2000) Santomense / Saotomense Santome (Hagemeijer, 2007)	forro / fôlô lungwa santome (Rougé, 2004)
crioulo dos Angolares angolar	Angolar CP Angolar (Lorenzino, 1998)	(lungwa) n'gola
principense	Príncipe CP	lun'gwiye lun'gie (Rougé, 2004)

Quadro 1

Por estas razões, decidiu-se usar os nomes geográficos, tal como na bibliografia de Reinecke *et al.* (1975). No entanto, o problema mantém-se sempre que tivemos de nos referir às diferentes variedades dentro de cada secção ou numa anotação em particular. Seria também possível utilizar o mesmo termo do que o autor da referência citada, correndo o risco de causar alguma confusão por estes termos poderem ser diferentes de autor para autor⁸.

3.2. A bibliografia de Reinecke *et al.* (1975): atualização e correções

A bibliografia que a autora elaborou deve obviamente muito a publicações anteriores, mais particularmente à bibliografia de Reinecke *et al.* (1975) mas, por razões práticas e teóricas, deve ter certos limites. Não pode ser uma mera compilação de trabalhos bibliográficos anteriores com uma atualização das referências que têm aparecido desde então. Em vez disso, considerámos ser mais útil rever estas obras, escolhendo as referências mais importantes que não podem ser omissas, e acrescentar novas entradas indispensáveis. Estas escolhas não foram fáceis se considerarmos a quantidade de artigos publicados por editoras com pouca divulgação, de difícil acesso, conhecidas apenas em

⁸ Ver também: Holm e Madeira (2009).

círculos restritos. Por exemplo, muitas missões publicaram alguns folhetos em pequenas edições para uma circulação limitada⁹.

30 Certas escolhas e mudanças foram necessárias. Uma delas, como já referido anteriormente, o facto de Reinecke *et al.* (1975) não ter qualquer secção sobre o português vernáculo de Angola e o de Moçambique. Por esse motivo, constituíram-se duas secções inteiramente novas (secções 4.1 e 4.2 respetivamente) nesta bibliografia tendo em conta que os estudos em ambas as áreas apenas se desenvolveram a partir do final dos anos 1980 ou no princípio dos anos 1990.

Por outro lado, aumentou-se substancialmente o número de referências cruzadas, já que eram limitadas em Reinecke *et al.* (1975); por exemplo, a entrada Morais-Barbosa (1966) aparece apenas uma vez na secção 17 (“Cape Verde Islands”):

MORAIS-BARBOSA, Jorge (or BARBOSA, Jorge de Morais). 1966. ‘Cabo Verde, Guiné e São Tomé e Príncipe. Situação linguística’, in *Cabo Verde, Guiné e São Tomé e Príncipe. Curso de extensão universitária. Ano letivo de 1965-66* [1966], p. 149-164. Also as an offprint, Lisboa: Universidade Técnica. 22p.¹⁰

Contudo, Lopes (1959) encontra-se na secção 2 (“Collective works”):

LOPES, Manuel, *et al.* (eds.). 1959. *Colóquios cabo-verdianos*. Lisboa. Xv, 182 p. (Junta de Investigações do Ultramar. Centro de Estudos Políticos e Sociais. Estudos de Ciências Políticas e Sociais, 22)¹¹

bem como na secção 17 (“Cape Verde Islands”):

LOPES, Manuel. 1959. ‘Reflexões sobre a literatura cabo-verdiana ou a literatura nos meios pequenos’, in *Colóquios cabo-verdianos*, p.1-22.¹²

⁹ É possível que a dispersão de todo este material tenha contribuído para que os crioulos de base portuguesa não tivessem sido devidamente estudados durante anos.

¹⁰ In Reinecke *et al.* (1975: 86), item 17-58.

¹¹ Idem, p. 21, item 2-58.

¹² In Reinecke *et al.* (1975: 86), item 17-51.

Na presente bibliografia, as referências cruzadas estão devidamente indicadas com abreviaturas no final da anotação que remetem para as secções em questão.

Outro problema encontrado em Reinecke *et al.* (1975) diz respeito aos itens¹³ não vistos por, pelo menos, um dos membros da equipa de compiladores e que estão marcados com o símbolo ° antes da data de publicação. Por exemplo:

BARRENA [MORINO], Natalio (C.M.F.). °n.d. [Diccionario español-annobonés y annobonés-español.]

Noted in Streit & Dindinger from *Almanaque 1923*; no further information.¹⁴

Como Reinecke declarou, o problema é o seguinte:

“Among unseen items included on the word of previous searchers there are doubtless some of doubtful pertinence; a few may even be nonexistent! Titles and other details of unseen items are subject to correction. Judging by many items seen after they had been copied from other bibliographies, the corrections will be numerous indeed.”¹⁵

Na sua introdução¹⁶, Reinecke também admite que muitas referências são muito raras, que algumas desapareceram completamente e que outras não são facilmente acessíveis aos leitores. Contudo, foi decidido guardar algumas referências indicadas em trabalhos anteriores, mesmo que a compiladora não tenha tido acesso a elas, por se considerar que algumas cópias possam ainda existir e serem localizadas.

¹³ 1/4 dos itens em todas as secções dedicados aos crioulos portugueses está marcado como não tendo sido visto. No entanto, na secção 17 sobre as ilhas de Cabo Verde, mais de 1/3 dos itens não foram vistos, o que é considerável.

¹⁴ In Reinecke *et al.* (1975: 94), item 21-2.

¹⁵ *Idem*, p.xxii.

¹⁶ *Ibidem*.

Alguns itens de menor valor ou de valor marginal foram incluídos em Reinecke *et al.* (1975) “Em parte para evitar ao utilizador deste livro ter de procurar no meio de muita verbosidade de qualidade inferior como os compiladores tiveram de o fazer”, para indicar “o interesse demonstrado na língua” (1975, p. xxii). Apesar de tudo, sentiu-se uma necessidade real em questionar a inclusão deste tipo de material bem como itens como o que se segue:

CARDOSO, Nuno Catharino. 1963. ‘O crioulo da ilha de Santo Antão de Cabo Verde’, in *Atas do 1º Congresso de Etnografia e Folclore promovido pela Câmara Municipal de Braga (de 22 a 25 de junho de 1956)*, 2:351-356 (Lisboa: Biblioteca Social Corporativa).

Nearly valueless; mainly a word list.¹⁷

Perante algumas escolhas mais difíceis, algum material teve de ser omitido: material da imprensa diária ou semanal; obras que forneçam dados históricos, demográficos e sociológicos do país em que a língua se desenvolveu são obviamente interessantes e essenciais para qualquer investigação na área das línguas crioulas e do português parcialmente reestruturado, dado serem demasiado numerosas para serem todas incluídas, o que alargaria igualmente o âmbito deste trabalho. De facto, poucas obras destas foram incluídas em Reinecke *et al.* (1975).

4. Conclusão

Em 2009, no âmbito da sua tese de Mestrado, a presente autora conduziu um primeiro trabalho de correção e atualização das referências publicadas até à data em e sobre os crioulos de base portuguesa em África, bem como o português vernáculo de Angola, o de Moçambique e o português

¹⁷ In Reinecke *et al.* (1975: 84), item 17-13. Não nos podemos esquecer que a avaliação deste item pode ter sido feito do ponto de vista de quem procura estabelecer o grau de reestruturação de uma variedade e não do ponto de vista de um lexicógrafo.

dos Tonga, intitulado *Towards a Bibliography of Restructured Portuguese in Africa*. Revelou-se ser um trabalho árduo. O próprio Reinecke acaba por admitir que: “Os estudiosos de uma área em particular encontrarão certamente aqui muitas lacunas para serem colmatadas.” (1975, p. xxii). De igual modo, a presente autora está consciente do facto de que, apesar de todos os seus esforços, um certo número de referências lhe tenham escapado. É, por isso, óbvio que mais pesquisas tinham e ainda têm de ser empreendidas não só para atualizar a presente bibliografia como também para melhorá-la, mesmo sabendo que uma bibliografia é, por natureza, um trabalho nunca acabado, o que levou a autora a disponibilizar o seu estudo no site da Associação dos Crioulos de Base Lexical Portuguesa e Espanhola (www.acblpe.org).

33

Com a autorização de Jean-Louis Rougé e com a ajuda de Emmanuel Schang e Hugo Cardoso, que gerem o site da ACBLPE sediado na Universidade de Orléans (França), foi possível colocar a bibliografia em questão on-line, tornando-a assim acessível aos membros da *Associação: Crioulos de Base Lexical Portuguesa e Espanhola* e da *Associação Brasileira de Estudos Crioulos e Similares* durante cerca de um ano até hoje, de modo a que estudiosos e investigadores em cada área pudessem fazer as correções necessárias e/ou acrescentar itens novos ou involuntariamente omissos. Infelizmente, apesar de o documento ter sido consultado mais de 740 vezes, nenhum comentário, sugestão ou contribuição foi apresentado.

Vimos esta participação no 1º Encontro de Pós-Graduação em Linguística do CELGA como mais uma oportunidade para apresentar e dar a conhecer este projeto à comunidade científica nacional e internacional que trabalha diretamente com estas variedades linguísticas de África, de modo a obter a sua ajuda na identificação e apreciação de todo o material pertinente a incluir ou já incluído nesta bibliografia.

A função primeira de uma bibliografia é “facilitar o trabalho moroso de recolha de referências ao estudioso que pretenda fazer investigação em qualquer área científica” (Tomás 1992: 10). Como Reinecke e os seus colegas o sentiram, este é, na verdade, um projeto que devemos empreender juntos.

5. Referências

- Baker, Philip e Peter Stein (1976). A supplementary bibliography of French-based Indian Ocean creoles within the framework of *A Bibliography of Pidgin and Creole Languages*. *Journal of Creole Studies* 1:2, 237-280.
- Hazaël-Massieux, Marie-Christine (1991). *Bibliographie des études créoles – Langues, cultures, sociétés*. Aix-en-Provence / Paris: Institut d'Etudes Créoles / Didier Erudition.
- Hagemeijer, Tjerk (2007). *Clause Structure in Santome*. Dissertação de doutoramento, Faculdade de Letras da Universidade de Lisboa.
- Holm, John (2000). *An Introduction to Pidgins and Creoles*. Cambridge: Cambridge University Press.
- Holm, John (2004). *Languages in Contact: The Partial Restructuring of Vernaculars*. Cambridge: Cambridge University Press.
- Holm, John e Sandra Madeira (2009). À propos des noms des variétés du portugais restructuré en Afrique. In: C. de Feral (org.). *Le nom des langues III. Le nom des langues en Afrique sub-saharienne: pratiques, dénominations, catégorisations. Naming Languages in Sub-Saharan Africa: Practices, Names, Categorizations*. Louvain-La-Neuve: Peeters / BCILL.
- Kihm, Alain (1994). *Kriyol syntax – The Portuguese-based Creole Language of Guinea-Bissau*. Amsterdam: John Benjamins.
- Lorenzino, Gerardo A. (1998). *The Angolar Creole Portuguese of São Tomé: Its Grammar and Sociolinguistic History*. München / Newcastle: LINCOM Europa.
- Madeira, Sandra (2009). *Towards a Bibliography of Restructured Portuguese in Africa*. Tese de Mestrado, Universidade de Coimbra.
- Morais-Barbosa, Jorge (org.) (1967). *Estudos linguísticos crioulos*. Reedição de artigos publicados no Boletim da Sociedade de Geografia de Lisboa. Lisboa: Academia Internacional de Cultura Portuguesa.
- Pereira, Dulce (2006). Argumentos históricos e linguísticos contra a oposição entre crioulo de Barlavento e de Sotavento, em Cabo Verde. Comunicação apresentada no VI Encontro da ACBLPE de 28 a 30 de junho de 2006, na Faculdade de Letras da Universidade de Coimbra.
- Reinecke, John *et al.* (eds.) (1975). *A Bibliography of Pidgin and Creole Languages*. Honolulu: The University of Hawaii Press.
- Rougé, Jean-Louis (2004) *Dictionnaire étymologique des créoles portugais d'Afrique*. Paris: Karthala.
- Valdman, Albert, Robert Chaudenson e Christine Hazaël-Massieux (1983). *Bibliographie des études créoles – Langues et littérature*. Bloomington: Indiana University Creole Institute.
- Tomás, Maria Isabel (1992). *Os crioulos portugueses do Oriente: uma bibliografia*. Macau: Instituto Cultural de Macau.

SECÇÃO 2:
LÉXICO

(Página deixada propositadamente em branco)

ASPECTOS DA APRENDIZAGEM DE COLOCAÇÕES EM PORTUGUÊS L2

1. Introdução

Segundo Pawley e Syder (1983), atingir uma proficiência numa Língua Segunda (L2)^{18/19} que se aproxime da de um falante nativo requer não só o conhecimento de um sistema de regras que gere um número infindável de enunciados, mas também o conhecimento de “sequências memorizadas” e de “frases lexicais”. Assim, do vasto e heterogéneo conjunto de unidades plurilexémicas fixas ou semifixas – fraseologias ou frasemas – que constituem o fraseoléxico (como os idiomatismos, as comparações estereotipadas, os provérbios, os pares fraseológicos, as construções com verbo-suporte, para

¹⁸ Este trabalho foi levado a cabo no âmbito da minha dissertação de Mestrado em Linguística e Ensino, área de especialização em Linguística Aplicada, apresentada à Faculdade de Letras da Universidade de Coimbra em 2009, sob a orientação das Professoras Doutoras Cristina dos Santos Pereira Martins e Maria Francisca Mendes Queiroz-Pinto de Athayde.

¹⁹ Língua Segunda (L2) é um termo que pode ser tomado com acepções distintas, mas, neste caso, será considerado equivalente a Língua Não Materna (LNM). As LNM's podem ainda subdividir-se em Língua Segunda (LS) e em Língua Estrangeira (LE). Genericamente, LS classifica “a aprendizagem e uso de uma língua não-nativa dentro de fronteiras territoriais em que ela tem uma função reconhecida” (por exemplo, habitantes de ex-colónias africanas que têm como L1 uma língua africana e como L2 Português) (Leiria 2004:1); o termo LE é usado para designar a aprendizagem (normalmente formal) e uso de uma língua em espaços onde essa língua não tem qualquer estatuto sociopolítico (por exemplo, o ensino do português como disciplina curricular na Alemanha). No entanto, no caso dos alunos tidos em consideração para esta investigação, tal distinção não se revela completamente satisfatória. Estes jovens aprendentes decidiram passar algum tempo num país estrangeiro, neste caso Portugal, e aprender português. Encontram-se, pois, numa situação de imersão, beneficiando, igualmente, do ensino formal da língua, factores que, combinados, tendem a acelerar o processo de aquisição. Uma vez que esta situação é um pouco *sui generis*, considerou-se ser mais adequado adoptar o termo L2 (Richards 1987: 7 *apud* Leiria 2004: 7).

mencionar apenas alguns²⁰), este trabalho debruçar-se-á sobre as chamadas colocações²¹.

38

A escolha recaiu sobre estas estruturas porque, apesar de ocorrerem com bastante frequência nas línguas, nem sempre são reconhecidas como combinatórias estáveis (com maior ou menor grau de fixidez estrutural e/ou lexical, que devem ser memorizadas de forma holística pelos aprendentes. Trata-se de entidades complexas, em regra, semanticamente transparentes (uma vez que o seu significado pode ser deduzido do significado dos constituintes que a compõem e da forma como estes se combinam), que não colocam, por isso, graves problemas de descodificação/compreensão a aprendentes de uma L2. A escolha de um colocativo por parte da base²² é, todavia, na generalidade dos casos, arbitrária/convencionalizada e semanticamente não-motivada, podendo causar dificuldades de codificação/produção e, nessa medida, levantar problemas no ensino-aprendizagem de línguas não-maternas.

Este trabalho deve, pois, entender-se como uma tentativa para encontrar possíveis respostas às seguintes questões: que mecanismos de apreensão são utilizados pelos alunos estrangeiros na aprendizagem de colocações? Serão estes mecanismos dependentes da LM dos aprendentes ou existem outros, independentes das LMs e transversalmente mobilizados por todos os alunos de Português Língua Segunda (PL2)?

Para procurar dar resposta a estas interrogações foi necessário enquadrar a questão da aprendizagem de colocações em PL2 à luz de contributos da Psicolinguística. Para tal, foi imperativa uma breve introdução ao conceito de interlíngua. Este termo foi cunhado por Selinker (1992 [1972]) para designar os sistemas linguísticos “intermédios” que os aprendentes tardios de uma L2 constroem enquanto, progressivamente, se aproximam dessa L2.

²⁰ Para uma tipologia do fenómeno fraseológico, cf., entre outros, Burger (2003) e Burger *et al* (2007).

²¹ Neste trabalho, tal como em Athayde (2011), definir-se-á “colocação”, de acordo com Grossmann e Tutin (2002: 3), como “a co-ocorrência lexical privilegiada de dois (ou mais) elementos linguísticos que entre si estabelecem uma relação sintáctica”.

²² Hausmann esclarece que o estatuto dos constituintes da colocação não é semelhante, existindo entre eles uma relação hipotáctica (ou de dependência). Este autor chama base (*Basis*) à palavra que determina a escolha do outro elemento co-ocorrente e colocativo (*Kollokator*) ao constituinte determinado.

Este sistema linguístico, independente da LM e da L2, é constituído através de estratégias que reflectem o recurso à respectiva LM, como a “transferência linguística”, e, por outro lado, por estratégias que serão independentes da LM do aprendente, como a “sobregeneralização”, a “transferência de instrução” e a “simplificação”.

2. Trabalho empírico

Para explicitar até que ponto a LM dos informantes influencia a aprendizagem das colocações do português, levei a cabo uma investigação com alunos de dois níveis de PL2 (A2 e C1), falantes de diferentes LMs: alemão, espanhol, inglês e italiano (cf. quadro 1). Para o meu estudo, e tendo em consideração uma análise feita aos vários programas de cada unidade curricular frequentada pelos sujeitos da amostra, procedi a uma divisão dos informantes, seguindo os critérios do Quadro[Q] Europeu[E] Comum[C] de Referência[R] para as Línguas: a saber, A2 e C1²³.

LM	A2	C1	Total
Alemão	3	19	22
Espanhol	13	12	25
Inglês	5	9	14
Italiano	15	10	25
	Total	Total	Total
	36	50	86

Quadro 1: Número de informantes em função da LM e do nível de Português

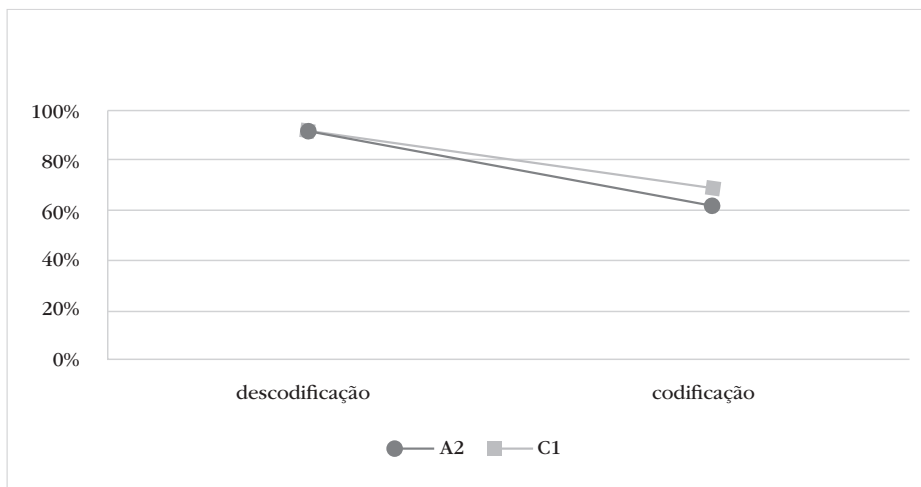
²³ Para uma caracterização mais exaustiva de cada um destes níveis do QECR, cf. a página http://sitio.dgidec.min-edu.pt/recursos/Lists/Repositio%20Recursos2/Attachments/724/Quadro_Europeu_total.pdf.

Estes alunos foram submetidos a um teste escrito com a duração de 1 hora, sendo o referido teste composto por duas partes: a primeira consistia na tradução de 20 colocações com uma estrutura sintáctica Verbo + Nome (por exemplo, *lavar os dentes*) e Nome + Adjectivo (por exemplo, *chuva torrencial*), do português à respectiva LM (descodificação); a segunda parte do teste requeria o exercício inverso (codificação).

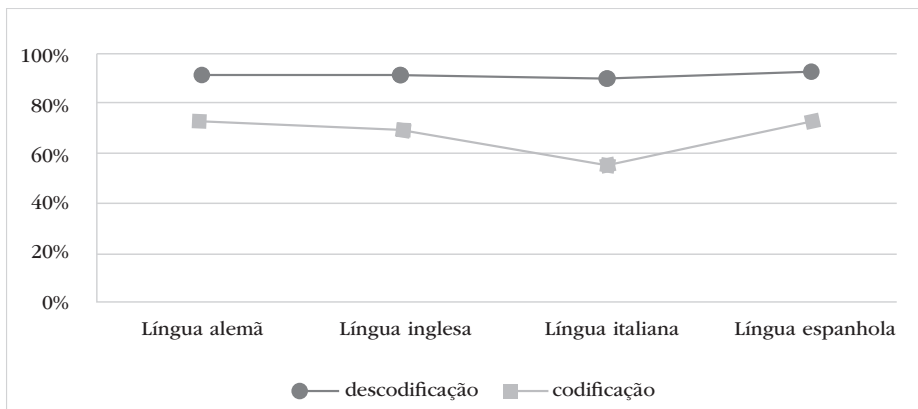
3. Resultados

Independentemente do nível de PL2 e da LM do aluno, os resultados da tarefa de descodificação revelaram-se sempre melhores do que os da tarefa de codificação (cf. quadros 2 e 3). Estes resultados demonstram bem a já mencionada dupla característica das colocações: a transparência semântica e a imprevisibilidade do ponto de vista da codificação.

Como podemos observar no quadro 2, os alunos do nível C1 obtêm melhores resultados nas duas tarefas propostas do que os alunos do nível A2. Estes dados demonstram que quanto maior for o grau de proficiência na L2 dos alunos melhores são os resultados no processamento de colocações, o que de resto seria inteiramente de esperar.



Quadro 2: Respostas correctas em função do nível de PL2 dos sujeitos



Quadro 3: Respostas correctas em função da LM dos informantes nas duas partes do teste

Foram os alunos de LM espanhola que obtiveram os melhores resultados, sustentando a hipótese de que falantes de línguas tipologicamente mais próximas podem mais facilmente capitalizar o conhecimento da LM na assimilação das colocações da L2. No entanto, os piores resultados obtidos pelos alunos de LM italiana parecem não apoiar esta interpretação. Foi, portanto, necessário realizar uma análise mais aprofundada da tipologia de respostas dadas pelos alunos de PL2.

3.1. Tipologia de fenómenos atestados nas respostas

Para um melhor enquadramento desta questão, proponho a seguinte tipologia de respostas:

Transferência da LM (de itens, de regras) – trata-se das respostas que se devem ao recurso à LM face a uma lacuna dos conhecimentos linguísticos neste domínio específico. Um tal mecanismo pode resultar positivamente, caso haja correspondência entre a LM e a L2, ou negativamente, caso não exista equivalência entre a LM e a L2 (ex: (ing) *strong coffee* – *café forte* (pt); *fare un esempio* (it) – (pt) **fazer um exemplo*, respectivamente).

Transferência de outra língua – trata-se de respostas atribuíveis ao recurso a outra língua conhecida do aprendente. Neste estudo foram inventariados poucos casos, uma vez que uma tal situação é bastante difícil de identificar. No entanto, podemos referir a tradução *sacar uma foto* para a colocação *tirar uma foto* por parte dos alunos de LM italiana.

Sobregeneralização do material linguístico da LA – Neste caso, a resposta resulta do facto de o aluno proceder à extensão de uma regra da L2 a contextos onde ela não se aplica. Exemplos deste comportamento são as respostas **tomar uma visita para fazer uma visita* ou **tomar a mesa para pôr a mesa*.

Transferência de instrução – trata-se da transferência de elementos que faziam parte do exercício. O aluno parte do princípio que será igual. Um exemplo de recurso a este mecanismo será o caso de alunos de LM espanhola que traduzem *café forte* por **café rigoroso*. Uma vez que a colocação *Inverno rigoroso* fazia parte da primeira parte do teste, poderá tal resposta ter servido de inspiração.

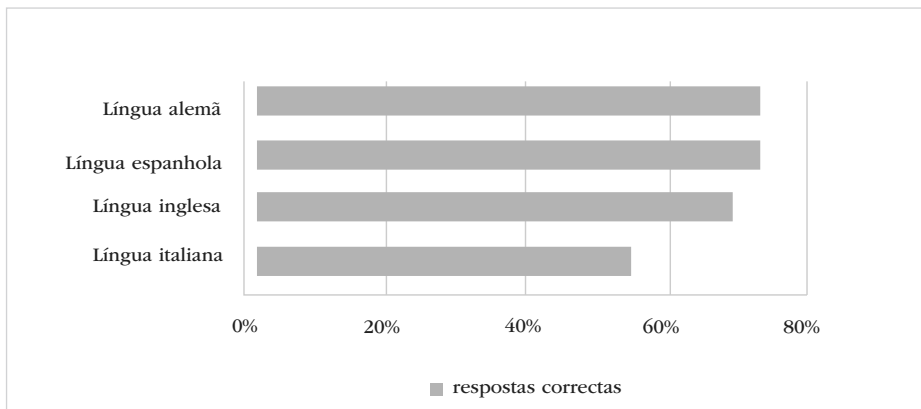
Simplificação – trata-se da redução da L2 a um sistema mais simples, tanto a nível sintáctico como semântico. Nesta análise, a evitação foi também considerada uma instância da estratégia de simplificação. Respostas como *não apanhar o comboio* para a tradução da colocação *perder o comboio* e *muito calor* para *alta temperatura* foram considerados exemplos de simplificação.

Outras respostas que não foram passíveis de classificar (*estar a vidro* (ir ao quadro) (ing); *apanhar o telefone* (alemão)).

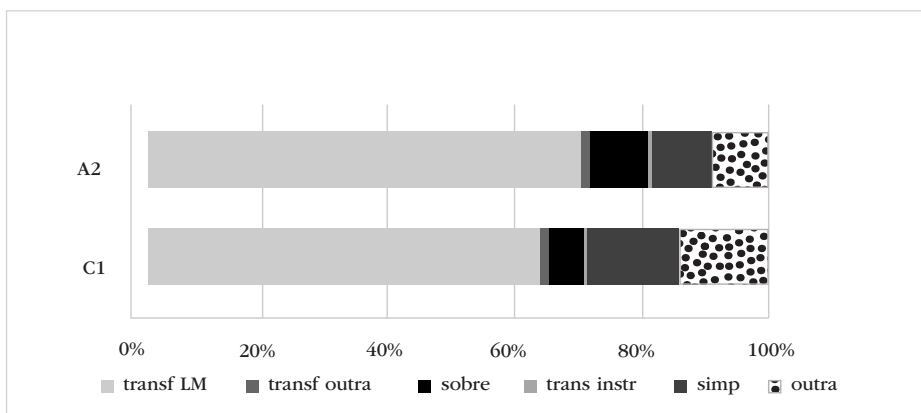
4. Discussão dos Resultados

Relembremos, então, o desempenho dos alunos das quatro LMs, no âmbito da tarefa de codificação proposta, e atentemos na tipologia de respostas dadas pelos alunos das quatro LMs diferentes.

Como podemos concluir da comparação dos gráficos 4 e 5, a transferência e o sucesso na realização das traduções parecem estar relacionados. Os alunos que obtêm os melhores resultados são os que mais confiam na LM e os alunos com piores resultados são, inversamente, os que menos confiam na LM. Valerá a pena ponderar sobre os motivos que levaram os



Quadro 4: Resultados da tarefa de codificação dos alunos das quatro LMs



Quadro 5: Tipologia de respostas dos alunos das quatro LMs na segunda parte do teste

alunos de LM espanhola a confiarem e os alunos italianos a não confiarem na respectiva LM, apesar de ambas as línguas serem consideradas tipologicamente mais próximas do português do que o inglês ou o alemão. Parece-me que uma explicação plausível para este comportamento poderá residir no facto de serem os alunos a decidir o que deve ou não ser transferido. Será, pois, a percepção que o aluno tem de quanto a L2 se

assemelha à sua LM que leva a que este transfira ou não uma estrutura. Os alunos de LM espanhola têm a percepção de que existe proximidade tipológica entre o português e o espanhol, partindo, então, de um pressuposto de equivalência entre as colocações deste par de línguas. Os alunos italianos, por sua vez, terão entendido as colocações como estruturas específicas da sua língua, idiossincráticas, que, provavelmente, não teriam equivalentes totais em português. Entendendo, portanto, que as colocações não seriam passíveis de uma simples transposição, estes aprendentes acabaram por apoiar-se mais em mecanismos de sobregeneralização e evitação do que os alunos de outras LMs. A sobregeneralização manifesta-se, sobretudo, no recurso ao verbo *tomar* – veja-se **tomar uma pausa*, **tomar amigos*, **tomar uma foto*, **tomar uma visita*, enquanto a evitação ter-se-á revelado pela opção de não resposta.

A transferência será, provavelmente, e como os dados apresentados neste estudo parecem indicar, o mecanismo mais poderoso no processo de aprendizagem das colocações, e, sobretudo, no da sua codificação na L2. Contudo, estes mesmos dados revelam o poder operante de outros mecanismos de aprendizagem, nomeadamente as sobregeneralizações e as simplificações, que deverão ser levadas em linha de conta no desenvolvimento de estratégias pedagógicas no âmbito do ensino destas combinatórias.

5. Conclusões

Tal como foi apurado, a estratégia mais frequentemente usada pelos alunos estrangeiros para colmatar as falhas na codificação de algumas colocações portuguesas é a transferência linguística. Também Granger (2001: 158) parece partilhar esta opinião, sublinhando que os aprendentes de uma L2 não são “phraseologically virgin territory: they have a whole stock of prefabs in their mother tongue which will inevitably play a role – both positive and negative – in the acquisition of prefabs in the L2”. Parecendo incontornável o papel da LM, vários autores, como Howarth (1998b: 14), sugerem que, tanto quanto possível, as colocações devem ser ensinadas através de uma metodologia contrastiva entre a LM a L2. Nesta medida,

uma perspectiva didática que confronte as colocações da LM dos aprendentes e as colocações correspondentes da L2, chamando a atenção para os casos em que existem quer semelhanças, quer diferenças, pode controlar a influência potencialmente negativa que a LM dos alunos poderá exercer no processo de assimilação destas estruturas.

Embora a comparação entre línguas seja uma das estratégias fundamentais para o ensino das colocações em contextos de sala de aula, será igualmente necessário chamar a atenção do aluno estrangeiro para a natureza fraseológica – isto é, uma natureza fixa ou semifixa – das colocações. Estas combinatórias lexicais estáveis passam frequentemente despercebidas aos olhos do aluno, não sendo sentidas como entidades fraseológicas. Não raras vezes, os professores de línguas centram a sua atenção nos dois extremos do espectro, i.e., nas combinatórias lexicais livres e nos idiomatismos ou provérbios, fazendo crer aos alunos que existem apenas dois modos distintos de combinação: a aplicação de regras morfossintáticas e semântico-lexicais em combinatórias livres e a completa “cristalização”, no caso das expressões fixas de natureza mais idiomática, dedicando menos atenção a fenómenos de transição como o são as colocações. Para uma melhoria na aprendizagem, é necessária uma categorização do fenómeno fraseológico bem como a compreensão, por parte dos alunos, de que a inaceitabilidade de um enunciado pode advir, não de uma quebra das regras gramaticais ou semânticas, mas também de um desvio nas regras fraseológicas, de natureza predominantemente lexical (Howarth 1998b: 3).

Sublinho ainda que, apesar de autores como Mackin (1978 *apud* Bahns 1993: 59) revelarem cepticismo em relação à possibilidade de se ensinarem colocações, dado o seu elevado número, outros investigadores (fundamentalmente, que se debruçam sobre a aprendizagem do inglês L2), como Channel (1981) e Morton (1977 *apud* Bahns 1993: 58), enfatizam a necessidade de estas serem, ainda assim, ensinadas, advogando que a mera exposição à língua-alvo não é suficiente para um aluno de nível avançado adquirir o conhecimento dos sintagmas convencionalizados. Será imperativo recordar que, visto estarmos perante estruturas aprendidas através da memória declarativa, a exposição e o uso reiterado são necessários para o sucesso da apreensão das colocações. Quanto mais vezes uma estrutura for

ouvida/lida e utilizada, mais facilmente será retida na memória de longo prazo declarativa. Noutros termos, quanto maior a exposição e uso destas estruturas, melhor será a proficiência do aluno (McCarthy 1984: 12).

Como nota final, gostaria de salientar que, não obstante este trabalho se ter proposto lançar luz sobre alguns aspectos da aquisição e da aprendizagem das colocações em contextos de português como L2, outras questões, a equacionar em futuros trabalhos, ficaram ainda por explorar. Refiro-me, nomeadamente, a um alargamento do estudo a alunos de outras LMs ou, ainda, a uma investigação mais exaustiva sobre colocações construídas em torno da mesma base ou do mesmo colocativo.

6. Referências

- Athayde, Maria Francisca (2011). Fazer uma pergunta, hacer una pregunta, poser une question, eine Frage stellen, to ask a question, fare una domanda – colocações em línguas europeias: a ilusão da semelhança?. In: Mingocho, M. Teresa, M. Fátima / Gil e M. Esmeralda Castendo (eds.). *Miscelânea de Estudos em Homenagem a Maria Manuela Gouveia Delille*. Coimbra: Minerva Coimbra.
- Bahns, J. (1993). Lexical Collocations: A Contrastive View. *ELTJ* 47/1, 56-63.
- Burger, Harald (2003). *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 2., überarbeitete Auflage. Berlin: Erich Schmidt Verlag.
- Burger, Harald, Dmitrij Dobrovolskij, Peter Kühn e Neal R. Norrick (eds.) (2007). *Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research*, Volume 1, 2. Berlin/New York: de Gruyter.
- Carter, Ronald e Michel MacCarthy (eds.) (1988). *Vocabulary and Language Teaching*. London: Longman.
- Channell, Joanna (1988). Psycholinguistic considerations in the study of L2 vocabulary acquisition. In: Carter, Ronald/MacCarthy, Michael (eds.), 83-96.
- Cowie, A.P. (ed.) (2001[1998]). *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press.
- Granger, Sylviane (2001 [1998]). Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. In: Cowie, A. P. (ed.), 145-160.
- Grossmann, Francis e Agnès Tutin (2002). Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif. *Revue Française de Linguistique Appliquée* 7, 7-25. (online: <http://w3.u-grenoble3.fr/tutin/collocations/RFLA.pdf>) (último acesso: 22 de junho de 2009).
- Hausmann, Franz Josef (1985). Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. In: Bergenholtz, Henning e Mugdan, Joachim (Hrsg.). *Lexikographie und Grammatik*, Tübingen: Max Niemeyer Verlag, 118-129.
- Howarth, Peter (1998). *Phraseology and Second Language Proficiency*. (online: http://www.leeds.ac.uk/languages/contact/people/howarth_docs/howarth1998b.doc) (último acesso: 1 de outubro de 2007).

- Howarth, Peter (2001 [1998]). The Phraseology of Learners' Academic Writing. In: Cowie, A. P. (ed.), 161-186.
- Leiria, Isabel (2004). Português língua segunda e língua estrangeira: investigação e ensino. *Idiomático. Revista Digital de Didática de PLNM*, nº 3. Centro Virtual Camões, 1-7. (online: <http://www.instituto-camoes.pt/cvc/idiomatico/03/portuguesLSeLE.pdf>) (último acesso: 22 de junho de 2009).
- McCarthy, Michael (1984). A new look at vocabulary in EFL. *Applied Linguistics* 5, 12-22.
- Pawley, A. e F. Syder (1983). Two Puzzles for linguistic theory: nativelike selection and nativelike fluency. In: J. Richards and R. Schmidt (eds.). *Language and Communication*. London: Longman.
- Quadro Europeu Comum de Referência para as Línguas* (2001). Lisboa: Edições Asa.
- Selinker, Larry (1972) La interlingua. In: Muñoz Licerias, Juana (ed.) (1992). *La adquisición de lenguas extranjeras*, Madrid: Visor, 79-101.

(Página deixada propositadamente em branco)

ENTRAR DE CABEÇA / SAUTER À PIEDS JOINTS: ANÁLISE CONTRASTIVA
DE SOMATISMOS EM PORTUGUÊS E EM FRANCÊS

1. Introdução

O corpo humano é fonte de inspiração para a linguagem humana, quer pelo simbolismo que carrega, quer pela sua função de principal instrumento de apreensão física e cognitiva da realidade²⁴. É através dos sentidos que conhecemos e apreendemos o mundo, para finalmente estarmos em condições de verbalizar experiências de vida, apetências, desejos, anseios, toda uma multiplicidade de facetas que caracterizam o ser humano.

Svensson (2004) dá conta, no seu trabalho, de *expressões fixas*, *unidades fraseológicas*, *fraseologias* ou *frasemas*²⁵, caracterizadas pelas particularidades que revelam de cada língua, bem como pelo colorido que imprimem à comunicação, dados os processos de transformação figural (metáfora, metonímia, sinédoque e eufemismo, entre outros) que estão na sua origem, bem como a sua mais-valia retórica.

Expressões do tipo *se sauver à toutes jambes*, *ter o sangue quente* não serão, com certeza, desconhecidas para os falantes que têm o Português e o Francês como línguas não-maternas. Trata-se, com efeito, de estruturas fixas ou semifixas – fraseologias – que fazem parte da comunicação quotidiana de todas as línguas, mas que podem passar despercebidas ao

²⁴ Texto elaborado no âmbito da Dissertação de Mestrado em Linguística e Ensino, especialidade em Linguística Aplicada, apresentada à FLUC em novembro de 2009, sob a orientação da Professora Doutora Maria Francisca Athayde e da Professora Doutora Graça Rio-Torto.

²⁵ Internacionalismos que se vão impondo na literatura específica, sendo que a utilização do último termo prevalece na última série HSK de Burger *et al.* (2007)

falante comum pelo grau de enraizamento que evidenciam na sua comunidade linguística.

50 Não é nosso objetivo aprofundar a definição das construções que se enquadram no âmbito do fenómeno fraseológico, bem como a diversidade do seu escopo e as dificuldades que se colocam aos trabalhos de investigação desenvolvidos nessa área²⁶. Procuraremos enunciar as principais propriedades das fraseologias, servindo esta breve incursão para refletir sobre as relações de “forma-significado” que se estabelecem entre os pares de expressões que serão objeto de descrição, com o objetivo de apurar divergências e convergências entre os somatismos²⁷ – doravante também designados SO – do par de línguas em contraste, num contributo para a procura de universais fraseológicos que atestam a crescente globalização das línguas e das culturas.

2. Propriedades das fraseologias

O tratamento das propriedades e da tipologia do fenómeno fraseológico²⁸ envolve, como se percebe, dois tópicos intimamente relacionados: um primeiro ligado aos traços definitórios das fraseologias – na sua relação com as unidades unverbais e as combinatórias não-fraseológicas – e um segundo dedicado ao enquadramento tipológico das fraseologias.

²⁶ Procurámos refletir, sempre que oportuno, o pensamento teórico de autores consagrados nesta matéria e que ajudaram a configurar o nosso trabalho, tais como Zuluaga (1980), Fleischer (1997), Larreta Zulategui (2001), Sanromán (2001), Burger (2003), Mellado Blanco (2004), Svensson (2004) e Burger *et al.* (2007).

²⁷ A designação ‘somatismo’ tem origem no Grego *soma*, que denominava o “corpo”. Consideramos que os SO são fraseologias que incluem, como componente, um lexema que designa uma parte da anatomia humana ou animal. Neste estudo, cingir-nos-emos, no entanto, a SO que se reportem a partes do corpo humano, uma vez que constituem, como facilmente se compreende, a maioria dos SO inventariados.

²⁸ A nossa opção pelo modelo de Burger (2003) foi motivada pelo facto de o linguista suíço propor uma conceção lata e flexível de Fraseologia enquanto subdisciplina da Lexicologia, tentando compatibilizar conhecimentos já consolidados nesta esfera de trabalho e aspetos da investigação mais recente, resultantes, entre outros, de diversos estudos no âmbito da Linguística do Corpus. Não obstante o suporte da língua alemã, os exemplos, as propriedades e as subclasses apresentadas de forma clara e sistemática pelo autor têm aplicabilidade na descrição do material fraseológico português e de outras línguas.

Devido à heterogeneidade do fenómeno fraseológico, as propriedades²⁹ elencadas por Burger (2003: 15-32) – *Polilexikalität, Festigkeit* e *Idiomatizität*, ou seja, a natureza pluriverbal, a fixidez e a idiomaticidade – deverão ser sobretudo vistas como traços que definem, com maior ou menor amplitude, diferentes subtipos de expressões (Athayde 2007: 40)³⁰.

Por outro lado, nem todas estas propriedades funcionam como traços definitórios das fraseologias em relação ao ‘discurso livre’ ou outras unidades do Léxico, sendo por isso relevante o contributo de Svensson (2004: 139) para a definição daquelas que são condições necessárias e suficientes, para podermos identificar combinatórias lexicais estáveis.

51

2.1. Natureza pluriverbal

É o traço mais evidente das fraseologias – constituídas por mais do que uma palavra –, não sendo, todavia, condição suficiente para a definição do nosso objeto de estudo (Svensson 2004: 142). Constitui somente um indício, uma vez que não distingue univocamente as fraseologias do ‘discurso livre’³¹.

Se a questão da fronteira mínima não parece suscitar dúvidas, o “limite máximo” tornou-se difícil de precisar, tendo Burger (2003: 15), incluído, neste quadro teórico, estruturas fráscicas ou textuais memorizadas pelos

²⁹ Fleischer (2007) veicula, a nosso ver, uma conceção menos dinâmica do que aquela proposta por Burger (2003) e, por isso, mais conservadora de ‘fraseologia’ e, consequentemente, de ‘Fraseologia’, ao colocar numa zona periférica expressões que evidenciem um menor grau de fixidez ou que não sejam idiomáticas. No entanto, o seu manual complementa o trabalho de Burger (2003), uma vez que fornece ao leitor um vasto conjunto de exemplos. Não poderíamos deixar de salientar os trabalhos de Hundt (1994) e Burger *et al.* (2007) na área da investigação fraseológica. Hundt (1994) constitui uma obra de referência sobre a Fraseologia do Português, ilustrada com numerosos exemplos. O mesmo acontece com Burger *et al.* (2007) que procura, numa síntese alargada das ideias apresentadas em Burger (2003), abordar questões de índole semântica e textual, de crucial importância no tratamento das fraseologias.

³⁰ É o caso da fixidez pragmática que caracteriza unicamente as fraseologias comunicativas, fórmulas de rotina ou fraseologias pragmáticas.

³¹ Poder-se-ão detetar algumas marcas de superfície que ajudarão essa distinção, como sejam os “componentes únicos” – (pt.) *nem chus nem bus*, (fr.) *rester bouche bée* –, particularidades como (pt.) *ser de arrepiar os cabelos*, e não **ser de arrepiar o cabelo*, (fr.) *avoir les foies (blancs)*, e não **avoir le foie (blanc)* ou incompatibilidades a nível dos traços sémicos dos componentes – (pt.) *ter um coração de pedra*, *ter a cabeça em água*, *ter um coração de ouro*; (fr.) *avoir un cœur de pierre*, *avoir un noeud dans la gorge*, *avoir du sang bleu*, por exemplo.

falantes da comunidade, enquanto Fleischer (1997) só considera um núcleo restrito, puramente idiomático.

2.2. Fixidez

52 Tido por muitos autores como o denominador comum entre todas as fraseologias – entre outros, por Zuluaga (1980: 15), Burger (2003: 16), Mellado Blanco (2004), Svensson (2004) –, o caráter estável das fraseologias resulta da fixação em bloco de um grupo livre de palavras, sendo a estabilidade da sua forma e o caráter constante da sua composição lexical que determinam a sua reprodução holística no discurso³².

Fixidez psicolinguística	A ‘memorização’ (Svensson 2004: 45 e ss.), propriedade partilhada por todas as fraseologias, implica o processamento de estruturas pré-fabricadas na memória do falante e, posteriormente, a sua recuperação e reprodução em contexto comunicativo ³³ .
Fixidez estrutural	Critério simultaneamente necessário e suficiente ³⁴ revela, ainda que em grau variável, fixidez morfossintática, por um lado, e fixidez semântico-lexical, por outro.
Fixidez pragmática	Deflui da fixação de determinadas fraseologias a contextos comunicativos específicos (Burger 2003: 29 e ss.). São disso exemplo as fórmulas de cumprimento, de agradecimento, de despedida e de felicitação.

Quadro 1 - Propriedades das fraseologias

³² No entanto, como refere Zuluaga (1980: 24), nem todas as reproduções discursivas levam à fixação em bloco de palavras que normalmente coocorrem. São, na verdade, repetições com uso social concreto e generalizado a todos os contextos de comunicação que originam as fraseologias. Ficam assim excluídas as “fraseologias familiares”, por não se estenderem a uma comunidade linguística mais vasta.

³³ Testes psicolinguísticos com fraseologias incompletas que os falantes deverão preencher (Lückentests) – mencionados por Burger (2003: 17-18) e exemplificados por Svensson (2004: 47 e ss.) – provam a natureza holística das fraseologias (Athayde 2007: 45). Com efeito, os resultados dos testes não só corroboram que o mesmo preenchimento ocorre repetidamente com um grande número de falantes, o que pressupõe que são reconhecidas intuitivamente por estes, como também permitem avaliar o grau de enraizamento das fraseologias na comunidade linguística.

³⁴ Propriedade escalar que permite distinguir as fraseologias do “discurso livre”. Encontraremos, por outras palavras, expressões mais fixas do que outras em função dos bloqueios paradigmáticos ou transformacionais que sobre elas impendem: quanto mais bloqueios sofrerem ou anomalias apresentarem, mais fixas (estáveis) serão as fraseologias em análise.

Recorrendo a Burger (2003: 16 e ss.), percebe-se facilmente que a fixidez é uma propriedade escalar e multidimensional – com caráter psicolinguístico, estrutural (morfo-sintático e semântico-lexical) e pragmático – que define as fraseologias por oposição à coocorrência livre de palavras, motivo pelo qual as abordaremos com algum pormenor no Quadro 1.

2.3. Idiomaticidade

Propriedade entendida como sinónima de não-composicionalidade, sendo que não existe nenhum mecanismo que nos permita, a partir da leitura literal, deduzir o significado do idiomatismo. É, de acordo com Fleischer (1997: 30 e ss.) e Burger (2003: 31-32), a discrepância entre o significado extrafraseológico e o sentido figurado da expressão que está na origem de combinações de três tipos:

Expressões parcialmente idiomáticas em que parte dos componentes mantém o significado que carrega fora da fraseologia:

- (1) *Receber de braços abertos*
- (2) *Savoir sur le bout du doigt*

Expressões totalmente idiomáticas, marcadas, globalmente, pelo seu sentido figurado:

- (3) *Dar o braço a torcer*
- (4) *Manger [qqn.] des yeux*

Expressões não-idiomáticas em que os componentes preservam o seu valor referencial extrafraseológico:

- (5) *Lavar os dentes*
- (6) *Se peigner les cheveux*

Discutidas as propriedades do fenômeno fraseológico, convém lembrar que a investigação se debate ainda com problemas na definição e classificação do seu objeto de estudo, uma vez que os limites delineados no *continuum* entre o ‘livre’ e o ‘fixo’ são muito ténues.

54 3. Somatismos do Português e do Francês

Como se pode compreender, a complexidade e a heterogeneidade das expressões que se enquadram no escopo fraseológico motivou a necessidade de estabelecer um microparadigma para este trabalho no seio da classe das fraseologias de valor referencial³⁵.

Fixámo-nos nas fraseologias denominativas, mais detalhadamente nos SO³⁶ e, ainda no âmbito destes, nos cinegramas³⁷ (*levantar os ombros, abanar a cabeça, arregalar os olhos; hausser les épaules, baisser les bras, tendre l'oreille*), deixando assim de parte provérbios³⁸ (*Deus dá nozes a quem não tem dentes; il faut tourner sa langue sept fois dans sa bouche avant de parler*). Abordaremos também alguns pares e algumas comparações fraseológicas (*sem pés nem cabeça, se cacher la tête comme une autruche, être bête comme ses pieds*) no domínio dos “modelos fraseológicos”.

³⁵ Na evidente impossibilidade de seguir um critério único, a investigação fraseológica reconhece a importância das “classificações mistas” (*Mischklassifikationen*), herdadas da tradição soviética, que combinam diferentes critérios. A tipologia de Burger (2003: 36 e ss.) serviu de enquadramento ao nosso trabalho, sendo a clareza, o grau de sistematicidade e o caráter inovador, o motivo por que se generalizou no âmbito da Germanística.

³⁶ Sobre a origem e formação dos SO, destacamos os importantes contributos de Földes (1985), Čermák (1998) e Mellado Blanco (2004).

³⁷ Incluídos no domínio das “classes especiais” por Burger (2003: 46), os cinegramas evocam um comportamento – concebido e codificado linguisticamente com base no simbolismo que assumiu num momento específico. As fraseologias inseridas nesta categoria mantêm ativas, ao contrário do que acontece com os restantes idiomatismos (Burger 2007: 101), a leitura literal e a leitura fraseológica de uma expressão.

³⁸ As expressões idiomáticas do nosso *corpus* têm a forma de sintagma, sendo que a sua utilização na comunicação admite livre preenchimento do sujeito e dos complementos.

3.1. Designações de partes do corpo usadas em expressões idiomáticas

Mediante a análise dos dados recolhidos, o *corpus* em estudo revelou, como seria de prever, alguma convergência no que respeita às referências e analogias traçadas no uso de SO em ambas as línguas.

Registou-se que, por exemplo, o conceito de “amor” e a imagem que a este associamos (bondade, ternura, força, generosidade, sinceridade, etc.) são comumente referenciadas pelo “coração”, quer no Português, quer no Francês, facto que explica ocorrências como³⁹:

55

(7) *fazer [qc.] do coração, ir direito ao coração [de alg.], ter alegria no coração, ter (bom) coração, ter um coração de leão / forte, ter um coração de ouro*

(8) *avoir du cœur, avoir la main sur le cœur, parler à cœur ouvert, porter [qqn.] dans son cœur, ouvrir son cœur*

No conjunto dos SO recolhidos, pudemos verificar que as denominações das partes do corpo humano mais produtivas são substancialmente as mesmas em ambas as línguas, como ilustra o Gráfico 1:

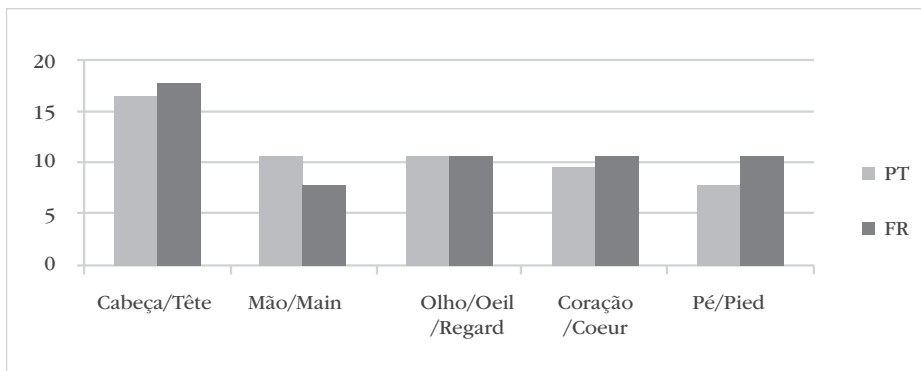


Gráfico 1 - Designações de partes do corpo humano com uma utilização superior a 8%

³⁹ Contudo, o uso deste componente nem sempre serve a expressão de sentimentos eufóricos, mas também disfóricos, que aludem, claramente, a sentimentos de medo, receio e preocupação: *ter um aperto no coração; estar / ficar com o coração nas mãos*. Encontrámos ainda exemplos que exprimem, respetivamente, a necessidade de ultrapassar dificuldades, nostalgia, tristeza e dor causada em outrem: *pôr o coração ao largo; avoir le cœur serré, briser / déchirer le cœur [de qqn.], crever le cœur [de qqn.]*.

Ao analisarmos estes dados, podemos concluir que a maior ou menor frequência de uso de determinada designação de órgão ou membro prende-se, em grande parte, com a sua potencialidade semântica. Com efeito, são mais utilizadas designações e órgãos que desempenham um papel mais relevante nas nossas capacidades motoras, sensoriais e cognitivas.

Verificámos também que alguns dos lexemas para referir uma parte do corpo ou atividade se referem à mesma parte do corpo, ou seja, são variantes lexicais ao nível dos SO como *coração / cor*⁴⁰, *orelha / ouvido*, *face / figure*, *doigt / pouce*⁴¹, *œil / regard*, ainda que designent partes do corpo distintas na realidade extrafraseológica.

A lista dos componentes que remetem para partes do corpo humano é bastante extensa. Ao analisarmos os dados recolhidos, pudemos concluir que as denominações das partes do corpo humano mais produtivas são substancialmente as mesmas em ambas as línguas: *cabeça/tête*, *mão/main*, *coração/cœur*, *olho/œil*, *pé/pied* e *boca/bouche*. A maior ou menor frequência de uso de determinada designação de órgão ou membro prende-se, em grande parte, com a sua potencialidade referencial. Com efeito, são mais utilizadas designações de órgãos que desempenham um papel mais relevante nas nossas capacidades motoras, sensoriais e cognitivas. A “cabeça” simboliza o intelecto, o esclarecimento, o centro da atividade cognitiva do ser humano. Por sua vez, a “mão” remete para qualquer atividade, o poder e o domínio. O “coração” é o órgão central do indivíduo, apontando, de modo geral, para os sentimentos, a inteligência e a intuição (centro de afetividade e intelectualidade). O “pé” corresponde ao apoio do corpo, sendo por isso símbolo de apoio e poder. Por último, a “boca” é capaz de destruir, matar, confundir e rebaixar, pois, como sabemos, a boca derruba tão depressa como edifica os seus castelos de palavras, como se observa em (pt.) *andar de boca em boca*, *andar nas bocas do mundo*, *dizer à boca cheia*, *ser só da boca para fora*, *ter má boca*; (fr.) *courir de bouche en bouche*, *dire à pleine bouche*, *être sur toutes les lèvres*.

⁴⁰ Palavra derivada do étimo latino *cor*, *cordis* que preserva ainda, no Português contemporâneo, a sua forma arcaica.

⁴¹ Consideramos *doigt* e *pouce* numa mesma rubrica, uma vez que estão intimamente relacionadas pela proximidade espacial e, naturalmente, pela relação de metonímia estabelecida na expressão *ne pas bouger d'un pouce*, em que esse último – *polegar / pouce* – representa um dedo específico.

A comparação entre as duas línguas revelou ainda que algumas designações não denotem tão grande produtividade como as referidas anteriormente. É o caso de *fígado*, ilustrado, no caso do Português, com uma única ocorrência, numa clara alusão ao “(mau) humor”:

(9) *ter maus fígados*

57

No caso do Francês, o mesmo componente integra apenas quatro SO que apontam para o “medo” e para o “humor”:

(10) *avoir les foies blancs*

(11) *donner les foies*

(12) *avoir les jambes en pâté de foie*

(13) *manger les foies*

Com base nos exemplos (9) a (13), interessante se torna verificar que, ao contrário do que sucede em Francês, não se registam em Português idiomatismos com o componente *fígado* que reenviem para a noção de “medo”.

3.2. Análise contrastiva

Refletiremos, neste capítulo, sobre as relações de “forma-significado” que se estabelecem entre os pares de expressões que serão objeto de descrição, fundamentando-nos principalmente na proposta de análise contrastiva de Larreta Zulategui (2001: 69).

O linguista avalia, no quadro do Espanhol e do Alemão, a relação “forma-significado” dos SO recolhidos com base em:

- aspetos morfossintáticos, sendo consideradas a isomorfia ou a anisomorfia de estruturas;
- aspetos lexicais, sendo avaliado o grau de congruência do componente lexical;

- aspetos semânticos, que avaliam o conteúdo dos SO, isto é, a equivalência do significado fraseológico nas expressões recolhidas no par de línguas em estudo.

58 Tendo em conta estes critérios, encontraremos casos de equivalência total ou absoluta, parcial ou nula, sendo também incontornável a referência a raros casos de pseudo-equivalência ou equivalência aparente, que seguidamente abordaremos.

3.2.1. Equivalência total

Relação entre expressões das duas línguas que veiculam o mesmo conteúdo semântico e que utilizam os mesmos recursos equivalentes e cognatos, como ilustra o Quadro 2⁴².

Português	Francês
<i>Abrir a boca</i>	<i>Ouvrir la bouche</i>
<i>Ser o braço direito [de alg.]</i>	<i>Être le bras droit [de qqn]</i>
<i>Meter [qc.] na cabeça [de alg.]</i>	<i>Mettre [qqcb] dans la tête [de qqn]</i>
<i>Ter a cabeça fria</i>	<i>Avoir la tête froide</i>
<i>Agarrar a ocasião pelos cabelos</i>	<i>Saisir l'occasion par les cheveux</i>
<i>Ser o calcanhar de Aquiles [de alg.]</i>	<i>Être le talon d'Achille [de qqn.]</i>
<i>Virar/voltar as costas [a alg./qc.]</i>	<i>Tourner le dos [à qqn./qqcb.]</i>
<i>Estar armado até aos dentes</i>	<i>Être armé jusqu'aux dents</i>
<i>Estar em boas mãos</i>	<i>Être en bonnes mains</i>
<i>Ter (uma) memória de elefante</i>	<i>Avoir une mémoire d'éléphant</i>
<i>Meter o nariz [em qc.]</i>	<i>Mettre le nez [dans qqcb]</i>
<i>Olhar [alg.] por cima do ombro</i>	<i>Regarder [qqn] par-dessus l'épaule</i>
<i>Arrebitar a orelha</i>	<i>Dresser/tendre l'oreille</i>
<i>Estar em/num tête-à-tête [com alg.]⁴³</i>	<i>Être en tête-à-tête [avec qqn]</i>

Quadro 2 – Equivalentes totais

⁴² A relação de equivalência total encontra a sua justificação nas funções desempenhadas pelas partes do corpo humano designadas – comuns a todos os seres humanos – e pelo simbolismo que a elas se atribui – fruto, por sua vez, da base cultural comum ao par de línguas que nos propomos estudar.

⁴³ Note-se, no caso do Português, a manutenção do original, o galicismo “tête-à-tête”.

3.2.1.1. Casos problemáticos no quadro da equivalência total

Detetámos SO que não correspondem plenamente à definição de ‘equivalência total’, dado que evidenciam diferenças de superfície que não cumprem o grau máximo de congruência pretendido⁴⁴.

Dos diferentes casos identificados⁴⁵, exemplificaremos apenas o da sinonímia verbal por ser, a nosso ver, o mais problemático, merecendo, por esse motivo, um tratamento específico.

59

Português	Francês
<i>Encher a barriga/o papo</i>	<i>Se remplir le ventre</i>
<i>Andar/correr de boca em boca</i>	<i>Aller/courir de bouche en bouche</i>
<i>Cair/meter-se na boca do lobo</i>	<i>Se jeter dans la gueule du loup</i>
<i>Virar a cabeça [a alg.]</i>	<i>Tourner/reverser la tête [à qqn] Faire tourner la tête [de qqn]</i>
<i>Escapar por entre os dedos</i>	<i>Glisser entre les doigts</i>
<i>Mostrar/arreganhar os dentes</i>	<i>Montrer les dents</i>
<i>Não conseguir pregar olho a noite toda</i>	<i>Ne pas pouvoir fermer l'œil de la nuit⁴⁶</i>
<i>Levar [alg.] pela ponta do nariz</i>	<i>Mener [qqn.] par le bout du nez</i>
<i>Encolher os ombros</i>	<i>Hausser les épaules</i>

Quadro 3 - Casos problemáticos no quadro da ‘equivalência total’: a sinonímia verbal

Exemplos do Quadro 3 demonstram que, embora não sejam totalmente equivalentes, os verbos utilizados em ambas as línguas veiculam um mesmo

⁴⁴ A categorização de alguns casos como aqueles que passaremos a enunciar não é consensual na área da investigação, por estes serem fenómenos de fronteira. Hundt (1994) e Batista (2006) inserem estas diferenciações interlinguísticas nos equivalentes parciais, mesmo que não afetem nem a estrutura sintática nem o componente lexical básico, elementos-chave que definem a relação de equivalência interlinguística.

⁴⁵ Identificámos, nesta zona de fronteira, SO em que as variantes ou covariantes apresentam diferenças a nível do artigo, do modificador adjetival ou até mesmo da adição de um componente nominal, por um lado, e a ocorrência de preposições distintas em cada uma das línguas que não permite paralelismos sistemáticos de uma língua para outra, por outro lado.

⁴⁶ O preenchimento dos complementos apresenta ainda algumas divergências, sendo de salientar o valor enfático de “toda” na expressão portuguesa.

sentido, evocam uma mesma representação mental, justificando, na nossa opinião, a sua inclusão no grupo dos equivalentes totais.

São exemplo de discrepâncias que não podem deixar de ser abordadas num trabalho de índole contrastiva, por constituírem um importante núcleo de trabalho nas aulas de LE.

60

3.2.1.2. Equivalência total defetiva

Podem ainda atestar-se casos de equivalência total defetiva em que as diferenças revelam anisomorfias tipológicas entre línguas comparadas. São, segundo Larreta Zulategui (2001: 103-104), discrepâncias sistemáticas, fruto de idiosincrasias gramaticais de cada língua.

Uma das especificidades do Português é a possibilidade de contração da preposição e do determinante, apenas presente no Francês nos casos pontuais de *des* (de+les), *aux* (à+les) e *du* (de+le):

(14) *ter um nó na garganta*

(15) *avoir un nœud dans la gorge*

Outra anisomorfia tipológica envolve o uso do partitivo em Francês, sem correspondente explícito no Português:

(16) *chorar lágrimas de sangue*

(17) *pleurer des larmes de sang*

A aparente divergência no que respeita a realização da categoria ‘número’ é outro caso de equivalência total defetiva:

(18) *ser um quebra-cabeças⁴⁷*

(19) *être un casse-tête (chinois)*

⁴⁷ Temos em *quebra-cabeças* uma palavra composta que, em Português, não admite singular, provando assim que estamos perante um equivalente total.

Por último, Földes (1996: 122-123) e Larreta Zulategui (2001: 125 e ss.) referem ainda casos especiais de hiponímia / hiperonímia que não deverão ser considerados exemplos inequívocos de equivalentes totais ou parciais. Situados na fronteira entre dois subtipos de relações, são casos como os apresentados seguidamente que nos permitem concluir que a equivalência total fica comprometida por força da presença de traços sémicos adicionais numa das fraseologias contrastadas:

61

(20) *saltar à vista, andar nas bocas do mundo*

(21) *sauter aux yeux, être sur toutes les lèvres*

3.2.2. Equivalência parcial

Grande parte dos SO recolhidos enquadram-se na categoria dos equivalentes parciais, motivo pelo qual sentimos a necessidade de estabelecer subdivisões, na medida em que, partindo do mesmo significado denotativo, de imagens muito próximas, foram encontrados diferentes graus de isomorfia estrutural e de congruência do componente lexical.

No entanto, dada a descrição extremamente fina que desenvolvemos e o elevado número de SO analisados, seguiremos, de uma forma mais abrangente, as categorias de sinónimos estruturais, dos sinónimos ideográficos e dos sinónimos funcionais propostos por Larreta Zulategui (2001) e Földes (1996).

3.2.2.1. Sinonímia estrutural

Trata-se de uma relação em que os SO partilham um mesmo significado denotativo um modelo sintático semelhante, mas cuja estrutura apresenta algumas diferenças a nível do componente lexical e/ou a nível morfossintático⁴⁸.

⁴⁸ Centrámó-nos, no caso das discrepâncias a nível morfossintático, nos pares de fraseologias que registam: divergências no uso de determinante – na sua presença ou na sua ausência – (*ter mão de ferro / avoir une main de fer*), divergências no uso da preposição (*pôr o dedo na ferida / mettre le doigt sur une plaie*) e divergências a nível da categoria de ‘número’ (*bater o dente / claquer des dents*).

Os exemplos apresentados nesta secção do nosso trabalho constituíram casos de difícil classificação no âmbito da investigação, uma vez que a utilização de componentes lexicais diferentes em ambas as línguas suscita uma imagem também ela diferente nos falantes ou ouvintes (Athayde 2007: 133).

62 Dado a extensão dos SO incluídos neste grupo, destacamos seguidamente alguns exemplos no quadro da variação lexical⁴⁹, mais precisamente as discrepâncias que consistem sobretudo na alteração do componente nominal, também com alguma influência no componente adjetival, por serem as mais produtivas e singulares no nosso *corpus*.

Português	Francês
<i>Ficar de boca <u>aberta</u></i>	<i>Rester bouche <u>bée</u></i>
<i>Fazer [qc.] de <u>cabeça</u> fria</i>	<i>Faire [qqcb] de <u>sang-froid</u></i>
<i>Não ter <u>cabeça</u> [para qc.]</i>	<i>(Ne pas) avoir l'<u>esprit</u> [à faire qqcb]</i>
<i>Não chegar aos <u>calcanhares</u> [de alg.]</i>	<i>Ne pas arriver à la <u>cheville</u> [de qqn]</i>
<i>[Qc.] custar os olhos da <u>cara</u></i>	<i>[Qqcb] Coûter les yeux de la <u>tête</u></i>
<i>Partir <u>a cara</u>/os dentes [a alg.]</i>	<i>Casser le <u>nez</u> [à qqn]</i>
<i>Ser <u>a cara</u> (chapada) [de alg.]</i>	<i>Être le <u>portrait</u> (craché) [de qqn]</i>
<i>Pôr o coração ao <u>largo</u></i>	<i>Mettre le cœur de <u>côté</u></i>
<i>Ter um coração <u>de leão</u>/forte</i>	<i>Avoir un <u>brave</u> cœur</i>
<i>Escolher [alg.] <u>a dedo</u></i>	<i>Choisir [qqn] de l'<u>œil</u></i>
<i>Ter <u>um nó</u> na garganta</i>	<i>Avoir la gorge <u>serrée</u></i>
<i>Saber na <u>ponta da língua</u></i>	<i>Savoir sur le bout du <u>doigt</u></i>
<i>Ter [qc.] entre as <u>mãos</u></i>	<i>Avoir [qqcb] sur les <u>bras</u></i>
<i>Ter <u>mãos</u> de fada</i>	<i>Avoir des <u>doigts</u> de fée</i>
<i>Ter ouvidos de <u>lísico</u></i>	<i>Avoir de <u>bonnes</u> oreilles</i>
<i>Não arredar <u>pé</u></i>	<i>Ne pas bouger d'un <u>pouce</u></i>
<i>Sem <u>pés</u> nem cabeça</i>	<i>Sans <u>queue</u> ni tête</i>
<i>Gritar a <u>plenos pulmões</u></i>	<i>Crier à <u>tue-tête</u></i>

Quadro 4 - Sinónimos estruturais - divergências no componente nominal ou adjetival

⁴⁹ No conjunto do *corpus* recolhido, detetámos divergências a nível do componente verbal (*fazer* braço de ferro/ *avoir* un bras de fer; *puxar* pela cabeça / *se creuser* la tête), na divergência de um ou mais lexemas (*apanhar* [alg.]com a *boca* na *botija* / *prendre* [qqn.] la *main* dans le *sac*; fugir a *sete pés* / *se sauver* à *toutes jambes*) e na diferença no número de componentes (*suar sangue* / *suer sang* *et eau*; *esfregar as mãos* *de contente* / *se froter les mains*).

3.2.2.2. Sinonímia ideográfica

Defendemos, na senda de Larreta Zulategui (2001: 144), que a sinonímia ideográfica é também um caso de equivalência parcial. Para que sejam considerados sinónimos ideográficos, a imagem mental e o significado veiculado pelos idiomatismos de ambas as línguas deverão ser forçosamente idênticos, ainda que o modelo sintático possa ser distinto (anisomorfia estrutural). Contudo, estas fraseologias manifestam alguma congruência lexical, uma vez que o elemento-chave (o componente nominal) e todo o seu simbolismo são equivalentes em ambas as línguas.

63

Detetámos, no *corpus* recolhido, alguns sinónimos ideográficos, mais do que inicialmente esperávamos, como são exemplo:

(22) *dormir com um olho aberto e outro fechado, andar com a cabeça à roda, estar por um cabelo*

(23) *ne dormir que d'un œil, avoir la tête qui tourne, ne tenir qu'à un cheveu*

Verificamos, a título ilustrativo, que *dormir com um olho aberto e outro fechado* e *ne dormir que d'un œil* suscitam um mesmo significado denotativo, confirmado pelo recurso não só ao mesmo verbo, como também à designação de um mesmo órgão do corpo humano, os olhos. Se o Português coloca a tónica no papel desempenhado por cada olho, um aberto – a vigiar – e outro fechado – a descansar –, o Francês opta por valorizar o facto de “dormir” apenas afetar um olho, uma vez que o outro e encontra em alerta.

3.2.2.3. Sinonímia funcional

Neste caso, a relação estabelecida entre os SO de ambas as línguas resume-se à equivalência semântica: portadores de um mesmo significado,

os sinónimos funcionais apresentam um preenchimento lexical e estruturas distintas nas línguas contrastadas⁵⁰:

(24) *ficar de barriga para o ar, escapar por entre os dedos, não ter mãos a medir, fazer [qc.] com uma perna às costas*

(25) *se tourner les pouces, passer sous le nez, ne pas avoir cinquante bras, réussir / gagner les doigts dans le nez*

64

3.2.3. Equivalência através de estruturas não-idiomáticas

Contrariamente ao que poderíamos pensar, a ausência de equivalentes em termos de idiomatismos não é sinónimo de ‘ausência de correspondência’. Os idiomatismos de uma língua nem sempre têm, como equivalente na outra língua, um idiomatismo, motivo por que serão, por um mecanismo de compensação interlinguística (Eismann, 1995), vertidas sob a forma de lexema, de um outro tipo de estruturas fraseológicas não-idiomáticas (como sejam, por exemplo, as construções com verbo-suporte) ou através de sintagmas livres que funcionam como suas paráfrases:

(26) *culpar [alg.], ter as costas quentes, ter maus fígados*

(27) *mettre [qqcb.] sur le dos de [qqn.], être sous la protection de [qqn.], être de mauvaise humeur*

3.2.4. Pseudo-equivalência

Comumente designados por ‘falsos amigos’⁵¹, estes pares são expressões semelhantes a nível estrutural e sintático, mas veiculam significados distintos.

⁵⁰ Hundt (1994) inclui-os, por esse motivo, no grupo dos equivalentes parciais. No entanto, não poderemos deixar de referir a perspetiva de Baptista (2006: 105-106), que opta por tratá-los num grupo independente da equivalência parcial por constituírem expressões totalmente distintas, quer de um ponto de vista formal, quer da perspetiva da imagem que transmitem.

⁵¹ Conceito desenvolvido pela Teoria da Tradução e pela Didática de Línguas Estrangeiras.

O falante ou aprendente de LE poderá, por esse motivo, ser induzido em erro, acreditando estar perante estruturas sinónimas: *estar com a corda ao pescoço / se mettre la corde au cou, tomar corpo / prendre corps*.

4. Considerações finais

65

Como foi oportunamente referido, quanto maior o grau de idiomaticidade das expressões, mais dificuldades de codificação e de descodificação suscitam nos aprendentes de uma LE. Assim, em atividades de produção, e desconhecendo a construção equivalente na língua-alvo, o aluno tenderá a produzir enunciados fraseológicos decalcados da sua LM, mecanismo que poderá resultar em erros, dada a frequente ausência de relações de equivalência total entre unidades fraseológicas das diferentes línguas. Sendo um domínio do Léxico que coloca, em regra, algumas dificuldades ao aprendente de uma LE, a sua inclusão nas aulas de língua é, todavia, indispensável se tivermos em mente que o objetivo é aproximar a proficiência do aprendente estrangeiro da do falante nativo. A natureza muitas vezes lúdica, o caráter expressivo e as suas potencialidades retóricas podem, contudo, transformar a aquisição de fraseologias num tópico que motiva o aluno para a aprendizagem da LE e lhe proporciona não apenas conhecimentos de índole linguística, mas também de caráter sociocultural e intercultural.

A análise estatística dos dados revelou, como seria de prever, informações que já tinham sido adiantadas. Assim, pudemos verificar que existe, de facto, maior número de equivalentes parciais, ainda que, quando comparados com os equivalentes totais, tenhamos valores que excedem ou que estão muito próximos de metade do *corpus* recolhido para este tema. De referir ainda que o número de casos identificados como problemáticos no quadro da 'equivalência total' é bastante significativo, quando comparado com o número de casos que não mereceram reflexões específicas. Por último, a análise do *corpus* permitiu-nos adicionalmente concluir que, no quadro da relação de 'equivalência parcial', prevalece o grupo dos sinónimos estruturais – com o claro predomínio de casos que apresentam variação lexical –, seguido pelos

sinónimos ideográficos, nos quais, como referimos oportunamente, a designação da parte do corpo presente no SO é idêntica no Português e no Francês.

66 A par do confronto de unidades desenvolvido, o contraste interlinguístico assumiu, cumulativamente, a dimensão de um contraste de sistemas, permitindo-nos destacar a existência de “europeísmos” e “universais fraseológicos”. Com efeito, o facto de os SO serem construídos a partir das experiências motivadas por um mesmo elemento em ambas as culturas – o corpo humano – e a partilha de tradições comuns a ambas as comunidades linguístico-culturais explicam a existência de expressões muito semelhantes. O mesmo será dizer que todos os signos linguísticos, incluindo os signos fraseológicos, carregam uma dimensão cultural partilhada por muitas línguas, provando que grande parte do nosso património linguístico e sociocultural deflui de um fundo cultural e civilizacional comum, como sejam a tradição greco-latina, as mesmas origens judaico-cristãs, a Literatura Universal, transmitidas aos longo dos tempos, de geração em geração, de comunidade em comunidade, de continente em continente.

No entanto, SO com um preenchimento totalmente distinto encontram justificação nos diferentes esquemas de apreensão do mundo envolvente, subjacentes à transposição do plano real – da experiência vivida – para o plano figurado, verbalizado, neste caso, através da fraseologia de cada uma das línguas.

A finalizar, gostaríamos de salientar que, não obstante este trabalho se ter dedicado especificamente à análise contrastiva de SO, outras questões, a equacionar em futuros trabalhos, ficaram ainda por explorar. Referimo-nos, por exemplo, a um alargamento do objeto de estudo a outras subclasses de fraseologias com componentes somáticos – como os provérbios, entre outros – ou a um aprofundamento de temas relacionados com a Didática do Fraseoléxico que envolvessem propostas concretas de material didático a utilizar – cadernos de exercícios, capítulos de manuais escolares ou de manuais de apoio ao professor de LE.

Referências

- Athayde, Maria Francisca (ed.) (2006). *Estudos sobre Léxico e Gramática*. Centro Interuniversitário de Estudos Germanísticos. Coimbra: Minerva Coimbra.
- Athayde, Maria Francisca (2007). *Programas, conteúdos e métodos do ensino teórico e prático das matérias da disciplina semestral do Seminário de Língua*. Relatório apresentado ao concurso para provimento de uma vaga de Professor Associado do 4º Grupo (Estudos Germanísticos) da Faculdade de Letras da Universidade de Coimbra.
- Baptista, Sofia Margarida (2006). "Dos pés à cabeça / Von Kopf bis Fuß. Análise contrastiva de somatismos no Português e no Alemão". In: Athayde, Maria Francisca (ed.) *Estudos sobre Léxico e Gramática*, Centro Interuniversitário de Estudos Germanísticos. Coimbra: Minerva Coimbra, 81-110.
- Burger, Harald (2003). *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 2., überarbeitete Auflage. Berlin: Erich Schmidt Verlag.
- Burger, Harald, Annelies Häcki Buhofer, Gertrud Gréciano (Hrsg.) (2003). *Flut von Texten – Vielfalt der Kulturen. Ascona 2001 zur Methodologie und Kulturspezifität der Phraseologie*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Burger, Harald, Dmitrij Dobrovol'skij, Peter Kühn, Neal R. Norrick (Hrsg.) (2007). *Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung*, Volume 1, 2. Berlin/New York: de Gruyter.
- Čermák, František (1998). Somatic Idioms Revisited. In: Eismann, Wolfgang (Hrsg.), 109-119.
- Corpas Pastor, Gloria (1996). *Manual de fraseologia espanhola*. Madrid: Gredos.
- Cortès, Colette (1999). Phraséologie et corps humain. Étude comparative du corps humain dans les expressions phraséologiques en français et en allemand. *Cahier du CIEL* (1998-1999), 85-109.
- Eismann, Wolfgang (1995). Pragmatik und kulturelle Spezifik als Problem der Äquivalenz von Phraseologismen. In: Baur, Rupprecht S. und Christoph Chlosta (Hrsg.). *Von der Einwortmetapher zur Satzmetapher. Akten des Westfälischen Arbeitskreises Phraseologie, Parömiologie*, 94/95. Bochum: Universitätsverlag Dr. N. Brockmeyer, 95-119.
- Eismann, Wolfgang (Hrsg.) (1998). *Europbras 95 - Europäische Phraseologie im Vergleich*. Bochum: Brockmeyer.
- Fleischer, Wolfgang (1997). *Phraseologie der deutschen Gegenwartssprache*, Tübingen: Max Niemeyer Verlag.
- Földes, Csaba (1985). Über die somatischen Phraseologismen der deutschen, russischen und ungarischen Sprache. Versuch einer konfrontativen Analyse. *Germanistische Jahrbuch DDR-UVR* 4, 18-40.
- Földes, Csaba (1996). *Deutsche Phraseologie kontrastiv: intra- und interlinguale Zugänge*. Heidelberg: Groos.
- Hundt, Christine (1994). *Untersuchungen zur portugiesischen Phraseologie*. Wilhelmsfeld: Egert Verlag.
- Larreta Zulategui, Juan Pablo (2001). *Fraseologia contrastiva del alemán y el español. Teoría y práctica a partir de un corpus bilingüe de somatismos*. Frankfurt am Main: Peter Lang.
- Mejri, Salah (2003). La stéréotypie du corps dans la phraséologie: approche contrastive. In: Burger, Harald et al. (Hrsg.) (2003), 219-228.
- Mellado Blanco, Carmen (2004). *Fraseologismos somáticos del alemán*. Frankfurt am Main: Peter Lang.
- Sanromán, Álvaro Iriarte (2001). *A unidade lexicográfica. Palavras, colocações, frases, pragmatemas*. Universidade do Minho, Centro de Estudos Humanísticos.

- Steyer, Kathrin (Hrsg.) (2004). *Wortverbindungen – mehr oder weniger fest*. Institut für Deutsche Sprache, Jahrbuch 2003. Berlin/New York: Walter de Gruyter.
- Svensson, Maria Helena (2004). *Critères de figement. L'identification des expressions figées en français contemporain*. Omslag: Tommy Sund.
- Zuluaga, Alberto (1980). *Introducción al estudio de las expresiones fijas*. Frankfurt am Main: Peter Lang.
- Zuluaga, Alberto (2002). Los “enlaces frecuentes” de María Moliner: Observaciones sobre las llamadas colocaciones. *Philologie im Netz*, 22/2002.

SECÇÃO 3:
MORFOLOGIA E SINTAXE

(Página deixada propositadamente em branco)

VARIAÇÃO DA PRODUTIVIDADE DE SUFIXOS
AGENTIVOS EM IDADE ESCOLAR

To what extent are children miniature scientists:
testing hypotheses and checking up on what they
know about particular words or constructions?
Eve Clark, *First Language Acquisition* (2003)

1. Enquadramento teórico-descritivo

O presente estudo centra-se no uso produtivo de sufixos agentivos por indivíduos nativos do Português Europeu (L1) em idade escolar, questionando-se a ocorrência de variação da produtividade de tais sufixos em função da faixa etária e do nível de escolarização de quem os utiliza⁵².

Este estudo surge num contexto marcado pela escassa investigação nas áreas da Linguística e da Psicolinguística direcionada para processos derivacionais, em particular para a formação de agentivos (sobretudo deverbais) no Português Europeu atual (L1), envolvendo indivíduos em idade escolar.

⁵² Texto elaborado no âmbito da dissertação de Mestrado em Linguística e Ensino, especialidade de Linguística Aplicada, intitulada *Produção de Nomes Agentivos Deverbais em Português Língua Materna (L1) por Alunos do Ensino Básico*, concluída em 2009 à Faculdade de Letras da Universidade de Coimbra, sob a orientação da Professora Doutora Cristina dos Santos Pereira Martins e Professora Doutora Graça Maria Rio-Torto.

1.1. Definição de produto agentivo

72

Importa, em primeiro lugar, refletirmos sobre a definição de produto agentivo. Trata-se de uma construção, tradicionalmente conhecida como *nomen agentis*, que designa um ser ou uma entidade pela prática ou exercício de uma ação. Normalmente o produto agentivo surge associado à noção de agente humano – “aquele que faz X”, sendo que X pode traduzir uma função, profissão ou atividade.

No entanto, há linguistas (Booij 2005, Rio-Torto 1998, Basílio 1995) que fazem referência a determinados casos (“leitor”, “operador” e/ou “servidor”) em que o agente pode ser humano ou não humano. Como tal, propõe-se que a interpretação do produto agentivo seja parafraseável pela expressão “aquele ou aquilo que V”, sendo que V designa a ação concreta realizada pelo agente.

1.2. Regra de Formação de Nomes Agentivos Deverbais

Segundo Rio-Torto (2007), qualquer Regra de Formação de Palavras (RFP) depende necessariamente de condicionalismos de vária ordem impostos pela L1.

A Regra de Formação de Nomes Agentivos Deverbais (RFP AG Dev) obedece, em parte, a restrições de categoria sintática quer da base, quer do produto. Assim, partimos necessariamente da combinação de um verbo (base) com um sufixo específico para formar um nome agentivo (produto), o qual pode adquirir um funcionamento adjetivo: $V \rightarrow N$ (N/A). Cf. Exemplo:

(1) [concorre-] *+nte* \rightarrow *concorrente* (N/Adj).

A RFP AG Dev aplicada ao Português Europeu prevê também condicionalismos ao nível morfológico, ou seja, que implicam a seleção de um radical ou tema verbal consoante o sufixo a utilizar. Por exemplo, os operadores *-dor* e *-nte* selecionam apenas temas verbais, enquanto *-or* e *-ista* se agregam a radicais verbais.

Em termos da estrutura argumental dos verbos que servem de base aos produtos agentivos, é condição determinante o recurso a verbos com argumento externo, denotando agentes controladores da ação, que normalmente correspondem a verbos transitivos e alguns inergativos.

No processo de formação de agentivos, podem ocorrer, também, restrições semânticas impostas pela base e sobre a base. Rodrigues (2006) refere-se à necessidade de coadunação dos traços semânticos da base e do sufixo, constituintes internos dos produtos agentivos. Por exemplo, *-dor* e *-deiro* selecionam temas de verbos transitivos ou inergativos, portadores de valores aspetuais relativos a atividades e *accomplishments*. Não se associam a temas de verbos intransitivos, que podem traduzir valor existencial, meteorológico, de movimento ou de mudança de estado.

Finalmente, restrições de ordem semântico-referencial determinam que um produto agentivo apenas será aceite no léxico da L1 se contrastar, em termos de referente e de significado, com os demais produtos construídos a partir do mesmo paradigma genolexical.

73

1.3. Aquisição e produção de nomes agentivos na infância: conhecimento atual

No decorrer das últimas três décadas alguns especialistas têm publicado estudos a propósito da aquisição infantil no domínio derivacional. No entanto, o conhecimento sobre o desenvolvimento linguístico desde idades precoces, sobretudo no que concerne aos produtos agentivos, ainda necessita de ser instigado e aprofundado, o que requer um maior número de estudos experimentais envolvendo vários sistemas linguísticos, de modo a fundamentar os dados atualmente obtidos neste domínio.

Segundo os estudos experimentais de Clark & Cohen (1984) e Lo Duca (1990), será sensivelmente a partir dos três anos de idade (dependendo da tipologia da língua de *input*) que as crianças geralmente começam a analisar e formar os primeiros derivados, nomeadamente nomes agentivos. Verificou-se que os primeiros sufixos utilizados são os mais produtivos e mais frequentes no discurso dos adultos usuários da L1. Desde muito cedo as

crianças estabelecem novos usos dos sufixos mais produtivos, por exemplo, *-er* no inglês e *-aio* no italiano (equivalentes ao *-dor* em português), muito antes de manipularem outros sufixos agentivos disponíveis na L1.

Por conseguinte, ao reconhecerem a presença de regularidades associadas a certos sufixos na sua língua, elas tendem a produzir "neologismos infantis" (designação de Lo Duca 1990), ou seja, nomes agentivos que podem ou não corresponder aos convencionalizados no discurso adulto. Cf. Exemplos:

- (2) **?bicycler* (utilizado em vez de *bicyclist*) para designar 'ciclista';
- (3) **?cooker* (em vez de *cook*) para se referir a 'cozinheiro'.

Verifica-se, portanto, que a aquisição de produtos agentivos em línguas como o inglês e/ou o italiano permite, numa primeira fase, desenvolver a criatividade linguística da criança, tendo em conta os princípios universais de transparência semântica, simplicidade formal e produtividade.

No entanto, a percepção da criança sobre a produtividade associada ao processo de formação de agentivos tende a modificar-se com o avançar da idade e quanto maior for a quantidade de palavras e estratégias convencionalmente disponíveis na L1 para expressar o significado pretendido.

Com efeito, Clark & Cohen (1982) e Lo Duca (1990) constataram que uma criança, entre os quatro e cinco anos de idade, realiza usos inovadores com outros sufixos agentivos, igualmente produtivos, tais como, *-ist*, *-ian* (no inglês) e/ou *-tore*, *-ista* e *-iere* (no italiano). Estes dados sugerem uma provável ordem de aquisição e de uso de sufixos semanticamente próximos, disponíveis na L1, sendo possível identificar um padrão de variação de produtividade ao longo do tempo.

Como consequência do aumento de conhecimento sobre as opções sufixais disponíveis na L1, a criança tende a substituir, de forma espontânea e gradual, os seus neologismos pelas formas convencionais correspondentes (Clark 2004). Assim, os princípios de convencionalidade e contraste passam a condicionar intensamente o processo de construção de agentivos, atenuando a capacidade criativa da criança que, numa fase inicial, fora bastante estimulada.

1.4. Hipóteses explicativas do processo ontogénico de construção de derivados agentivos em L1

1.4.1. Produtos agentivos atestados vs não atestados, possíveis vs impossíveis

Tal como observa Clark (2003), o discurso das crianças, especialmente o das mais novas, tende a envolver uma mistura entre produtos legítimos e ilegítimos, estes últimos decorrentes do estabelecimento de novas associações bases-sufixos. Ou seja, vão sendo produzidos produtos atestados e não atestados na L1, podendo estes últimos ser considerados possíveis ou impossíveis.

Os produtos atestados integram o léxico convencional da língua, logo são conhecidos pela comunidade linguística e encontram-se listados em livros de referência como os dicionários.

Os produtos agentivos não atestados e possíveis, embora não sejam utilizados pela comunidade linguística, obedecem aos padrões regulares morfológicos que nela vigoram.

Em contrapartida, os produtos agentivos não atestados e impossíveis (também designados pseudoformas) contrariam os paradigmas genolexicais, daí que não sejam gramaticalmente aceitáveis e, por isso mesmo, não podem integrar o léxico virtual de um sistema linguístico.

1.4.2 Atuação da memória declarativa vs. memória procedimental

Colocamos, então, a seguinte questão: - Como se explica a criação de nomes agentivos legítimos e ilegítimos, convencionais e não convencionais pelas crianças?

Para compreendermos o processo de aquisição e produção de agentivos na infância, evocamos o modelo teórico de Ullman (2003, 2001) que pressupõe que o desempenho linguístico depende de duas capacidades do indivíduo: uma primeira relacionada com o léxico mental e a outra com a gramática mental. Segundo Ullman, a memória declarativa é responsável pela

aprendizagem, representação e processamento do léxico mental, enquanto a memória procedimental possibilita a aquisição e o uso implícito de regras mentais (de ordem morfológica, sintática e possivelmente fonológica).

76 Tendo em conta o pressuposto de que o processamento de derivados envolverá uma ativação paralela dos dois subsistemas de memória, Ullman (2001) postula que sempre que um produto for identificado pela memória lexical, a computação por parte do sistema de regras será inibida. Quando a identificação do produto não for passível de ser feita pela memória lexical, será então processada uma construção entendida como regular. No nosso entender, esta ativação paralela da memória declarativa e da memória procedimental deverá ocorrer em crianças de idade precoce, cujo léxico mental ainda se encontra reduzido e que iniciam a construção de derivados agentivos, recorrendo a sufixos muito produtivos e recorrentes no discurso dos adultos (convencional).

Segundo o mesmo autor, o facto de, por vezes, não se conseguir aceder a um produto memorizado pode ter como consequência «erros de sobrerregularização», criando-se produtos agentivos possíveis tais como **?cooker* (**?cozinbador*) e **?trumpeter* (**?trompetador*).

Clark (2003) explicita que as crianças desconhecem, de início, o que é percecionado como léxico convencional ou novo ao nível das associações forma-significado e base-afixo. Só mais tarde, com o avanço da idade, e a par do aumento do seu leque de experiências linguísticas e conhecimentos sobre o léxico convencional, é que conseguem realizar um processo de "pré-esvaziamento" (*preemption*), que consiste na substituição dos neologismos pelos produtos convencionais. Passam então a adotar as formas utilizadas pela comunidade linguística (*cook – cozinheiro; trumpelist – trompetista*).

Espera-se que a criança acuse uma progressiva diminuição ou quebra de criatividade, atribuível à ação da «gramática mental», na formação derivacional, já que, e segundo sugere Ullman, a capacidade da memória declarativa vai provavelmente, com a idade, prevalecendo sobre a memória procedimental. Nestas circunstâncias será de prever uma progressiva preferência pelos derivados convencionais em detrimento dos não convencionais.

1.4.3. Outros fatores: frequência de formas convencionais e conhecimento metalinguístico

Na verdade, verifica-se um aumento gradual e cumulativo do léxico mental da criança ao longo de diferentes faixas etárias (desde os primeiros anos de vida até aos primeiros anos escolares) – logo, em momentos distintos do desenvolvimento ontogénico da memória declarativa e da memória procedimental (conceitos explicitados em 1.4.2.) que estarão envolvidos no processo derivacional.

77

É, sobretudo, a partir do momento que a criança inicia o estudo formal da L1 e durante todo o seu percurso escolar que a criança passa a estar mais desperta para a frequência de produtos e de sufixos utilizados no discurso adulto, e exposta a uma crescente quantidade de dados analisáveis no *input*.

Em virtude da maior exposição da criança ao conhecimento do convencional, cremos que, por um lado, o desenvolvimento da capacidade da sua memória declarativa permite que, de modo mais sistemático e eficaz, sejam bloqueados produtos não atestados em prol dos convencionais. Por outro lado, e em simultâneo, a criança tende a automatizar as regularidades mais produtivas da L1.

Entendemos ainda que o comportamento derivacional e o conhecimento metalinguístico, que o ensino formal ajuda a promover (Leiria 2006), tendem a influenciar-se mutuamente. Assim, após os seis anos, a criança aprende a desenvolver um saber analítico explícito sobre propriedades semânticas e sintáticas dos constituintes internos de nomes agentivos (bases e sufixos) e a aplicar esse saber com progressiva eficiência ao nível da construção e da interpretação de produtos regulares (compatíveis com as RFP AG dominantes na L1), sejam estes considerados atestados e/ou não atestados.

2. Proposta de estudo empírico: variação de produtividade de sufixos agentivos em idade escolar?

O nosso estudo empírico pretendeu testar a hipótese de variação de produtividade de sufixos agentivos aplicados ao Português Europeu (L1),

com base em produções agentivas criadas por indivíduos nativos de diferentes faixas etárias e níveis escolares.

Com este estudo, pretendemos avançar conhecimentos na área da aquisição ontogénica da morfologia lexical agentiva e abrir caminho para estudos semelhantes sobre derivação na L1.

78

2.1. População a ser estudada

O nosso estudo empírico envolveu uma diversidade de participantes situados em contextos escolares autênticos. Foram selecionados, aleatoriamente, os seguintes grupos e níveis escolares:

- a) uma turma do 2º ano, composta por um total de 21 alunos (13 elementos do sexo masculino e 8 do sexo feminino), com idades compreendidas entre os 6-7 anos;
- b) uma turma do 3º ano, formada por um total de 22 elementos (11 do sexo masculino e 11 do sexo feminino), situados na faixa etária dos 8-9 anos;
- c) uma turma do 4º ano, constituída por um total de 20 alunos (13 do sexo masculino e 7 do sexo feminino), com idades entre 9-10 anos;
- d) uma turma de 6º ano, que incluiu um total de 20 elementos (10 rapazes e 10 raparigas), na faixa etária dos 11-13 anos;
- e) uma turma de 9º ano que integrou um total de 18 elementos (6 rapazes e 12 raparigas), na faixa etária dos 14-16 anos.

Na totalidade, a amostra de estudo englobou 101 elementos, pertencentes a cinco níveis de escolaridade distintos (três turmas de 1º ciclo, uma turma de 2º ciclo e uma turma de 3º ciclo do ensino básico), sendo todos falantes nativos de Português Europeu e frequentando escolas oficiais. Esta amostra de população também teve em conta uma diferenciação de momentos de maturação cognitiva.

2.2. Métodos de recolha de dados

Numa primeira fase, foi aplicado um exercício escrito que envolveu todos os discentes participantes no estudo, no qual se solicitava o registo de todas as formas possíveis a partir da combinação de verbos e formas sufixais apresentados. Numa fase posterior, foram realizados questionários orais/escritos aos mesmos sujeitos sobre as formas agentivas registadas e respetivos constituintes internos, metodologia que forneceu dados complementares aos inicialmente obtidos.

79

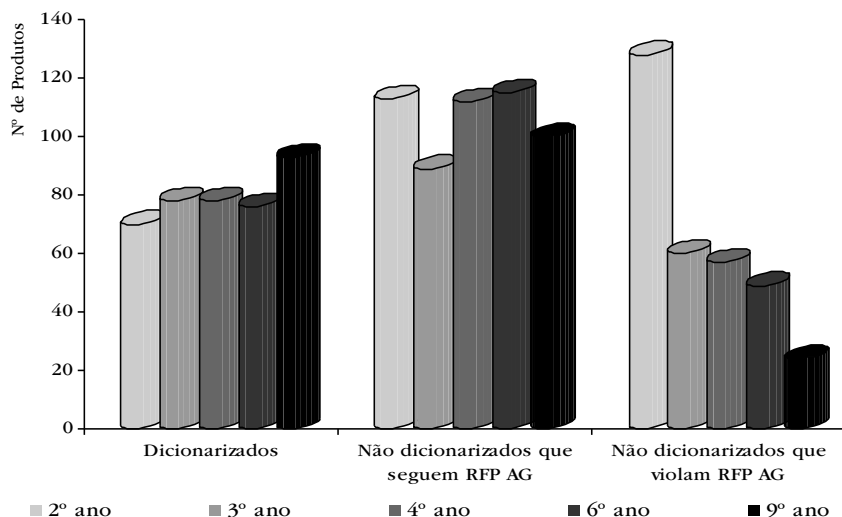
2.3. Variáveis em análise

O presente estudo teve em consideração a análise das seguintes variáveis:

- a) produtividade e frequência de uso das formas sufixais agentivas no Português Europeu atual;
- b) disponibilidade e não disponibilidade de nomes agentivos deverbais regulares (dicionarizados e não dicionarizados) no léxico mental do aluno participante;
- c) influência de conhecimentos linguísticos e metalinguísticos em diferentes anos escolares.

2.4. Resultados

Atendendo ao fator «nível escolar» (e subjacente a este, a idade dos sujeitos), podemos constatar resultados divergentes no que respeita os registos de agentivos dicionarizados e não dicionarizados.



Quadro 1: Totais dos registos de Nomes Agentivos Deverbais dicionarizados e não dicionarizados por ano escolar

Mediante a leitura do gráfico apresentado (quadro 1), podemos verificar que, em relação aos nomes agentivos deverbais dicionarizados, se regista um crescente aumento ao nível da sua produção desde o 2º ano até ao 9º ano, o que indicia uma estabilização progressiva do léxico agentivo convencional, sobretudo no último ano escolar em análise.

Por sua vez, detetámos um elevado número de registos de nomes deverbais não dicionarizados que seguem RFP AG, suplantando as restantes categorias de produtos agentivos em análise.

Por outro lado, observámos um claro decréscimo do número de registos de agentivos não dicionarizados que violam RFP AG na L1, a partir do 3º ano e, sobretudo, no 9º ano.

Assim, ao longo do seu percurso escolar, os alunos tendem a aplicar corretamente a RFP AG em novas palavras, descartando, de modo progressivo, derivados incompatíveis com a referida RFP AG.

Tal como seria de prever, são os alunos dos primeiros anos de escolaridade, que se situam na faixa dos 6-7 anos, os que produzem um número mais

elevado de nomes agentivos não dicionarizados transgressores da RFP AG, manifestando uma maior instabilidade do seu léxico mental.

Observemos, em seguida, o seguinte quadro:

Exemplos de Nomes Agentivos Dicionarizados	Ano Escolar / N° Registos				
	2º ano (6-7)	3º ano (8-9)	4º ano (9-10)	6º ano (11-14)	9º ano (14-16)
<i>consumidor</i>	5	8	14	15	16
<i>consumista</i>	1	3	9	9	13
<i>desenbador</i>	9	10	16	20	16
<i>desenbista</i>	2	0	3	5	6
<i>tecedor</i>	6	5	3	12	11
<i>tecedeira</i>	2	1	6	5	5
<i>tecelão</i>	0	1	5	2	3

81

Quadro 2: Convencionalidade Sufixal/ Frequência de Ocorrência Variável

Observa-se um aumento do conhecimento do léxico dicionarizado disponível ao longo do percurso escolar dos sujeitos, como revela o crescente número de registos de agentivos dicionarizados.

Por outro lado, é também interessante constatar que, dentro do léxico convencional, podem ocorrer diferenças na frequência de uso de sufixos agentivos. Na verdade, uma base idêntica pode ser combinada com sufixos agentivos distintos, obtendo-se produtos convencionais com registos de ocorrência variável não apenas entre diferentes anos escolares (e faixas etárias), mas também num mesmo ano escolar (ou faixa etária). Por exemplo, no 2º ano, *consumidor* obteve cinco registos e *consumista* obteve apenas um único registo. No 3º ano, *consumidor* obteve oito registos, enquanto *consumista* apenas três registos. Nos anos escolares seguintes ocorre a mesma diferenciação de registos ao nível de usos sufixais. Nomes agentivos como *desenbador* e *tecedor* foram identificados com maior recorrência do que outros construídos com base idêntica (*desenbista*, *tecedeira* e/ou *tecelão*). Estes dados sugerem uma provável maior frequência de utilização de sufixos em detrimento de outros disponíveis na L1, sendo de destacar o uso bastante produtivo do sufixo *-dor*.

Através dos resultados apresentados pelo nosso estudo, podemos distinguir os seguintes grupos de sufixos consoante o seu nível de produtividade e independentemente do ano escolar:

Sufixos + Produtivos	Sufixos - Produtivos	Ø
-dor/-dora -ista -nte	-ão -deira, -ino/-ina -deiro,-eiro/-eira -ã,-ona, -or/-ora	-sor/-sora -tor/-tora

Quadro 3: Produtividade sufixal em Agentivos deverbais dicionarizados (independentemente do ano escolar)

No que respeita os agentivos deverbais regulares não dicionarizados, também é possível observar variações de produtividade de sufixos utilizados por ano escolar e faixa etária, o que se poderá comprovar pela leitura do seguinte quadro-síntese:

Ano Escolar/ Idade	2º ano (6-7)	3º ano (8-9)	4º ano (9-10)	6º-9º ano (11-16)
Sufixos + Produtivos	-dor/-dora	-eiro	-or/-ora -eiro	-ista/-nte -deiro
Sufixos + Produtivos	-eira/-deira/-ão/-ona			
Sufixos - Produtivos	Ø ⁵³	-ino/-ina/-sor		
	-ã/-tor/-tora/-sora			

Quadro 4: Produtividade sufixal em Agentivos Deverbais regulares (não dicionarizados) por ano escolar.

A título ilustrativo, seguem-se alguns produtos agentivos identificados no nosso estudo que permitem atestar a variação de produtividade sufixal por ano escolar e faixa etária:

⁵³ Ø Significa, de acordo com o estudo de capô realizado, que os alunos do 2º ano não utilizaram os sufixos -ino, -ina e/ou -sor. A utilização destes sufixos por alunos apenas se verificou a partir do 3º ano escolar em diante.

Ano escolar/ Idade	Produtos agentivos deverbais regulares não dicionarizados
2º ano (6-7)	*? <i>assistidor</i> , *? <i>figurador</i> , *? <i>intrigadora</i> , *? <i>remetedora</i> ...
3º ano (8-9)	*? <i>baileiro</i> , *? <i>socorreiro</i> , *? <i>bailino</i> , *? <i>cantino</i> ...
4º ano (9-10)	*? <i>bailora</i> , *? <i>figurora</i> , *? <i>pedeiro</i> , *? <i>remeteiro</i> ...
6º ano (11-16)	*? <i>assistinte</i> , *? <i>socorrente</i> , *? <i>bailista</i> , *? <i>separista</i> , *? <i>comentadeiro</i> , *? <i>escrevedeiro</i> ...
9º ano (14-16)	*? <i>assistista</i> , *? <i>tecista</i> , *? <i>copiadeiro</i> , *? <i>pedincheiro</i> , *? <i>copiante</i> , *? <i>desenhante</i> ...

Quadro 5: Exemplos de variação da produtividade sufixal por ano escolar

3. Conclusões do estudo

O nosso estudo empírico aponta para a ocorrência de um uso sufixal diferenciado durante o percurso escolar, sendo de salientar uma notória diferenciação de padrão de comportamento entre alunos do 2º ano e os dos anos escolares subsequentes. Efetivamente, entre o 2º e o 3º ano ocorreu uma diminuição drástica da prática de geração de produtos considerados impossíveis, de acordo com os padrões de formação de palavras do Português Europeu atual.

Sugerimos uma possível convergência de fatores-chave que explicitam esta mudança de comportamentos.

Assim, a escolarização das crianças permite facilitar uma maior exposição da criança a formas dicionarizadas e a aquisição de todo um conjunto de experiências propícias ao desenvolvimento da competência metalinguística (aplicada a componentes distintos da nossa língua e sua gramática), em função de critérios como a idade e, simultaneamente, o grau de escolarização. Estratégias de promoção de hábitos de leitura e exercícios de funcionamento da língua são desenvolvidos, sobretudo, a partir do 3º ano escolar e intensificados nos anos escolares subsequentes, favorecendo obviamente o processo de aperfeiçoamento de operacionalização do subsistema de memória declarativa e, conseqüentemente, reforçando o efetivo conhecimento lexical na L1 por parte da criança ou do jovem.

Por outro lado, é de salientar que as construções possíveis ocorrentes em anos iniciais do ensino básico, resultantes provavelmente da ativação do subsistema de memória procedimental, tendem a ser abandonadas com maior convicção por indivíduos que frequentam o 4º ano em diante, dando estes primazia ao uso dos itens dicionarizados correspondentes, entretanto disponíveis na memória declarativa. Este facto leva-nos a comprovar uma quebra da criatividade em prol da componente convencional da língua, dado que ao longo do percurso escolar a criança/o jovem aprende a estabilizar combinatórias entre base e sufixo agentivo e adquire uma maior eficiência no seu uso.

Neste sentido, o nosso estudo apresenta fortes indícios de um crescente predomínio da memória declarativa sobre a memória procedimental na formação de léxico na L1.⁵⁴

Referências

- Almeida, Ana Lúcia Andrade da Silva (2009). Produção de Nomes Agentivos Deverbais em Português Língua Materna (L1) por Alunos do Ensino Básico. Dissertação de mestrado em Linguística Aplicada, Universidade de Coimbra.
- Basílio, Margarida (1995). O fator semântico na flutuação substantivo/adjetivo em português. In: Jürgen Heye (org.). *Flores Verbais – uma homenagem linguística literária para Eneida do Rego Monteiro Bomfim*. Rio de Janeiro: Editora 34, 177-192.
- Booij, Geert E. (2005). Polysemy and construction morphology. In: Antonio Fábregas, Irene Gil e Paul O'Neill (eds.). *Cuadernos de Lingüística*, Vol. XII. Madrid: Instituto Universitario Ortega y Gasset, 17-26.
- Clark, Eve L. (2004). How language acquisition builds on cognitive development. *Trends in Cognitive Sciences* 8:10, 472-278.
- Clark, Eve V. (2003). *First Language Acquisition*. Cambridge, New York: Cambridge University Press.
- Clark, Eve L. & Cohen, Sophia R. (1984). Productivity and memory for newly formed word. *Journal of Child Language* 11, 611-625.
- Leiria, Isabel (2006). Léxico, aquisição e ensino do português europeu língua não materna. Dissertação de doutoramento em Linguística Aplicada. Lisboa: Fundação para a Ciência e Tecnologia e Fundação Calouste Gulbenkian.
- Lo Duca, Maria Giuseppa (1990). *Creatività e regole, Studio sull'acquisizione della morfologia derivativa dell'italiano*. Bologna: Il Mulino.
- Rio-Torto, Graça M. (1998). *Morfologia derivacional. Teoria e aplicação ao português*. Porto: Porto Editora.

⁵⁴ Para mais informações sobre o tema tratado neste estudo, consulte-se Almeida (2009).

- Rodrigues, Alexandra F. S. (2006). Formação de substantivos deverbais sufixados em português. Dissertação de doutoramento na área de Línguas e Literaturas Modernas, especialidade de Linguística Portuguesa, Universidade de Coimbra.
- Silva, Ana Lúcia Andrade da (2007). Reflexões acerca de léxico e morfologia derivacional em língua materna: análise do processo de formação de agentivos deverbais atestados e não atestados em português. Ms, Universidade de Coimbra.
- Ullman, Michael T. (2003). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231-270.
- Ullman, Michael T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistics Research* 30: 1, 37-69.

ANEXO I

Exercício escrito de testagem

1. Com as palavras da coluna A e os sufixos da coluna B constrói todas as palavras que são possíveis em português.

86

Parte 1

Coluna A	Coluna B
1. aldrabar	- ante / - ente / - inte ⁵¹
2. assistir	- ão / - ã / - ona
3. bailar	- eiro / - eira
4. campear	- ino / - ina
5. cantar	- ista
6. comentar	- or / - ora
7. consumir	- dor / - dora
8. contar	- deiro / - deira
9. copiar	- sor / - sora
10. defender	- tor / - tora
11. desenhar	
12. escrever	

Parte 2

Coluna A	Coluna B
13. figurar	- ante / - ente / - inte
14. intrigar	- ão / - ã / - ona
15. namorar	- eiro / - eira
16. pedir	- ino / - ina
17. projectar	- ista
18. remeter	- or / - ora
19. separar	- dor / - dora
20. servir	- deiro / - deira
21. socorrer	- sor / - sora
22. tecer	- tor / - tora
23. vender	

Exemplo: pescar – *pescador; pescadora; pesqueiro; pesqueira.*

⁵⁵ Embora na verdade o sufixo seja -nte, colocamos aqui as formas -ante, -ente e -inte de modo a tornar mais compreensível o seu uso.

ANEXO II

Guião de Perguntas

1º ciclo de ensino básico (turma de 2º ano)

Identificação da palavra formada	Neste exercício, construístes várias palavras. Quais são as partes que usaste para construir a palavra x?	O que significa a palavra x? Conheces outro(s) sentido(s) da mesma palavra?	Encontras alguma diferença de sentido entre a palavra x e a palavra y? Qual?	Costumas usar esta palavra na tua escrita e/ou fala diária? Já a ouviste a ser utilizada antes? E em que situação?
	Qual é a classe gramatical da palavra base? É um verbo ou um nome?	Qual é a classe gramatical da palavra final? É um verbo ou um nome?	Outras observações (ex. Lembras-te de outras palavras possíveis que não escreveste? Quais?)	

87

1º ciclo de ensino básico (turmas de 3º e 4º anos)

Identificação da palavra formada	Neste exercício construístes várias palavras. Quais são as partes que formam a palavra x? Sabes como se chamam?	Qual é a classe gramatical da palavra que usaste para construir a nova? É um verbo, nome ou adjectivo?	Qual é a classe gramatical da palavra que formaste? É um verbo, um nome ou um adjectivo?	O que significa a palavra que formaste? Conheces outro(s) sentido(s) da mesma palavra?
	Encontras alguma diferença de sentido entre a palavra x e a palavra y? Qual?	Utilizas esta palavra na escrita/fala diária? Já a ouviste a ser utilizada antes? E em que situação?	Outras observações: (ex. Lembras-te de outras palavras possíveis que não identificaste? Quais?)	

2º ciclo de ensino básico (turma de 6º ano)

88

Identificação da palavra formada	Neste exercício construístes várias palavras. Quais são as partes que formam a palavra x? Sabes como se chamam?	Qual a classe gramatical da palavra que usaste para construir a nova? É um verbo, um nome ou um adjetivo?	Qual a classe gramatical da palavra que formaste? É um verbo, um nome ou um adjetivo?	Que significado(s) tem a palavra que formaste? Conheces outro(s) significado(s) da mesma palavra?
	Existe alguma diferença entre o significado da palavra x e o da palavra y? Explica essa diferença. Na tua opinião, que sentido é dado pela forma sufixal que utilizaste na palavra x?	Utilizas esta palavra na tua escrita/ fala diária? É frequente ou não a sua utilização em discurso? Já a ouviste a ser utilizada antes e em que situação?	Outras observações: (ex. Lembra-te de outras palavras possíveis que não foram registadas? Quais?)	

3º ciclo de ensino básico (turma de 9º ano)

Identificação da palavra formada	Quais os constituintes internos da palavra x? Como se designam?	Qual a classe gramatical da palavra base?	Qual a classe gramatical da palavra x?	Que significado(s) tem a palavra final? Conheces outro(s) significado(s) da mesma palavra?
	Consideras que existe diferença de significado entre a palavra x e a palavra y? Explica essa diferença de sentido. Na tua opinião, qual o sentido atribuído pela forma sufixal que utilizaste na palavra x?	Utilizas esta palavra no teu discurso escrito/oral diário? Na tua opinião, qual a frequência de uso desta palavra em discurso: muito frequente/ ocasionalmente/ raramente/ nunca? Já a ouviste a ser utilizada antes e em que situação?	Outras observações: (ex. Lembra-te de outras palavras possíveis que não foram registadas? Quais?)	

PADRÕES NA APRENDIZAGEM DA ESTRUTURA
DE CONCORDÂNCIA EM PORTUGUÊS LÍNGUA ESTRANGEIRA

1. Introdução

A concordância em português apresenta inúmeras dificuldades para os alunos estrangeiros que estão a aprender português. Nesse sentido, propomos explorar neste trabalho algumas estruturas de concordância em português e, em seguida, com base na análise de um conjunto de produções escritas por alunos estrangeiros, verificar se existem padrões regulares na aprendizagem desta estrutura independentemente das línguas que conheçam e da sua proficiência.

Deste modo, tomámos como ponto de partida para a nossa investigação as seguintes questões: é possível definirmos um padrão na aprendizagem da estrutura de concordância em português por alunos estrangeiros, sem ter em conta o conjunto de línguas conhecidas pelo aprendente e respetivo nível de proficiência? Existirá, de facto, uma classe de palavras onde os alunos apresentem maior dificuldade em seguir as regras de concordância previstas para o português? E em que parâmetros (Número, Género ou Pessoa) apresentarão os alunos mais dificuldade?

Na tentativa de encontrar respostas para estas perguntas, analisámos um conjunto de produções escritas de alunos estrangeiros que frequentaram cursos de português para estrangeiros na Faculdade de Letras da Universidade de Coimbra entre maio de 2009 e maio de 2010 (Martins 2011). O *corpus* que serve de base para o nosso trabalho é constituído por um conjunto de textos escritos sobre as mais variadas temáticas livres (passatempos, transportes, comidas favoritas, apresentar-se, etc.), de que aleatoriamente

foram selecionados 35, com origens diferentes (turmas ERASMUS LP II, LP III, Cursos de Férias, Curso Anual)⁵⁶.

90 Procedeu-se à seleção aleatória das produções, uma vez que se pretende verificar, neste estudo, quais as principais dificuldades dos aprendentes em relação à concordância nominal, sem estabelecer uma correlação direta com o nível de proficiência linguístico ou a língua materna dos aprendentes. Feito o trabalho de recolha, procedemos à sua análise, atentando nos desvios de concordância que cada texto apresentava, de modo a verificar que dificuldades os alunos apresentam na estrutura de concordância, mais especificamente na concordância nominal. Assim, verificámos de que modo os aprendentes estabelecem a concordância entre os elementos dos constituintes nominais e em que estruturas os desvios são mais recorrentes. Devido à natureza das produções, é possível identificarmos, inclusive através do processo de autocorreção do próprio aluno, padrões de marcação da concordância.

2. A estrutura da concordância nominal

A concordância é uma estrutura linear para o português, i.e., consiste na colocação reiterada das marcas de concordância em todos os elementos da frase, mas não o é para outras línguas onde de facto não encontramos uma correspondência direta com esta estrutura. Define-se como um conjunto alargado de regras que permitem aos falantes a boa formação de frases, e pode ser de dois tipos: concordância verbal e concordância nominal.

Para o presente trabalho, centrámo-nos na estrutura da concordância nominal, ou seja, no tipo de concordância que ocorre entre sintagmas (predicativo e sujeito) e dentro dos Sintagmas Nominais.

Neste sentido, apoiámo-nos na definição de concordância nominal proposta por Mateus *et. al.* (2003) que identificam nesta estrutura a presença

⁵⁶ Os textos utilizados foram selecionados a partir de um conjunto alargado de produções que integram o “Corpus de Produções Escritas de Aprendentes de PL2” (PEAPL2) (coord. Cristina Martins), projeto iniciado em junho de 2008, no Centro de Estudos de Linguística Geral e Aplicada (CELGA), da Faculdade de Letras da Universidade de Coimbra. Disponível online em: <http://www.uc.pt/fluc/rcpl2/>.

de um sistema de valores que integra um conjunto de formas que se opõe entre si: o Parâmetro de Número (NUM) que tem dois valores opostos (Singular vs. Plural); e o Parâmetro de Género (GEN) igualmente com dois valores diferentes (Masculino vs. Feminino). Para marcar estes valores em português acrescentam-se, geralmente, índices temáticos no final das palavras. O Português é, por isso, uma língua que assinala o GEN e o NUM por marcação morfológica. As palavras que recebem estes morfemas são designadas por palavras variáveis que flexionam relativamente a um determinado atributo (NUM ou GEN). Assim, admitindo a marcação de GEN e/ou NUM no nome, centrámo-nos no levantamento dessas marcas nas palavras concordantes à esquerda e à direita do nome, ou seja, nos determinantes (artigos e quantificadores) e adjetivos.

91

Deste modo, verificamos que, no interior dos Sintagmas Nominais, a concordância se estabelece entre:

- i) O Núcleo Nominal e as expressões de determinação e quantificação que o antecedem – os artigos definidos ou indefinidos, os determinantes demonstrativos e possessivos e os quantificadores;
- ii) O Núcleo Nominal e os adjetivos integrados em Sintagmas Adjetivais que funcionam como modificadores ou apostos. (Mateus *et. al.* 2003)

3. Análise dos dados

3.1 A Marcação de Número

Feitas estas considerações gerais em torno da estrutura de concordância nominal, importa agora verificar que observações retirámos depois de analisar os dados do *corpus*. No que diz respeito à marcação de Número, constatámos que este parâmetro apresenta poucos casos problemáticos, isto é, comparado com os problemas que a marcação de Género apresenta, este parâmetro manifesta um número pouco significativo de problemas. Ainda assim, apesar dos poucos dados obtidos para a marcação deste parâmetro, foi possível verificar que os aprendentes apresentam maior dificuldade em

fazer a correta marcação do NUM na relação substantivo-adjetivo. Alguns textos apresentam desvios de concordância como⁵⁷:

- (1) **Informações novo*;
- (2) **Pessoas desonesta*;
- (3) **Os olhos verde*;
- (4) **Doces portugues*

92

Em boa verdade, são exemplos muito escassos, mas permitem identificar a tendência do aluno estrangeiro para marcar o NUM no nome e deixar o adjetivo na forma não-marcada, ou seja, no SING. Talvez, pelo facto de o substantivo ser o Núcleo Nominal é, no entender destes alunos, a palavra que deve sofrer essa marca de Plural. Tal padrão permite-nos constatar que o aprendente adquiriu a marca de NUM e a conseqüente distinção sintática entre o Singular/Plural, mas não domina ainda a noção de concordância no interior do Grupo Nominal. Estes casos são representativos de como o sentido é mais importante do que a forma. Estudos realizados nesta área revelam que durante o processo de aprendizagem de uma língua estrangeira os alunos tendem a “selecionar” a informação dada pelo *input* da Língua-Alvo processando, nas primeiras fases, unidades linguísticas sentidas com maior valor *comunicativo*, deixando de lado os itens gramaticais (cf. VanPatten 1996).

3.2 A Marcação de Género

No que concerne à marcação de Género, identificámos, com base na análise do *corpus*, que os desvios de marcação são muito frequentes. Neste ponto da investigação, questionámo-nos sobre qual seria a classe gramatical que apresentaria um maior número de casos de desvios de marcação de GEN. E neste sentido pudemos constatar que é a classe gramatical dos determinantes a que apresenta maiores dificuldades, precisamente a classe

⁵⁷ Cf. Quadro 1.

	Código da produção ⁵⁸	Desvio na marcação de NUM no DET	Desvio na marcação de NUM no NOME	Desvio na marcação de NUM no ADJ	Desvio na marcação de NUM no ADV
1	UC. ER.LPIII.F. 06.09.01.52.2L			(...) umas conjuntos de estudantes, mas não são tão popular (...)	
2	UC. ER.LPIII.F. 06.09.06.52.2L				(...) amigos pertos (...)
3	UC.CA.I.A. 01.10.04.69.3Q			(...) informações novo (...)	
4	UC.CF.EB. 02.07.09.02.33.1J	(...) três semana (...)			
5	UC.ER.LPII.F. 06.09.10.1.1A			(...) umas pessoas desonesta (...)	
6	UC.ER.LPII.F. 06.09.13.1.1A			(...) os olhos verde (...)	
7	UC.ER.LPII.F. 06.09.17.1.1A	O meus pais (...)		(...) cabelos longos e claro (...)	
8	UC.ER.LPII.A. 12.09.01.55.2M			(...) doces portugues (...).	

Quadro 1 – Desvios na marcação de NUM

de palavras que permite identificar o género do nome (o núcleo do sintagma). Os aprendentes, geralmente, não fazem a correta concordância do Determinante com o Nome e colocam-no preferencialmente na forma não marcada, ou seja, no Masculino. Um dos fatores que pode condicionar esta tendência pode ser o facto de esta ser uma classe de palavras que, fazendo parte da estrutura interna do Sintagma Nominal, não apresenta, no entender do aprendente, muita importância do ponto de vista comunicativo. De facto, estes resultados corroboram os argumentos defendidos por Bill VanPatten

⁵⁸ O código da produção corresponde a um código geral que abrange informações distintas sobre o aprendente e o texto produzido. Para uma descrição detalhada deste código, veja-se <http://www.uc.pt/fluc/rcpl2/metodologia/>.

(1996), ou seja, no processamento do *input* de uma língua estrangeira, os aprendentes preferem processar itens lexicais do que a processar itens gramaticais, desenvolvendo um conjunto de estratégias que lhes permitam ultrapassar as várias dificuldades.

94

Desta forma, que hipóteses poderemos levantar que justifiquem estes desvios de marcação? Por um lado, se a LM do falante apresenta na palavra equivalente um GEN diferente, isso pode condicionar o aprendente na marcação do GEN da palavra portuguesa. Por outro lado, a própria natureza da palavra pode condicionar a opção do aprendente, ou seja, um aluno estrangeiro durante o seu processo de aprendizagem da marcação do parâmetro de GEN distingue mais facilmente, numa primeira fase, o valor de GEN das palavras cuja distinção de GEN Masculino/Feminino corresponda a uma diferença de sexo, do tipo *o menino/a menina*, ou seja, palavras em que a marcação de GEN é semanticamente pertinente. No entanto, em palavras que não apresentem essa correspondência, já surgem mais problemas. Assim, numa segunda fase, o aluno vai assimilar que todas as palavras terminadas em *-o* são do GEN Masculino e que as palavras que terminam em *-a*, são do GEN Feminino. Nesse caso, quando se depara com palavras que não seguem esta “regra” que generalizou, tende a errar a marcação de GEN (ex. *as problemas*)

Assim sendo, analisámos o *corpus* a partir destas variáveis na tentativa de encontrar um padrão regular de desvios de marcação de GEN, e que nos permitiriam aferir um padrão na aprendizagem da estrutura de concordância. Deste modo, a generalização da regra, característica das primeiras fases de aprendizagem de uma língua, aparece no Determinante⁵⁹:

(5) **Alguns modificações;*

(6) **No capital;*

(7) **Esta dia;*

(8) **O possibilidade;*

(9) **Estes atividades;*

(10) **Os paredes;*

(11) **As suas costumes;*

⁵⁹ Cf. Quadro 2.

- (12) **Os árvores;*
- (13) **todas as problemas;*
- (14) **o seu cerveja;*
- (15) **num empresa*

Com base nestes exemplos, constatou-se que: i) quando a palavra termina em *-a*, o aluno, por ter assimilado que todas essas palavras pertencem ao GEN Feminino, coloca o artigo também nesse GEN (**esta dia/*todas as problemas*); ii) em palavras que não terminam nem em *-o* nem em *-a*, o aluno coloca o artigo preferencialmente na forma não-marcada, ou seja, na forma masculina.

Durante a análise dos dados encontramos casos muito particulares, como os seguintes: a) **Esto bairro*; b) **Esto bar*; c) **Estos transportes*; d) **Esta dia*. Nestes exemplos, parece-nos ter ocorrido um fenómeno muito interessante: os aprendentes mostram que assimilaram a 1ª fase de marcação do GEN, que é a de saber que o nome tem GEN inerente e que ele tem de ser marcado nas palavras concordantes - em especial no Determinante. Mas ainda não aprenderam que as palavras que não terminam em *-o* não são necessariamente masculinas, e que o Determinante demonstrativo assinala o masculino com *este / estes*. Dessa forma, uma explicação possível para estas produções é a de que optam por escrever o Determinante também com a terminação *-o* para assinalar que se trata de uma forma do Masculino. O confronto com o exemplo **esta dia* corrobora esta conclusão: por se tratar de uma palavra terminada em *-a*, logo sentida como sendo do GEN feminino, apresenta a marcação de Feminino *esta*. Para atestar esta hipótese, em futuras investigações, seria relevante analisar um conjunto de produções escritas do mesmo falante, inclusive registos orais.

Existe ainda um outro tipo de desvios de marcação, que demonstram como, durante a aprendizagem, os aprendentes generalizam que a marca final da palavra determina o seu GEN. Esses casos são os seguintes:

- (16) **Os semestros;*
- (17) **Mundo um bocado mais alegre;*
- (18) **Estou contenta* (trata-se de um aprendente Feminino)⁶⁰

⁶⁰ Cf. Quadro 2.

Assim em (16), o aprendente em vez de *semestres* colocou *semestros* (para assinalar que se trata de um nome Masculino); em (17) também o adjetivo uniforme *alegre* sofreu essa alteração (optou por colocar *alegro e não alegre* para marcar o Masculino e concordando com o nome *mun*do). Em (18), como está a falar de si, de um determinado estado psicológico em que se encontra, a aluna optou por marcar o predicativo *contente* com *contenta* (adicionando-lhe a marca final, que associa ao GEN feminino *-a*).

	Código da produção	Desvio na marcação de GEN no DET	Desvio na marcação de GEN no NOME	Desvio na marcação de GEN no ADJ
1	UC. ER.LPIII.F. 06.09.01.52.2L	(...) alguns modificações (...) (...) umas conjuntos (...)		
2	UC. ER.LPIII.F. 06.09.07.77.3T			(...) população simpático (...)
3	UC. ER.LPIII.F. 06.09.09.33.1J	(...) um outro coisa (...)		
4	UC. ER.LPIII.F. 06.09.11.33.1J	(...) no capital (...) (...) esta dia (...)		
5	UC. ER.LPIII.F. 06.09.12.33.1J	(...) o possibilidade (...) (...) estes atividades (...)		
6	UC. ER.LPIII.F. 06.09.12.77.3T	(...) esto bairro (...)		
		(...) Os paredes (...)		
		(...) o possibilidade (...)		
		(...) esto bar (...)		
		(...) outros coisas (...)		
7	UC. ER.LPIII.F. 06.09.13.33.1J			(...) um pequeno dor (...)
8	UC. ER.LPIII.F. 06.09.22.52.2L	(...) o seu cerveja (...)		
9	UC. ER.LPIII.F. 06.09.23.52.2L	(...) as suas costumes (...)		
10	UC. ER.LPIII.F. 06.09.24.33.1J	(...) os árvores (...)		
11	UC. ER.LPIII.F. 06.09.28.33.1J		(...) os semestros (...)	
		(...) num empresa (...)		
12	UC. ER.LPIII.F. 06.09.29.33.1J	(...) alguns atividades (...)		
		(...) alguns vezes (...)		
		(...) outros vezes (...)		
13	UC. ER.LPIII.F. 06.09.30.33.1J			(...) composições bonitos (...)
				(...) mun do um bocado mais alegro (...)
14	UC. ER.LPIII.F. 06.09.31.33.1J	(...) um outra passatempo é a fotografia (...)		

15	UC. ER.LPIII.F. 06.09.32.77.3T			(...) uma casa mais bonito (...)
16	UC. ER.LPIII.F. 06.09.33.52.2L	(...) Os vantagens (...)		Os vantagens divertidos (...)
17	UC.CA.I.A. 01.10.01.33.1J	(...) o viagem (...)		
18	UC.CA.I.A. 01.10.04.69.3Q	(...) um vantagem (...)		
		(...) na parque (...)		
		(...) o organização (...)		
		(...) da dia do voo (...)		
19	UC.CF.EB. 01.07.09.01.33.1J			(...) água do mar é muito limpo .
20	UC.CF.EB. 02.07.09.01.33.1J	(...) estos lugares (...)		
21	UC.CF.EB. 02.07.09.05.33.1J	(...) no Internet (...)		
22	UC.ER.LPII.F. 06.09.08.1.1A	Um parte (...)		(...) cabelo loira (...)
23	UC.ER.LPII.F. 06.09.10.1.1A	Outro cidade (...)		
24	UC.ER.LPII.A. 12.09.08.1.1A	(...) dois semanas (...)		
25	UC.ER.LPII.A. 12.09.13.75.3S	(...) todas as problemas (...)		
26	UC.ER.LPII.A. 12.09.15.75.3S	(...) estos transportes (...)		(...) um viagem muito pessado (...)
		(...) um viagem (...)		
		(...) o viagem (...)		
27	UC.ER.LPII.A. 12.09.20.1.1A			(...) estou contenta (...)
28	UC.ER.LPII.A. 12.09.25.1.1A	(...) o cidade (...)		(...) um cidade pequeno mas cheio .
		(...) um cidade (...)		

Quadro 2 – Desvios da marcação de GEN⁶¹

4. Considerações finais

Em suma, todos estes dados nos permitem concluir que é possível definirmos um padrão na aprendizagem da estrutura de concordância por alunos de português como língua não-materna e mostram igualmente que este é um processo constituído por duas fases distintas, sendo a primeira a da generalização da regra segundo a qual todos os nomes têm género

⁶¹ Ver nota 58.

marcado no Determinante e a segunda a da generalização da regra: “nomes em -o são Masculinos e nomes em -a são Femininos”.

98 Igualmente, pelos exemplos analisados se verifica que os aprendentes tendem a assinalar os determinantes, respetivamente, com -o e -a, de acordo com a terminação que encontra no nome (como em *esta *dia*). Nos restantes casos, há a tendência de o aluno para utilizar a forma masculina do determinante, i.e., na forma não-marcada. Além disso, os dados obtidos deste trabalho permitem confirmar a ideia de que o sentido tem prioridade sobre a forma na aprendizagem, tal como defendem os trabalhos de VanPatten (1996). Geralmente, os informantes tendem a “ignorar” as unidades gramaticais, valorizando, assim, as unidades lexicais sentidas como itens de maior valor “comunicativo”.

Referências

- Corder, S. P. (1992). La importancia de los errores del que aprende una lengua segunda. In: J.M. Liceras (1992). *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua*. Visor: Madrid, 31-40.
- Corbett, Greville G. (1991). *Gender*. Cambridge: Cambridge Univ. Press.
- Cunha, Celso e Luís Filipe Lindley Cintra (1984). *Nova Gramática do Português Contemporâneo*. Lisboa, Livraria Sá da Costa.
- Franceschina, Florencia (2005). *Fossilized second language grammars - the acquisition of grammatical gender*. Amsterdam / Philadelphia, John, Benjamins.
- Martins, Cristina (no prelo). O *Corpus* de Produções Escritas de Aprendentes de PL2 (PEAPL2/CELGA). Caracterização e desenvolvimento de uma infra-estrutura de investigação. In: Cristina Flores (ed.). *Português Língua Não Materna: Investigação e Ensino*. Lisboa, Lidel.
- Mateus, Maria Helena et al. (2003). *Gramática da Língua Portuguesa*. Lisboa, Caminho.
- Van Patten, Bill (1996). *Input Processing and Grammar Instruction*. Norwood, New Jersey: Ablex Publishing Co.

SECÇÃO 4:
SEMÂNTICA

(Página deixada propositadamente em branco)

OPOSIÇÕES ASPETUAIS EM PORTUGUÊS LÍNGUA SEGUNDA.
O CASO DOS FALANTES DE RUSSO EM CONTEXTO DE IMERSÃO

1. Introdução

O trabalho de investigação linguística levado a cabo procurou avaliar o reconhecimento de algumas formas de expressão do aspeto em Português por parte de falantes de Russo em contexto de imersão, isto é, inseridos na sociedade portuguesa, mas não necessariamente com instrução formal de Português. Para tal, foram consideradas, por um lado, algumas oposições aspetuais consideradas relevantes, descritas de acordo com as perspetivas teóricas de Comrie (1976) e de Cohen (1989). Por outro lado, para a investigação propriamente dita, foi feito um inquérito, em que avaliámos o reconhecimento de algumas formas de expressão do aspeto em Português por parte de falantes de Russo em contexto de imersão, para apurarmos facilidades ou dificuldades que pudessem revelar a respeito da aquisição do aspeto em Português. Procurámos, portanto, averiguar se estes informantes, possuidores de um conhecimento aspectual prévio, reconheceriam determinados valores aspectuais numa configuração linguística distinta. Além disso, relacionámos os juízos dos informantes sobre configurações do aspeto em Português com a seguinte condicionante - o contexto real de comunicação.

Dos resultados obtidos evidenciam-se possíveis vantagens propiciadas pelo contexto linguístico e social para o reconhecimento de valores e de oposições aspectuais do Português, por contraste com uma menor relevância da instrução formal. Da mesma forma, em alguns casos, o conhecimento da língua materna dos informantes mostra-se mais influente e decisivo do

que a aprendizagem formal da língua. Por outro lado, particularidades aspectuais do Português, que não encontram correlações com a língua materna, não são claras aos olhos dos informantes. Por fim, observam-se dificuldades que os informantes têm em discernir semanticamente frases com configurações aspectuais, sobretudo se não estiverem acompanhadas por outros elementos contextuais.

102

Concluiu-se que, apesar de estarem contidas noções aspectuais no sistema da sua língua materna, e independentemente de terem tido instrução formal de Português, os falantes de Russo revelam algumas dificuldades em reconhecer formas de expressão do aspeto expressas somente pela morfologia do verbo (nomeadamente o Pretérito Imperfeito). Pelo contrário, formas frequentemente utilizadas no contexto comunicativo quotidiano parecem ser mais perceptíveis, o que indicia uma relevância do contexto de imersão.

2. Aspeto(s)

Existe uma dificuldade em definir universalmente o termo *aspeto* no que respeita à sua expressão formal em sistemas verbais de línguas geneticamente afastadas, como o Português e o Russo, contrariamente à definição semântica que, em geral, é dada ao conceito: o modo como perspectivamos a estrutura temporal interna de uma situação (Comrie 1989: 13). Na verdade, alguns linguistas consideram que não existe aspeto senão em línguas eslavas, ao passo que outros caracterizam como aspectuais determinadas particularidades semânticas não introduzidas pelas outras categorias verbais (Cohen 1989: 7). Por essa razão, não existe uniformidade terminológica no uso do termo: é utilizado presentemente para definir oposições semânticas gerais, assim como, em seu sentido restrito, para exprimir particulares oposições gramaticais baseadas em determinadas distinções semânticas presentes em línguas como o Português (Comrie 1976: 6).

A tradicional atribuição de dois tipos de aspecto diferentes presentes em Português e Russo – aspecto gramatical e aspecto lexical, respectivamente – deve-se ao facto de, na primeira língua, a perspetivação da estrutura

temporal interna da situação poder ser marcada gramaticalmente, pelos tempos verbais Pretérito Perfeito e Pretérito Imperfeito, que expressam a distinção aspectual – perfeito/imperfeito – respeitante a uma situação acabada/inacabada (ou com limites definidos/não definidos); na segunda língua, por seu turno, a perspetivação da estrutura temporal interna da situação é marcada lexicalmente, através de processos de formação de palavras como a afixação, sendo a distinção aspectual perfeito/imperfeito a base definitória do sistema verbal; na terminologia do sistema russo, estes dois critérios de organização verbal são designados de aspeto perfeito e aspeto imperfeito.⁶²

Contudo, a designação de aspeto gramatical atribuída ao Português, parcialmente adequada, releva unicamente os processos gramaticais, de que fazem parte formas verbais (PP, PI, Pretérito Perfeito Composto), verbos de operação aspectual e adverbiais, menosprezando, assim, outros processos, lexicais, relativos à expressão do aspeto, como a classe aspectual do predicador e a formação de palavras.⁶³

No que toca a aproximações no campo do aspeto, as duas línguas partilham algumas funções aspetuais que são desempenhadas pelo tempo Presente (russo e português), e por outros tempos verbais (o Pretérito de aspeto imperfeito e o PI português; o Pretérito de aspeto perfeito e o Pretérito Perfeito Simples), o que poderá, de certa forma, facilitar o reconhecimento numa L2 (o Português) de valores aspetuais presentes na sua língua materna, ainda que estejam configurados diferentemente.

⁶² Em relação à natureza lexical da expressão do aspeto em Russo, é usado o termo *Aktionsart* (ou modos de ação) para representar a lexicalização da distinção aspectual, diferenciando-se do termo Aspeto como conceito de gramaticalização de uma distinção semântica. De acordo com esta abordagem, são invocadas razões de ordem lexical como forma de distinguir os dois conceitos (*vide* Comrie 1976: 6-7, nota 4; Cohen 1989: 31).

⁶³ Existem na língua portuguesa afixos na estrutura interna dos verbos que transmitem informação aspetual, por exemplo, o verbo *saltitar*, cujo interfixo (-it) tem o valor aspetual iterativo, ou o verbo *ensurdecer*, que contém afixos com valor incoativo (*en-...-ec*) (Silva 2009: 33).

2.1 Oposições aspetuais (perfetivo/imperfetivo; pontual/durativo)

104 Sobre a oposição dicotómica entre as noções aspetuais de perfetividade e de imperfetividade, Comrie (1976: 21-24) circunscreveu os referidos conceitos às seguintes aceções: perfetividade envolve a falta de referência explícita à constituição temporal interna de uma situação, enquanto imperfetividade caracterizar-se-á, aparentemente, por uma referência explícita à estrutura interna temporal de uma situação, e por uma compatibilidade com a noção de perfetividade.

Estas aceções esclarecem algumas ambiguidades generalizadas sobre critérios semânticos associados a caracterizações inadequadas dos conceitos (Comrie 1976: 16). Um dos critérios é o da duração da situação, que não é exclusiva da esfera da imperfetividade. Apesar de o conceito de perfetividade perspetivar a situação como um todo, sem explicitar a estrutura interna da situação, poderá incluir a expressão de duração da situação. De facto, no Russo, é-nos veiculada informação aspetual, distribuída por meios lexicais, relativa a fases da estrutura interna da situação, mesmo quando se aborda a situação como um todo, como é o caso de formas pertencentes ao aspeto perfeito⁶⁴.

Ainda para o linguista (1976: 18-19), outros termos que frequentemente se associam à função de perfetividade – pontual e completa – não são totalmente adequados a todas as expressões de perfetividade, no sentido em que essas vertentes semânticas colidem de alguma forma com a própria delimitação do conceito. Tendo em conta o exemplo russo, numa forma perfetiva, a referência não explícita sobre a constituição interna da situação não invalida que a distribuição lexical do significado referencie explicitamente uma duração ou uma sucessão de fases de um intervalo de tempo, ainda que a situação seja encarada como um todo e de se tratar de uma situação momentânea. O termo pontual não se aplica, portanto, em todas as formas perfetivas, já que reduz conceptualmente a situação a um único ponto excluindo a complexidade interna da situação. Também a designação de *completa* atribuída geralmente a uma forma perfetiva não é adequada, não podendo ser considerada quando se analisa a língua russa, uma vez

⁶⁴ Vide Silva (2009: 53, nota 66)

que a informação aspetual relativa ao *terminus* de uma situação contradiz a própria definição de perfeitividade, que por si só não dá conta da explicitação dos limites temporais internos da situação.

Em relação ao campo da imperfetividade, na esteira de Comrie, poderemos concluir que: a utilização do termo duração não será generalizável para a caracterização de um subtipo aspetual de imperfetividade (dado que a duração também poderá ser expressa, em Russo, através de meios lexicais, por algumas formas perfectivas; por conseguinte, não é exclusiva da expressão da imperfetividade); a pontualidade, traço semântico característico de situações que não têm estrutura interna, não será compatível com a imperfetividade.

105

Também a expressão da imperfetividade e de subtipos aspetuais adjacentes – habitual *vs* contínuo; progressivo *vs* não progressivo – varia consoante a língua em questão, estando a sua utilização relacionada com propriedades do conteúdo lexical, cujas características temporais internas caracterizam uma situação que envolverá uma certa duração ou uma sequência de fases. Assim, a imperfetividade, campo de expressão que explicita especificidades internas da constituição da situação, sustentado, naturalmente, pelo significado inerente de itens lexicais, obedece a critérios de organização de cada língua, sendo, pois, de difícil generalização.⁶⁵

3. Hipótese de trabalho: o reconhecimento de oposições aspetuais

Sendo o sistema verbal das línguas eslavas estruturado em torno da categoria Aspeto – um sistema aspectual definido pela oposição perfeitivo/

⁶⁵ No caso da construção aspetual progressiva do Português Europeu <estar a + infinitivo>, que apresenta características de duração e de incompletude, é-nos fornecido ainda um significado de estaticidade, o que contraria em parte a definição de progressividade dada por Comrie (1976: 35) – “the combination of progressive meaning and nonstativity meaning”. Repare-se que, de acordo com esta abordagem de Comrie, o significado aspetual progressivo apenas poderá ser expresso por verbos cujas características temporais, inerentes ao significado lexical, permitam o tratamento de uma situação não estática. Por isso, o facto de um verbo ser estático impede a sua expressão numa forma progressiva, já que a progressividade contradiz a noção de estaticidade. É neste ponto que esta construção progressiva se insere na definição de progressividade de Comrie, no sentido em que verbos estáticos não poderão ser expressos pelo operador aspetual progressivo. A esse propósito, confronte-se esta informação com a **nota 72** da presente comunicação, para se refletir sobre o fenómeno do desconhecimento das propriedades aspetuais inerentes a verbos estáticos por falantes de PLNM.

imperfeito⁶⁶ – procurámos saber se os inquiridos teriam facilidade em reconhecer esta oposição aspectual configurada em outro suporte, subordinada à categoria do *Tempo* e sustentada através da oposição entre o Pretérito Perfeito e o Pretérito Imperfeito.

Considerámos dois possíveis fatores de facilitação do reconhecimento de noções aspectuais: o contexto de imersão e a língua materna, ambos testados no inquérito.⁶⁷

3.1. O contexto de imersão

Em relação ao primeiro fator, surgiram as seguintes questões:

- (i) *O simples facto de estarem imersos linguisticamente na comunidade levará estes falantes a reconhecer e a adquirir o valor imperfeito contido em algumas formas verbais portuguesas?*
- (ii) *Tendo em conta uma suposta facilidade em reconhecer a noção de imperfetividade, será que se pode considerar que esse reconhecimento é resultante apenas do contexto de imersão? E que os inquiridos utilizariam essa noção e consequentemente a reconheceriam no Pretérito Imperfeito?*

Definimos, então, o contexto de imersão como: *o contexto social e linguístico no qual o aprendente adquire a L2 exclusivamente pela interação comunicativa*.⁶⁸ Trata-se de um contexto natural de aquisição em que a L2 é usada para propósitos comunicativos do dia a dia (a rua e o local de trabalho); consideramo-lo facilitador pela aprendizagem incidental focada na comunicação com falantes nativos. Por fim, é um fator de ordem externa, extralinguístico, que influencia/facilita a aquisição de L2 (em conjunto com outros fatores externos – interação e o *input*).

⁶⁶ Vide Comrie (1976: 4- 50) e Cohen (1989: 7- 31).

⁶⁷ Na dissertação de Mestrado que está na origem do presente trabalho (Silva 2009) esteve presente outra variável – a influência da aprendizagem formal da língua segunda – que resolvemos não incluir nesta comunicação, uma vez que não se mostrou tão relevante para a diferenciação de valores aspectuais quanto as outras variáveis aqui apresentadas.

⁶⁸ Sobre o contexto de imersão *vd* Silva (2009: 14- 18).

3.2. A influência da língua materna

Em relação ao segundo fator – a influência da língua materna⁶⁹ –, e sendo a oposição aspectual perfeito/imperfeito interna ao sistema verbal russo, procurámos resposta para as seguintes questões:

- (iii) Conseguiriam os informantes reconhecer mais facilmente a oposição aspectual Pretérito Perfeito Simples e Pretérito Imperfeito na língua não materna, isto é, em Português?
- (iv) Conseguiriam reconhecer formas imperfetivas de Português, quando contrastadas com formas perfetivas?
- (v) Esse reconhecimento revelaria um estabelecimento de nexos aspetuais relativos a pontos de contacto com a sua língua materna?

107

4. Metodologia: descrição do inquérito e dos informantes

A pesquisa levada a cabo teve por base dados extraídos de um inquérito distribuído a imigrantes falantes de Russo e aprendentes de Português língua segunda, do qual faziam parte várias perguntas, sendo as primeiras destinadas a uma caracterização, ainda que sumária, do perfil dos informantes, e as segundas destinadas à testagem das formas de expressão do Aspeto em Português. Para estas, foram utilizados critérios de ordem semântica: reconhecer o sentido de uma frase como idêntico ou dissemelhante ao/do de outra frase⁷⁰. Ou seja, para cada pergunta ou conjunto de perguntas, foi pedido aos informantes que assinalassem as interpretações possíveis de frases distintas, testando formas com sentido idêntico/estrutura diferente e com estrutura (aparentemente) idêntica e sentido diferente. Como dissemos,

⁶⁹ Partindo de uma sugestão de Corder (1993: 9), adotámos na dissertação de Mestrado o termo *influência da língua materna* para, de uma forma abrangente, “ caracterizar supostas aplicações de recursos da L1 ou de outras línguas sobre a L2”; “a conjectura de ocorrência dessa influência (positiva ou negativa) do conhecimento prévio do aprendente (...) partirá da suposição de existência de características estruturais e de relações de sentido que poderão ser a explicação para determinados fenómenos de correção ou de incorreção (...)” (Silva 2009: 26).

⁷⁰ Exemplo: “Eu *trabalhava* na Ucrânia. / Eu *trabalhei* na Ucrânia.”.

era nosso intuito apurar dificuldades (ou facilidades) que este tipo de informantes declarasse, sob um ponto de vista semântico.

Foram inquiridos 54 falantes adultos⁷¹ – um grupo heterogêneo que, na sua maioria, comunica em Português em contexto laboral e/ou relacional, sendo que 26 informantes afirmaram ter aprendizagem formal da língua e 28 não. A média de idades dos que tiveram aprendizagem formal encontra-

108 -se entre os 33 anos, enquanto a média dos que não a tiveram é de 39 anos. No que respeita ao contexto de comunicação e aos interlocutores, um grande número dos inquiridos respondeu que comunica em português no trabalho (44 informantes) e na rua (36 informantes). O contexto predominante no qual comunicam em Português continua a ser, no entanto, o contexto laboral: 42 inquiridos comunicam com colegas de trabalho e 38 responderam que comunicam com amigos ou conhecidos – considerámos este, portanto, um critério para a suposição da influência do contexto de imersão.

Assim sendo, os contextos sociais e linguísticos mais relevantes são o contexto laboral seguido do contexto informal – a rua.

5. Análise dos dados

Para a análise dos dados, procurámos estabelecer correlações entre índices de correção/ reconhecimento de valores e as duas variáveis acima referidas – o contexto social e linguístico que envolve os falantes e a influência da língua materna.

Tendo em conta as correlações acima anunciadas, podemos dizer, em primeiro lugar, que detetámos a preferência pelo valor de habitualidade contido numa forma de presente, e que fazia parte de um exercício de distinção entre “Ele *janta* em casa.” e “Ele *janta* todos os dias em casa.”. Este facto evidencia a *combinação das duas variáveis – a influência da LM e do contexto de imersão*, sendo provavelmente justificável por se tratar de

⁷¹ O trabalho apresentado na dissertação de Mestrado (Silva 2009) pretendia inicialmente fazer uma análise estatística das respostas, correlacionando, como foi dito em 2., as diferentes variáveis. No entanto, dado o número reduzido de respostas, as constatações não terão validade suficiente. Por esse motivo, o texto presente resume apenas o que consideramos serem tendências a explorar mais em detalhe.

um valor aspetual comum a ambas as línguas (a russa e a portuguesa) e de ser ainda bastante usado no contexto que rodeia o inquirido.

Pelas respostas dadas, verificámos que é fator de reconhecimento de valores aspectuais provavelmente com relevância do contexto de imersão a *presença de elementos contextuais lexicais*, isto é, de adverbiais temporais (“Ele janta *todos os dias* em casa.”). Assim, alguns índices de correção no reconhecimento de formas imperfetivas poder-se-ão justificar pela presença de elementos contextuais lexicais, como, por exemplo, um adverbial temporal, e ainda pela exposição ao uso frequente da perifrástica e facilidade em usá-la.⁷² Com efeito, os falantes ouvem mais vezes a perifrástica, o que deve também estimular o seu uso, circunstância essa que corrobora a relevância do contexto de imersão relacionada com um índice maior de correção em informantes sem aprendizagem formal.

109

Uma outra condicionante – o uso de expressões adverbiais de tempo (“*Enquanto* a Maria trabalhava, o Vladimir dormia.” / “*No momento em que* a Maria estava a trabalhar o Vladimir estava a dormir”) – confirma uma preferência por elementos contextuais de tipo lexical, que poderá ser típica de informantes sem aprendizagem formal (Silva 2009: 93), tendo sido observada, desta feita, uma influência do contexto de imersão para este tipo de informantes.

A influência da LM não é fácil de avaliar, desde logo porque alguns falantes poderão não ter sido absolutamente precisos no que toca ao seu conhecimento do russo. Por este motivo, considerámos a possibilidade de estabelecimento de correlações de forma/sentido entre os dois sistemas linguísticos, quando apresentassem estruturas semelhantes – tipicamente, é o caso de *transfer positivo* de um adverbial que nos parece facilitar o reconhecimento da noção de perfeitividade presente numa forma de Mais-que-Perfeito (um tempo verbal que não existe na LM do inquirido).⁷³

⁷² A facilidade em conjugar a perifrástica poderá justificar-se pelo facto de os falantes apenas necessitarem de conhecer a flexão verbal do verbo *estar* e de a conjugar com o infinitivo de qualquer verbo. Esta foi uma informação obtida através do Doutor Vladimir Pliassov, que nos chamou a atenção, ainda, para a ocorrência de um possível erro quando este tipo de falantes incorre naquilo a que se chama de *over-use*: **estou a saber* em vez de *sei* (Silva 2009: 81, nota 109).

⁷³ A possibilidade de ocorrência de *transfer positivo* deste adverbial de contraste “refuta um parecer da investigação levada a cabo por Dietrich *et alii*. (1995) sobre uma improbabilidade

Se traduzirmos para russo a frase que dá o mote a uma das questões, com uma forma de MQP, observamos a presença necessária do elemento adverbial *uzbe* (*já*) para situar temporalmente a situação: “Quando o telefone tocou, eu *tinha feito* o jantar.” / “Kogda zazvonil telefon, ja *uzbe* prigotovil uzhin.” Sendo, portanto, utilizado este adverbial temporal para expressar a perfeitividade de uma situação em relação a outra, não é de estranhar que os inquiridos reconheçam essa função quando configurada na L2. Assim, e atendendo à estrutura do contexto frásico, poder-se-ia deduzir que, além da semelhança da função, também uma semelhança da estrutura poderá estar em causa, uma vez que o adverbial se posiciona antes do verbo, como em português. Por esse motivo, julgamos ser possível o adverbial *já* funcionar como um ponto de referência explícito que auxilie o reconhecimento da anterioridade e da perfeitividade de uma situação em relação a outra (Silva 2009: 94). A par do recurso à LM está o contexto de imersão, que coopera para a associação de sentidos.

Não obstante, outros casos são mais difíceis de interpretar. Por um lado, os aprendentes revelam dificuldade em reconhecer noções aspectuais também presentes na sua L1 – é o caso de valores imperfetivos configurados na forma de Pretérito Imperfeito simples – sustentadas apenas na morfologia da forma verbal em L2 (“Eu *trabalhava* na Ucrânia”), sem a coocorrência de um adverbial temporal no contexto frásico que esclareça o seu sentido. Note-se que nesta questão quase metade dos informantes não distinguiu semanticamente esta forma de outra no Pretérito Perfeito. Por outro lado, independentemente da aprendizagem formal a que tenham sido submetidos, há muitos que revelam facilidade em reconhecer o valor aspectual habitual – o que poderá indiciar uma ativação de recursos semânticos da L1 aplicados à L2 (“Ele *janta* em casa / Ele *janta* todos os dias em casa.”). Além da presença de um adverbial temporal, essa influência da LM está combinada, no nosso entender, com o contexto de imersão.

de ocorrerem fenómenos de transferências de palavras que envolvessem a temporalidade” (Silva 2009: 94).

Por outro lado ainda, o facto de o valor iterativo presente na forma de Pretérito Perfeito Composto, combinado com predicadores basicamente eventivos, ser pouco perceptível e de difícil compreensão para falantes de Português L2, quando surge sem o reforço de outros elementos contextuais, poderá explicar-se, em parte, pelo facto de na LM tal valor não poder ser expresso por um tempo composto, formado com o auxiliar *ter* + PP. Será isso que explica a diferença na facilidade de reconhecimento do valor simplesmente durativo em “Eu *tenho estado desempregado*”, por contraste com a dificuldade evidenciada no reconhecimento do valor iterativo em “Eu *tenho encontrado* a Maria esta semana.”

Terá sido, portanto, a ausência de correlações com a L1 que poderá ter travado o reconhecimento de certos valores aspectuais da L2, apesar do grau de imersão do aprendente na comunidade e do seu tempo de permanência no país, sendo esta observação confirmada pela elevada taxa de erro observada na questão do Pretérito Perfeito Composto. A presença de um adverbial temporal a especificar o valor aspectual singular de uma forma de PPC mostrou-se ineficaz, contrariamente à marcante tendência que confirma a informação fornecida pelos adverbiais temporais como necessária à contextualização e especificação aspectual de uma situação. Referimo-nos concretamente à questão em que os inquiridos não reconheceram o valor de iteratividade na forma “Eu *tenho encontrado* a Maria esta semana”, dado não terem associado o seu sentido a um contexto frásico com um adverbial “Eu encontrei a Maria *muitas vezes* esta semana.”

Em relação à oposição imperfetivo/perfetivo, o valor aspectual imperfetivo expresso apenas morfologicamente pela flexão verbal (sem o acompanhamento de adverbiais temporais) parece não ser suficientemente explícito para falantes adultos, ainda que faça parte de uma das noções estruturantes do sistema verbal da LM. Assim, e para finalizar, parece poder refutar-se a hipótese de uma suposta facilidade por parte deste tipo de informantes em reconhecer este valor aspectual noutra configuração – Pretérito Imperfeito simples – tendo eles apenas aprendido incidentalmente a língua em contexto real de comunicação.

6. Conclusão

112 Tendo em conta as correlações acima estabelecidas que, embora estatisticamente não suficientes, evidenciam possíveis tendências a explorar, parece-nos que o trabalho realizado permite as seguintes conclusões: (i) o contexto de imersão e a influência da LM surgem como possíveis condicionantes de um reconhecimento de oposições aspectuais; (ii) são dois fatores de facilitação que atuam em conjunto; (iii) o contexto de imersão não funciona sozinho (pelo menos no caso do PPC). Por outras palavras, a combinação destas condicionantes parece ser não só mais determinante para diferenciações aspectuais do que a atuação isolada de cada uma delas como também se revelou extremamente necessária, pois a operacionalidade singular de uma está, pelos vistos, interligada com a outra. Vejam-se os casos em que o contexto de imersão *per se* não poderá influenciar particularmente o falante se este não puder estabelecer pontos de contacto com a sua LM, ou com outra língua estrangeira que conhece, como por exemplo no reconhecimento do valor iterativo presente em algumas formas do PPC. Na verdade, a influência da LM (variável mais decisiva do que a influência da aprendizagem formal) parece ter mais peso do que o papel do contexto de imersão, uma vez que foram observados alguns casos possíveis de transferências positivas da LM.

Referências

- Cohen, D. (1989). *L'aspect verbal*. Paris: Presses Universitaires de France
- Comrie, B. (1976). *Aspect. An introduction to the study of verbal aspect and related problems*. 3ª Edição. Cambridge: Cambridge University Press.
- Corder, S. P. (1993). A Role for the mother tongue. In: S. Gass e L. Selinker (eds), *Language transfer in language learning*. Amsterdam: John Benjamins, 18 -30.
- Dietrich, R., W. Klein e C. Noyau (1995). *The acquisition of temporality in a second language*. Amsterdam: John Benjamins.
- Silva, Catarina Pereira (2009). Oposições aspectuais em Português Língua Segunda. O caso dos falantes de russo em contexto de imersão. Tese de Mestrado, Faculdade de Letras da Universidade de Coimbra.

PARA O ENSINO DO “*PRESENT PERFECT*”
A ALUNOS FALANTES NATIVOS DO PORTUGUÊS EUROPEU

1. Introdução

Neste texto é feita uma reflexão sobre o valor semântico do “Present Perfect” (doravante designado PP) com vista à construção de estratégias de didatização adequadas ao ensino daquele tempo verbal a alunos falantes nativos do português europeu⁷⁴. Para tanto, reavaliam-se as propriedades do PP, designadamente na sua dimensão aspeto-temporal. Por outro lado, o perfil dos aprendentes e os fatores externos que configuram a situação de ensino-aprendizagem são igualmente analisados.

O facto de o PP ser um conteúdo gramatical que faz parte dos programas de inglês a partir do 7º ano, nível III (isto é, o terceiro ano de aprendizagem da Língua Estrangeira I) em Portugal, e o facto de ser um tempo verbal com o qual professores e alunos se sentem desconfortáveis e inseguros justificam o interesse pelo tema e a necessidade de o discutir. Além disso, são muito escassos os estudos centrados neste tempo verbal na perspetiva do aprendente falante nativo de português, sobretudo de português europeu.

As dificuldades sentidas na aprendizagem (e no ensino) deste tempo verbal estarão, por um lado, certamente relacionadas com as especificidades deste tempo verbal, com dificuldades decorrentes da situação de contacto de línguas em questão e, ainda, com dificuldades do foro interno da população alvo. Por outro lado, a ideia tradicional, defendida na generalidade das gramáticas

⁷⁴ O presente trabalho corresponde a uma síntese da Dissertação de Doutoramento, intitulada “O Ensino do *Present Perfect* a alunos falantes nativos do português europeu”, apresentada à Universidade de Coimbra, em julho de 2010.

e manuais, de que o PP tem diversos valores semânticos em nada favorece a clareza e em nada contribui para ultrapassar as dificuldades referidas.

2. Pressupostos teóricos

114

2.1 Valor Semântico do Present Perfect

O “Present Perfect” é um tempo verbal problemático, logo à partida devido à dificuldade em inseri-lo numa esfera temporal. Apesar de existirem vários aspetos sobre os quais os diferentes autores que estudam o PP têm uma posição divergente, existe um que colhe unanimidade: tendo como moldura o estudo do sistema verbal inglês realizado por Reichenbach (1947), o PP descreve uma situação que, de alguma forma, é marcada por uma relação entre o passado e o presente, dado que, por um lado, o tempo de referência tem uma relação de sobreposição com o momento de enunciação e, por outro, se convoca uma eventualidade que teve início na esfera do passado. Contudo, segundo este autor (e grande número de outros que a ele se seguiram), a eventualidade descrita pelo PP localiza-se num intervalo de tempo anterior ao momento de enunciação e a posição que neste trabalho se defende é a de que o PP focaliza a consequência no presente dessa eventualidade e não a eventualidade em si.

Por razões que se prendem com a compreensão do PP, mas sobretudo tendo em mente a aplicação pedagógica dessa compreensão, neste trabalho é assumida uma teoria integradora dos diferentes usos do PP e, consequentemente, defende-se a conceção do PP como um tempo verbal que tem *um* determinado valor semântico.

A nossa proposta é fundamentada de forma mais elaborada por Moens (1987) e Kamp & Reyle (1993), mas outros autores também a subscrevem (cf. Smith 1991; de Swart 2000, 2007; e Steedman 2005, entre outros).

A teoria proposta por Moens (1987) e Kamp & Reyle (1993) tem como um dos seus alicerces o núcleo aspetual primeiramente apresentado por Moens (1987), mas também adotado, entre outros, por Kamp & Reyle (1993), autores do esquema que aqui se reproduz:

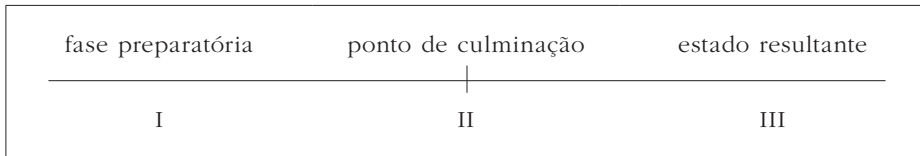


Figura 1 - Esquema tripartido

É com base neste esquema tripartido que autores como Moens (1987), Moens & Steedman (1988), Smith (1991) e Kamp & Reyle (1993) argumentam no sentido de que o PP focaliza um estado resultante (ou consequente, na terminologia de Moens). Aliás, já Comrie (1976) havia defendido que o sentido geral do PP era “the continuing relevance of a previous situation”, tendo ainda afirmado: “The Perfect looks at a situation in terms of its consequences” (Comrie 1976:64). A situação descrita pelo PP é representada pela fase III (embora a existência desta fase pressuponha a existência da fase II); às predicções estativas não se pode aplicar o esquema tripartido, mas Kamp & Reyle associam estas predicções a uma “culminação”⁷⁵, ou melhor a uma fronteira que ora marca o início do estado consequente, ora marca o fim da eventualidade que dá origem ao estado consequente. Deste modo, os princípios básicos da descrição semântica do PP são: (i) o PP focaliza a consequência de uma eventualidade e não a eventualidade em si; (ii) a consequência ou, por outras palavras, o estado resultante, inclui o momento da enunciação e a eventualidade que lhe dá origem ocorreu na esfera do passado; (iii) o estado resultante surge na sequência de uma culminação, razão pela qual os predicadores de “achievements” e de “accomplishments”⁷⁶ possuem as condições necessárias para se combinarem com o PP; relativamente às atividades e aos estados, é necessária a intervenção

⁷⁵ De realçar que, quando se trata de um predicador tético, a culminação é inerente ao predicador; quando estamos perante um estado, não se trata rigorosamente de uma culminação, é antes a terminação (evento pontual) do estado que dá origem a um estado resultante.

⁷⁶ Neste trabalho é usada a terminologia e a classificação aspetual de Vendler (1967).

de diferentes dispositivos linguísticos ou de contexto discursivo que consigam dotar a predicação de uma culminação com estado consequente⁷⁷.

Vejam os de que forma a exposição que acabámos de apresentar se aplica a enunciados aspetualmente distintos. Começemos por analisar o exemplo (1) em que a estrutura predicativa é tética e dá origem a um estado resultante:

116 (1) Max has built a house. (Moens 1987:70)

“Construir uma casa” é um sintagma verbal que denota um “accomplishment”. Preenche, por isso, os requisitos para poder coocorrer com o PP, já que descreve uma situação que contém um ponto de culminação, a partir do qual se configura uma consequência:

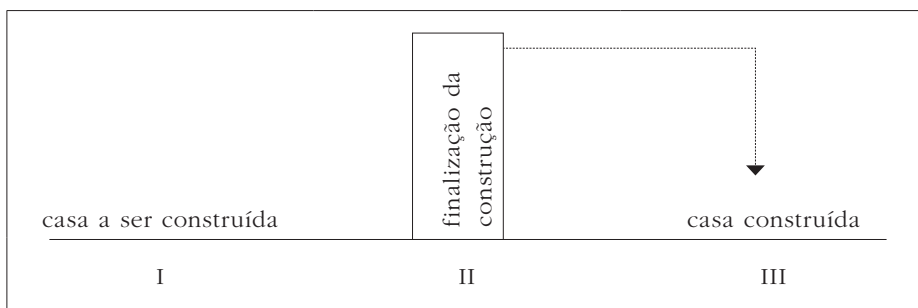


Figura 2 - Aplicação do esquema tripartido ao exemplo (1).

Quanto aos predicadores basicamente estativos, Kamp & Reyle (1993) assentam sobretudo a sua reflexão no facto de estes predicadores representarem situações que não envolvem a fase II do esquema tripartido e, conseqüentemente, também não envolvem a fase III. Por esta razão, Kamp & Reyle consideram as predicções estativas um caso à parte, não lhes aplicando o esquema tripartido. No entanto, estes autores reconhecem que “stative verbs admit the perfect just as readily as non-statives” (Kamp & Reyle 1993:567), mas referem também que “English perfects of stative verbs can be used in

⁷⁷ Trata-se de uma classificação aspetual dinâmica, na medida em que essa classificação é determinada pela interação de diversos fatores, entre eles a estrutura argumental, a coocorrência de adverbiais e aspetos de natureza pragmática.

two quite different ways” (Kamp & Reyle 1993:567). A estas duas maneiras, isto é, a estes dois tipos de leitura que um enunciado onde ocorre o PP pode ter, chamaremos “leitura prototípica” e “leitura idiossincrática” (do inglês)⁷⁸. Para ilustrar as duas leituras, Kamp & Reyle dão o seguinte exemplo:

(2) Mary has lived in Amsterdam for three years. (Kamp & Reyle 1993:567)

117

O exemplo (2) serve para demonstrar que um enunciado em que um predicador estativo está combinado com o PP pode ter (i) uma leitura que permite considerar-se que Mary ainda está, neste momento, a viver em Amesterdão (leitura idiossincrática) e (ii) uma outra leitura segundo a qual a Mary viveu nessa cidade durante um período de três anos, no passado (leitura prototípica). Neste último caso, o PP associa o predicador a uma eventualidade pontual que configura a fronteira final do estado e a fronteira inicial do estado consequente:

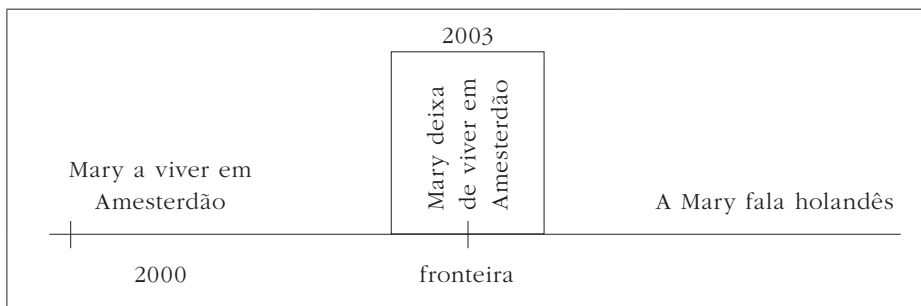


Figura 3 - Aplicação do esquema tripartido a um enunciado com leitura prototípica

⁷⁸ Recorremos à expressão “leitura prototípica” uma vez que o PP se combina prototípica e naturalmente com predicadores tólicos e a leitura decorrente de uma predicação tólica é a leitura em que a culminação marca o fim do evento ao mesmo tempo que marca o início do estado resultante.

Relativamente à expressão “leitura idiossincrática”, baseámo-nos, mais uma vez em Kamp & Reyle (1993:568) quando afirmam que “this use of the perfect proves to be the odd one out, for it appears to be something of an idiosyncrasy of English. Dito de outro modo, “leitura idiossincrática” será aquela que pressupõe que é o início do estado que dá origem ao estado resultante.

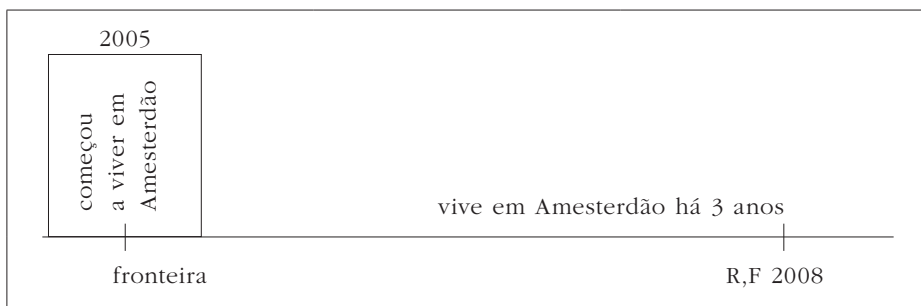


Figura 4 - Aplicação do esquema tripartido a um enunciado com leitura idiossincrática

Neste caso, o estado resultante é consequência do evento pontual que configura o seu próprio início e corresponde ao estado denotado pelo predicador. Podemos, assim, dizer que também nestes casos o estado consequente está associado a um evento pontual — o evento que configura o início do estado.

Kamp & Reyle (1993) descrevem os traços temporais e aspectuais do PP segundo o quadro que a seguir se apresenta⁷⁹:

⁷⁹ Neste quadro, a coluna A corresponde às designações dos tempos verbais; a coluna B diz respeito ao traço temporal que determina a relação entre o Ponto de Perspetiva Temporal e o momento de enunciação; a coluna C diz respeito à relação de ordem entre a eventualidade e o Ponto de Perspetiva Temporal, pelo que se refere ao tempo verbal, que pode ser Presente, Passado ou Futuro; a coluna D indica se se trata ou não de uma predicação estativa; a coluna E refere-se à existência ou não de um estado consequente.

A	B	C	D	E
Present	-PAST	pres	+STAT	-PERF
Simple Past	-PAST	past	+/-STAT	-PERF
	+PAST	pres	+STAT	-PERF
Present Perfect	-PAST	pres	+STAT	+PERF

Figura 5 - Traços tempororo-aspectuais de três tempos verbais ingleses (Kamp & Reyle 1993)

Relativamente ao tempo verbal em apreciação neste trabalho, a tradução deste quadro é a de que o PP (i) descreve um estado resultante (colunas D e E), (ii) representa uma relação de sobreposição com o momento de enunciação (coluna C), e que (iii) o seu Ponto de Perspetiva Temporal⁸⁰ é coincidente com o momento de enunciação (coluna B). Note-se que os valores temporais apresentados para o PP coincidem com os do “Present”, o que contribui para a assunção de que o PP é um tempo verbal da esfera do presente; é o traço aspetual [+PERF] (existência de um estado consequente) que o distingue do “Present”.

2.2 Tradução do “Present Perfect”

Apurado o valor semântico do PP, constatámos que não existe na língua materna da população alvo um tempo verbal com igual valor semântico e verificámos existirem mesmo três tempos verbais portugueses (sobretudo o Pretérito Perfeito Simples (PPS), mas também o Presente do Indicativo (PI) e o Pretérito Perfeito Composto (PPC)) que podem ser utilizados para traduzir enunciados onde ocorre o PP. De realçar ainda que o tempo verbal português morfologicamente correspondente ao PP é justamente aquele que, na tradução de enunciados onde ocorre o PP,

⁸⁰ O Ponto de Perspetiva Temporal corresponde à noção original de tempo de referência de Reichenbach (1947).

deverá ser menos vezes convocado. O quadro seguinte resume o que acaba de ser dito:

120

	INGLÊS	PORTUGUÊS	
Classe aspetual	Enunciado onde ocorre o PP	Tempo verbal	Tradução
Estados	I've been here for 2 hours.	PI	Estou aqui há duas horas.
	Paul has been ill.	PPC	O Paul tem estado doente.
	Clara has lived in Lousã.		O Paul esteve doente. A Clara viveu na Lousã.
Atividades	I've worked for 3 hours.	PPS	Trabalhei durante 3 horas.
"Accomplishments" e "Achievements"	Max has built a house. The President has resigned.		O Max construiu uma casa. O Presidente demitiu-se.

Figura 6 - Tempos verbais portugueses convocados para a tradução do PP

Ao traduzirmos o PP pelo PPS, verificamos a possibilidade de ocorrência de "já" na tradução. Veja-se um exemplo:

- (3) I've found my ring.
- (4) *Já* encontrei o meu anel.

Lopes (2003:417), a propósito da combinação de "já" com descrições de eventos, refere-se à "descrição perfeitiva ou acabada do evento" e "ao estado consequente que se focaliza e se apresenta como relevante".

Verifica-se que em português existe uma estrutura que, por incluir igualmente o particípio passado, "traduz" cabalmente o valor semântico do PP. Trata-se das passivas adjetivais (também denominadas, precisamente, "passivas de estado" ou "passivas resultativas") que "focalizam o estado resultante da transição sofrida" (Duarte 2003: 535). Vejam-se os seguintes enunciados:

- (5) He has won the race.
- (6) A corrida está ganha.

2.3 Contexto em que ocorre o processo de ensino-aprendizagem do “Present Perfect”

As condições de aprendizagem das línguas estrangeiras em Portugal não favorecem a sua assimilação. Na verdade, a exposição ao “input” é predominantemente feita em contexto instrucional, pelo que se trata de um “input” escasso e insuficiente para uma aprendizagem de Língua Estrangeira (LE) eficiente, já que, e de acordo com grande número de autores, o sucesso da aprendizagem de uma LE está em relação direta com a quantidade de “input” a que o sujeito está exposto (cf. Lightbown 2003, White 2003, R. Ellis 2005, Gass & Selinker 2008, entre outros). Por outro lado, trata-se fundamentalmente de um “input” não-nativo (professor) e/ou mesmo agramatical (pares), existindo uma reduzidíssima ou nula interação dos aprendentes com falantes nativos, pois são muito raras, ou mesmo inexistentes, as oportunidades de produção em contextos comunicativamente relevantes e naturais.

121

A população sob escrutínio teve a sua primeira exposição estruturada à LE por volta dos dez anos, circunstância esta que é bastante pertinente, uma vez que convoca outro tipo de reflexão, concretamente a que diz respeito a condições de desenvolvimento cognitivo e psicolinguístico dos sujeitos aquando da sua exposição ao “input” de língua inglesa, no contexto e nas condições descritas.

Autores como Ullman (2001, 2004) e Paradis (1994), entre outros, atestam uma diferença na qualidade de assimilação de uma Língua Materna (LM) e de uma Língua Não-Materna (LNM). A aquisição linguística feita pelas crianças - “aquisição” (cf. Krashen 1981) – traduz-se num conhecimento implícito, adquirido “incidentally”, armazenado implicitamente e usado automaticamente. Um tipo de conhecimento distinto é o que predomina na assimilação de uma LNM e que é aprendido conscientemente, armazenado explicitamente, podendo ser usado de uma forma controlada (cf. Paradis 2004:61) – “aprendizagem” (cf. Krashen 1981). É um conhecimento do tipo “know that”, um “saber sobre” (Martins 2008). Quando o conhecimento é qualitativamente aprendido desta forma, o indivíduo está concentrado nos itens que está a aprender e evidencia, por essa razão, a capacidade de, ulteriormente, descrever esse conhecimento.

O conhecimento adquirido e o conhecimento aprendido distinguem-se ainda pelos sistemas de memória que os suportam: o conhecimento implícito é armazenado na memória de tipo procedimental, ao contrário do conhecimento explícito que é armazenado na memória, também de longo prazo, mas de natureza declarativa (cf. Ullman 2001). Assim sendo, a capacidade da população alvo de assimilar conhecimentos gramaticais por via procedimental será já menor do que em idades mais precoces, passando a memória declarativa a exercer um papel bastante mais preponderante neste processo assimilatório.

Para além do fator idade, hoje em dia, é comumente aceite que existem igualmente outros fatores que podem afetar o processo de aprendizagem de uma LE — a LM, particularidades da Língua-Alvo (LA), atitude, motivação, aptidão para as línguas, ou outras línguas já assimiladas pelos aprendentes.

3. Trabalho Experimental

O trabalho experimental foi realizado com uma amostra de 209 informantes falantes nativos de português europeu, de um universo de 2.151 alunos a frequentar, no ano letivo 2009/2010, o ensino regular em escolas oficiais no Concelho de Coimbra. Cerca de metade dos inquiridos estava a frequentar o 9º ano de escolaridade (ano de conclusão do Ensino Básico) e a outra metade o 11º (que, para a maioria dos alunos, representa o último ano do Ensino Secundário em que o Inglês faz parte dos seus currículos). Desde modo, todos tinham pelo menos quatro anos de exposição à língua-alvo (o inglês) em ambiente instrucional. Por outro lado, os informantes tinham, aquando da aplicação dos inquéritos, idades superiores a 13 anos, o que significa que tinham uma idade que não lhes permitia assimilar uma LE com a mesma qualidade com que assimilaram a sua LM, isto é, de forma automática e sem esforço, pela mera exposição ao “input” linguístico.

O trabalho fundamentou-se em cinco questionários, com dez perguntas cada⁸¹. No sentido de conseguirmos resultados mais genuínos, para cada

⁸¹ Apresentam-se em anexo os referidos inquéritos.

questionário foram feitas duas versões (A e B) que se distinguem entre si apenas pela ordem das opções. Convém ainda esclarecer que os resultados obtidos foram considerados corretos ou incorretos, seguindo critérios decorrentes de todos os princípios por nós defendidos no que diz respeito ao valor semântico do PP, a saber: o valor semântico do PP é o de focalizar um estado resultante de uma eventualidade ocorrida na esfera do passado, estado esse que tem relevância no presente (ou no momento da enunciação).

123

As tarefas que os informantes tinham de desempenhar em cada um dos questionários (I a V) eram as seguintes:

- I - escolher, de entre as três opções que lhes foram dadas, qual a frase em português que melhor se associava à situação descrita pela frase apresentada em inglês. Constavam deste questionário dez enunciados em inglês em que o PP coocorria com predicadores que, ao nível do seu valor aspectual básico, pertenciam a diferentes classes aspectuais;
- II - escolher, de entre três opções dadas, a tradução correta do enunciado apresentado em inglês. Em todos estes enunciados ocorria o PP, mas os predicadores pertenciam a classes aspectuais básicas diferentes;
- III - escolher uma de três opções no sentido de completar um enunciado inglês. Existia sempre (i) uma opção correta com o verbo no PP; (ii) uma opção agramatical em que a anomalia residia no auxiliar “have”; (iii) uma outra opção agramatical em que a anomalia residia no particípio passado;
- IV - escolher a melhor opção para retroverter a frase sublinhada de um enunciado em português. Estes enunciados originais variavam entre si pelo facto de a situação ser descrita por tempos verbais diferentes: PPS, PPC e PI. Por outro lado, para cada enunciado os informantes podiam sempre optar entre um enunciado onde ocorresse o PP, o “Past Simple” ou o “Present Simple”;
- V - completar semanticamente enunciados em português com uma das três opções dadas em inglês. Estas opções compreendiam sempre um enunciado onde ocorria o PP, outro onde ocorria o “Past Simple” e um outro onde ocorria um tempo verbal de uma esfera temporal diferente.

O trabalho experimental veio corroborar as hipóteses de trabalho apresentadas, designadamente de que (i) os informantes interiorizam o PP com o mesmo valor semântico do “Past Simple”; (ii) o conhecimento morfológico do PP é necessário, mas não suficiente para a assimilação da sua semântica; e (iii) a língua materna é uma fonte de transferência linguística, pelo que a não equivalência dos dois sistemas verbais em causa se manifesta de diversos modos, nomeadamente em tarefas centradas na tradução e na retroversão⁸².

4. Implicações didáticas

Na sequência dos fundamentos teóricos e do trabalho experimental apresentados, as propostas de atividades de didatização do PP tiveram em consideração os seguintes princípios: (i) o PP focaliza um estado resultante de uma eventualidade; (ii) não há correspondência direta entre o PP e um tempo verbal português; (iii) a motivação é essencial para o ensino eficaz de uma língua estrangeira; (iv) num contexto instrucional, a língua estrangeira deve ser explicitamente ensinada, não se negando, todavia, a importância de um ensino que vise promover a assimilação implícita de estruturas linguísticas de uma LNM (cf. N. Ellis 2005, 2007, 2008; Ellis & Bogart 2007, Gass & Selinker 2008, entre outros).

Para fazer face às dificuldades diagnosticadas e sustentadas pelo quadro teórico apresentado, as atividades didáticas propostas deverão ser exemplificativas de um ensino que consideramos motivador. Na verdade, Paradis (2004), R. Ellis (2005), mas também a nossa própria experiência como docente aconselham-nos a não negligenciar o fator “motivação” na construção e na utilização de recursos pedagógicos que pretendemos eficazes.

Assim sendo, com o objetivo primordial de dar a conhecer o valor semântico do PP e de esclarecer o conceito de “estado resultante”, propomos que o material pedagógico seja ilustrativo da relação de contingência que

⁸² Dados os limites de espaço definidos, remetemos o leitor interessado para a dissertação de doutoramento onde estão devidamente tratadas as respostas obtidas nos inquéritos.

existe entre o estado consequente e a eventualidade que lhe dá origem. Uma forma porventura eficaz de transmitir esta noção de estado resultante será associar um enunciado onde ocorre o PP a um outro enunciado onde ocorre o “Present Simple” e em que a localização da situação descrita na esfera do presente é mais clara. Será o caso de materiais em que uma frase em que ocorre o “Present Simple” e/ou uma imagem remetem para um estado resultante no presente de uma eventualidade ocorrida na esfera do passado. Veja-se um exemplo:

125

He (rob) a bank. He's behind bars.



Figura 7 – Apresentação/Aplicação do conceito de “estado resultante”.

No que diz respeito à oposição PP/“Past Simple”, e suportados pela descrição dos traços aspeto-temporais apresentada por Kamp & Reyle (1993), será positiva a construção de uma ficha informativa, em que fique claro o que distingue um e outro tempos verbais: (i) o PP é um tempo verbal da esfera do presente, ao passo que o “Past Simple” é um tempo verbal da esfera do passado; (ii) o PP refere-se a um estado consequente que está relacionado de forma contingente com uma culminação, ao contrário do “Past Simple”, que localiza a situação descrita num intervalo de tempo anterior ao momento de enunciação. Veja-se um par de exemplos ilustrativo dessa oposição:



He has broken his leg. It is in plaster now.



He broke his leg five years ago.

Figura 8 – Oposição PP/“Past Simple”

Para superar a dificuldade decorrente de não existir um tempo verbal português com o mesmo valor semântico do PP, consideramos essencial que seja explicitamente mostrado aos aprendentes que existem três tempos verbais que, conforme as predicacões, traduzem o PP. Mais uma vez, poderá ser construída uma ficha informativa na qual são dados exemplos com diferentes classes aspectuais de predicadores e com diferentes traduções. Vejam-se alguns desses enunciados com as respetivas traduções:

- (8) They have built a house.
 (9) Eles construíram uma casa. (PPS)
- (10) She has lived in New York since 2003.
 (11) Ela vive em Nova Iorque desde 2003. (PI)
- (12) He has been ill.⁸³
 (13) Ele tem estado doente. (PPC)
 (14) Ele esteve doente. (Pretérito Perfeito Simples)

Referências

- Comrie, B. (1976). *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: Cambridge University Press.
- Cunha, L.F. (2004). Semântica das predicacões estativas. Para uma caracterização aspectual dos estados. Tese de Doutoramento. Faculdade de Letras da Universidade do Porto, Porto.
- Duarte, I. (2003). A família das construções inacusativas. In: M.H.M., Mateus, A. M. Brito, I. Duarte & I. H. Faria (orgs). *Gramática da língua portuguesa*. Lisboa: Caminho, 533-535.
- Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in second language acquisition*, 27:2, 305-352.
- Ellis, N. (2007). The weak-interface, consciousness, and form-focused instruction: Mind the doors. In: S. Fotos e H. Nassaji (Eds.). *Form-focused Instruction and Teacher Education: Studies in Honour of Rod Ellis*. Oxford: Oxford University Press, 17-34.

⁸³ A possibilidade de podermos recorrer a dois tempos verbais portugueses para traduzir este enunciado advém do facto de “be ill” ser um predicador estativo pouco faseável. O conceito de “faseabilidade” deve-se a Cunha (2004) e está associado ao conceito de Rede aspectual de Moens (1987). Cunha considera que os estados com o traço [+faseável] são aqueles que “podem ser convertidos em processos (termo de Moens para “atividades”), ‘absorvendo’, assim, o traço [+dinâmico]” (Cunha 2004:86).

- Ellis, N. (2008). Implicit and explicit knowledge about language. In: J. Cenoz e N.H. Hornberger (eds.). *Encyclopedia of Language and Education* (Vol. 6). Heidelberg: Springer Scientific, 1-13.
- Ellis, N. & P. Bogart (2007). Speech and language technology in education: The perspective from SLA research and practice. *SLaTE*, 1-8.
- Ellis, R. (2005). Principles of instructed language learning. *System* 33, 209-224.
- Gass, S.M. & L. Selinker (2008). *Second Language Acquisition: An Introductory Course* (3^a ed.). New York: Routledge.
- Kamp, H. & U. Reyle (1993). *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer.
- Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.
- Lightbown, P. (2003). SLA research in the classroom/SLA research for the classroom. *Language Learning Journal* 28, 4-13.
- Lopes, A.C.M. (2003). Elementos para uma análise semântica das construções com *já*. In: Ivo Castro e I. Duarte (orgs.). *Razões e emoção* (Vol I). Lisboa: Imprensa Nacional-Casa da Moeda, 411-428.
- Martins, C. (2008). *Línguas em contacto: "saber sobre" o que as distingue*. Coimbra: Imprensa da Universidade de Coimbra.
- Moens, M. (1987). *Tense, aspect and temporal reference*. PhD dissertation, University of Edinburgh.
- Moens, M. & M. Steedman (1988). Temporal ontology and temporal reference. *Computational Linguistics* 14: 2, 15-28.
- Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: implications for bilingualism and SLA. In: N. Ellis (ed.) *Implicit and Explicit Learning of Languages*. London: Academic Press, 393-419.
- Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins, 7-61.
- Reichenbach, H. (1947). The tenses of verbs. In: H. Reichenbach (ed.). *Elements of Symbolic Logic*. New York: The MacMillan Company, 287-298.
- Smith, C.S. (1991). *The Parameter of Aspect*. Dordrecht: Kluwer
- Steedman, M. (2005). The productions of time: Temporality and causality in linguistic semantics. Ms, University of Edinburgh.
- Swart, H. de (2000) Tense, aspect and coercion in a cross-linguistic perspective. In: M. Butt and T. H. King (eds.) *Proceedings of the Berkeley Formal Grammar Conference*. Stanford, CA: CSLI.
- Swart, H. de (2007). A cross-linguistic discourse analysis of the Perfect. *Journal of Pragmatics* 39: 12, 2273-2307.
- Ullman, M.T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30: 1, 37-69.
- Ullman, M.T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231-270.
- Vendler, Z. (1967) Verbs and times. In: *Linguistics in Philosophy*. Ithaca, New York: Cornell University Press, 97-121.
- White, Lydia (2003). On the nature of interlanguage representation: Universal Grammar in the L2. In: Doughty, C. & M. H. Long (eds.). *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell Publishing, 53-83.

ANEXO

Os cinco questionários que se seguem destinam-se a um trabalho de investigação sobre o sistema verbal inglês. Cada questionário tem 10 perguntas de escolha múltipla. Responda, por favor, a todas as perguntas. Obrigada!

I - Indique qual das frases em português (a, b ou c) melhor se associa à situação apresentada na frase em inglês

128

1. **She has bought a red car.**
 - a. Ela já teve um carro vermelho.
 - b. Ela agora tem um carro vermelho.
 - c. Ela vai ter um carro vermelho.
2. **The house has been empty for several months.**
 - a. A casa está cheia de ervas daninhas à volta porque não vive lá ninguém há que tempos.
 - b. Com o preço que estão a pedir, depois de saírem a casa vai ficar abandonada muito tempo.
 - c. A casa esteve sem ninguém muito tempo — mais ou menos de Setembro a Fevereiro do ano passado.
3. **They have built a summer house in Mira.**
 - a. Eles passavam as férias em Mira, quando tinham lá uma casa.
 - b. Eles vão para Mira passar férias, assim que tenham lá uma casa.
 - c. Eles costumam passar as férias numa casa que têm em Mira.
4. **Mary has lived in Amsterdam.**
 - a. A Mary viveu na Holanda quando era ainda bebé.
 - b. A Mary fala bem holandês porque viveu na Holanda.
 - c. A Mary vai viver para a Holanda depois de casar.
5. **I've found my ring.**
 - a. Encontrei o anel, mas já o perdi outra vez.
 - b. Hei de encontrar o meu anel.
 - c. Encontrei o anel. Olha, não é bonito?
6. **I've read that book.**
 - a. Acho que li esse livro quando andava na Primária, mas já nem me lembro da história.
 - b. Comprei este livro para ler nas próximas férias.
 - c. Comprei-te este livro porque gostei muito de o ler.
7. **I've been here for two hours.**
 - a. Estive à espera dele desde as 5. Às 7 fui-me embora.
 - b. Estou aqui desde as 9. Já são 11. Vou-me embora.
 - c. Vou para lá às 2. Às 4 venho-me embora.
8. **The apples have all been eaten.**
 - a. — Dás-me uma maçã?
 — Já não há.
 - b. — Dás-me uma maçã?
 — Toma, comprei-as hoje de manhã porque já não havia.
 - c. — Dás-me uma maçã?
 — Toma duas para ver se acabam.
9. **Fernando has lived in Aveiro since 2005.**
 - a. O Fernando viveu em Aveiro, mas agora vive em Coimbra.
 - b. O Fernando tenciona ir viver para Aveiro.
 - c. O Fernando viveu em Coimbra, mas agora vive em Aveiro.
10. **Jane has broken her leg..**
 - a. A Jane já nem se lembrava que tinha andado de muletas quando era miúda.
 - b. A Jane caiu há dois dias e agora anda de muletas.
 - c. Se a Jane não tiver cuidado ainda cai e depois tem de andar de muletas.

II – Escolha a tradução correta das frases em inglês.

1. **George has lived in Lisbon for 5 years.**
 - a. O George viveu em Lisboa durante 5 anos.
 - b. O George tem vivido em Lisboa desde há 5 anos.
 - c. O George vive em Lisboa há 5 anos.
2. **Bill has won the race.**
 - a. O Bill tem ganho a corrida.
 - b. O Bill ganhou a corrida.
 - c. O Bill ganha a corrida.
3. **Peter has painted a marvellous picture.**
 - a. O Peter pintou um quadro maravilhoso.
 - b. O Peter pinta um quadro maravilhoso.
 - c. O Peter tem pintado um quadro maravilhoso.
4. **I've loved the film.**
 - a. Tenho adorado o filme.
 - b. Adoro o filme.
 - c. Adorei o filme.
5. **I've been here before.**
 - a. Já estive aqui.
 - b. Tenho estado aqui.
 - c. Estou aqui.
6. **My son has lost the key.**
 - a. O meu filho tem perdido achave.
 - b. O meu filho perdeu a chave.
 - c. O meu filho perde a chave.
7. **I've known him since 1984.**
 - a. Conheço-o desde 1984.
 - b. Tenho-o conhecido desde 1984.
 - c. Conheci-o em 1984.
8. **Anne has written a good composition.**
 - a. A Anne escreveu uma boa composição.
 - b. A Anne escreve uma boa composição.
 - c. A Anne tem escrito uma boa composição.
9. **She has known his password for years.**
 - a. Ela soube a senha dele há anos.
 - b. Ela sabe a senha dele há anos.
 - c. Ela tem sabido a senha dele há anos.
10. **John has lived in Paris.**
 - a. O John vive em Paris.
 - b. O John tem vivido em Paris.
 - c. O John viveu em Paris.

III - Preencha o espaço com a), b) ou c).

1. Peter ... a letter to his father.
 - a. has wrote
 - b. has written
 - c. have written
2. Jane ... a 5€ note.
 - a. have found
 - b. has finded
 - c. has found
3. Susan ... them for the party.
 - a. has invited
 - b. have invited
 - c. has invite
4. The Stuarts ... to the USA.
 - a. have been
 - b. have went
 - c. has been
5. My dog ... a vase.
 - a. has broken
 - b. have broken
 - c. has broken
6. Mr and Mrs Jones ... a new house.
 - a. have buyed
 - b. have bought
 - c. has bought
7. Many students ... the homework.
 - a. has done
 - b. have do
 - c. have done
8. Max ... in London for 10 years.
 - a. have lived
 - b. has live
 - c. has lived
9. Janet and Carol ... a cake.
 - a. has made
 - b. have made
 - c. have make
10. Fred ... in the story.
 - a. have believed
 - b. has believe
 - c. has believed

129

IV – Escolha a tradução correta das frases que estão sublinhadas.

1. O António saiu às 11.
 - a. António left at 11.
 - b. António has left at 11.
 - c. António leaves at 11.
2. O Luís vive no Porto há 10 anos.
 - a. Luís lives in Porto for 10 years.
 - b. Luís lived in Porto for 10 years.
 - c. Luís has lived in Porto for 10 years.
3. O Manuel viveu em Paris. Conhece quase todos os museus parisienses.
 - a. Manuel has lived in Paris.
 - b. Manuel lives in Paris.
 - c. Manuel lived in Paris.
4. A Mariana tem estado doente.
 - a. Mariana was ill.
 - b. Mariana has been ill.
 - c. Mariana is ill.
5. A Maria ganhou a corrida. Está no pódio a receber a medalha.
 - a. Maria has won the race.
 - b. Maria wins the race.
 - c. Maria won the race.
6. Encontrei os bilhetes ontem.
 - a. I've found the tickets yesterday.
 - b. I find the tickets yesterday.
 - c. I found the tickets yesterday.
7. Encontrei o livro. Podes levá-lo.
 - a. I find the book.
 - b. I've found the book.
 - c. I found the book.
8. O Miguel tem sido simpático.
 - a. Miguel has been nice.
 - b. Miguel was nice.
 - c. Miguel is nice.
9. Estou aqui desde as 5.
 - a. I'm here since 5.
 - b. I was here since 5.
 - c. I've been here since 5.
10. O Rui esteve doente há duas semanas.
 - a. Rui has been ill two weeks ago.
 - b. Rui is ill two weeks ago.
 - c. Rui was ill two weeks ago.

1. A Patrícia vai a Braga, mas precisava de saber o caminho da estação até à Universidade. Então, o Miguel sugere-lhe que fale com a Teresa, porque
 - a. Teresa has lived in Braga.
 - b. Teresa would live in Braga
 - c. Teresa lived in Braga.
2. A Joana quer levar o filho de 13 anos ao cinema, mas não tem a certeza se o filme é apropriado para a idade dele. Pergunta a opinião ao irmão mais velho. Ele acha que ela pode levar o filho e explica:
 - a. I'll see that film when I'm 12.
 - b. I've seen that film when I was 12.
 - c. I saw that film when I was 12.
3. A Ana pergunta ao João o que é que a Maria tem de diferente e o João responde:
 - a. She has had her hair cut.
 - b. She'll have her hair cut.
 - c. She had her hair cut.
4. O Luís costumava dar boleia ao Fernando porque este não tinha carro. Porém, um dia o Fernando diz-lhe que já não precisa de boleia e explica:
 - a. I would buy a car.
 - b. I've bought a car.
 - c. I bought a car.
5. A Maria trabalha numa empresa e o Paulo começou hoje a trabalhar lá. Apresentam-se e o Paulo pergunta-lhe se aquele foi o seu primeiro emprego. Ela diz que não e acrescenta:
 - a. I've worked in a shop last Christmas.
 - b. I worked in a shop last Christmas.
 - c. I was going to work in a shop last Christmas.
6. O Rui pergunta por que razão a Clara está a faltar. Respondem-lhe que ela está de licença porque:
 - a. Clara had a baby.
 - b. Clara would have a baby.
 - c. Clara has had a baby.
7. A Sandra e a Diana estão a ver fotografias da juventude da Sandra. Quando aparece uma com a Torre Eiffel, ela diz:
 - a. This is in 2002 when I visited Paris.
 - b. This is in 2002 when I've visited Paris.
 - c. This is in 2002 when I'll visit Paris.
8. A Catarina e a mãe passam por um jovem que as cumprimenta. A mãe da Catarina não o reconhece e pergunta quem é. A Catarina explica-lhe que é o Pedro que foi colega dela, há 15 anos, quando ambos tinham 5 anos e acrescenta: não o reconhece porque
 - a. He'll grow up.
 - b. He has grown up.
 - c. He grew up.
9. O Pedro conhece a Ana numa festa e esta diz-lhe que está a viver em Coimbra há poucos anos e explica
 - a. I lived in Faro until 2002.
 - b. I've lived in Faro until 2002.
 - c. I'm going to live in Faro until 2002.
10. A Paula diz à Carla: "Não te imaginava a tomar conta de crianças. Estás mais madura". A Carla responde:
 - a. Yes, I'll change.
 - b. Yes, I changed
 - c. Yes, I've changed

SECÇÃO 5:
PRAGMÁTICA

(Página deixada propositadamente em branco)

A ARGUMENTAÇÃO ORAL FORMAL EM CONTEXTO DE SALA DE AULA

1. Introdução

Algumas das investigações desenvolvidas no âmbito do ensino da expressão oral em meio escolar têm revelado que a abordagem deste domínio é pautada por constrangimentos de ordem diversa, que passam tanto pela falta de preparação dos professores, como pela dificuldade em desenvolver um tratamento sistemático do oral formal, como ainda pelas dificuldades que são inerentes ao próprio ensino da expressão oral, pelo facto de se tratar de uma produção sonora que exige ouvintes e que se caracteriza pela efemeridade⁸⁴.

No presente artigo, apresentamos alguns resultados de uma investigação em torno da argumentação oral formal em contexto escolar, que se centrou em três eixos fundamentais: uma reflexão científica em torno dos domínios estruturantes da argumentação; uma reflexão de cariz pedagógico-didático centrada na definição de um modelo de práticas a desenvolver; uma reflexão a partir de produções orais de alunos do ensino secundário. Com base nestas linhas orientadoras, o presente artigo visa os seguintes objetivos: elencar algumas possibilidades de organização estrutural dos blocos textuais que constituem um texto argumentativo; apresentar as relações de coerência centrais no plano da argumentação positiva e da argumentação negativa; propor um modelo escolar para o tratamento da argumentação oral formal, designado *jogo escolar*; apresentar algumas

⁸⁴ O presente trabalho centra-se no tratamento de alguns dos aspetos abordados na tese de doutoramento intitulada *Argumentação oral formal em contexto escolar*, apresentada à Faculdade de Letras da Universidade de Coimbra.

conclusões relativas ao estudo de textos orais produzidos sem preparação prévia em ambiente escolar. Pretende-se, assim, que, partindo da análise de algumas características de textos produzidos por alunos, se possa definir um modelo de abordagem suscetível de conduzir a uma didática mais eficaz no plano do texto argumentativo oral.

134

2. Relações estruturais no texto argumentativo

A argumentação é uma atividade linguística, fundada num processo intelectual, que surge a partir de uma situação de divergência de opiniões e que tem como objetivo nuclear defender, de forma persuasiva, um determinado ponto de vista, por meio de um texto argumentativo constituído, na sua forma elementar, por argumentos que sustentam uma determinada conclusão. Uma vez que exige a presença, explícita ou implícita, de dois intervenientes, um locutor e um alocutário, a argumentação é, por natureza, uma atividade dialógica, sendo os textos argumentativos inerentemente polifónicos.

Um texto argumentativo é composto por blocos textuais. Estes são segmentos de texto, que se organizam e estruturam em torno da função que assumem no interior do texto. Assim, um texto argumentativo, na sua forma mínima, é composto por dois blocos textuais considerados nucleares:

Argumento(s) – Conclusão⁸⁵

Esquema 1

⁸⁵ Os blocos textuais que consideramos nucleares – argumento(s) e conclusão - correspondem à noção de macroproposição defendida por Adam (2001). Referimo-nos em particular à P. arg.1 (données) e à P.arg.3 (conclusion). Neste texto, não utilizaremos, porém, o termo macroproposição, dado que, aquando da abordagem da estruturação textual, identificaremos outros blocos textuais que, embora não sejam nucleares nem definitórios do texto argumentativo, são utilizados pelos alunos na organização da estrutura argumental.

De acordo com esta linha de análise macro-textual, num texto argumentativo bem formado, só existe um bloco com função de *conclusão*. Consequentemente, todos os blocos textuais que se relacionem com o bloco *conclusão*, no sentido de o justificar ou de o refutar, terão a função de *argumento*. É evidente que uma análise centrada num outro plano poderá levar à defesa de que num mesmo texto argumentativo se identificam várias *conclusões*. Observemos o seguinte texto argumentativo, onde está presente um bloco textual *argumento complexo*, que desenvolve uma ideia nuclear com função de sustentar uma *conclusão*, que corresponde à posição defendida:

(1) [A sopa é ótima para as dietas,]_{conclusão} [[pois reduz o apetite,]
[o que ajuda a perder peso.]]_{argumento complexo}

Se analisarmos internamente o *argumento complexo*, verificamos que as sequências que o constituem não partilham a mesma função. Com efeito, [porque reduz o apetite,] afirma-se como um *argumento principal*, ou seja, aquele que encerra a ideia nuclear do *argumento complexo*, enquanto o segmento [o que ajuda a perder peso] funciona como um *subargumento*, na medida em que se relaciona diretamente com o *argumento principal*, sustentando-o. Por esta razão, é possível proceder a uma análise mais detalhada do *argumento complexo*:

(2) [A sopa é ótima para as dietas,]_{conclusão} [[pois reduz o apetite,]_{argumento principal} [o que ajuda a perder peso.]_{subargumento}]_{argumento complexo}

Não obstante, é também a partir da análise apresentada em (2) que poderemos confirmar que o texto argumentativo é construído por estruturas recursivas. Com efeito, se autonomizarmos o bloco textual *argumento complexo*, observamos que o *argumento principal* assume o estatuto de *conclusão*, passando o *subargumento* a funcionar como *argumento (principal)*:

(3) [A redução do apetite numa dieta é importante,]_{conclusão} [pois ajuda a perder peso.]_{argumento}

A recursividade que se regista nesta construção pode, como é evidente, ser reproduzida em n sequências encaixadas.

136

Os blocos considerados nucleares poderão, ainda, ser submetidos a uma análise de outro cariz que permitirá considerar diferentes tipos de *conclusão* e de *argumentos*, nomeadamente a *conclusão própria* e a *conclusão alheia*, os *argumentos próprios*, os *argumentos alheios* e os *contra-argumentos*. Este enfoque torna-se particularmente relevante nos textos argumentativos monogerados, isto é, produzidos por um único locutor, onde se assiste a um confronto de posições: um dado locutor verbaliza num mesmo texto, por um lado, a posição defendida por um oponente bem como o(s) *argumento(s)* que aquele invoca para a sustentar e, por outro, a sua *conclusão própria* e os seus *argumentos próprios*. Sabendo ainda que uma das possibilidades de movimento argumentativo consiste na apresentação de *contra-argumentos*, cuja construção se orienta no sentido da refutação de *argumentos alheios*, consideramos que é também pertinente distinguir o bloco textual *contra-argumento* do bloco *argumento próprio* pela função específica que aquele desempenha num texto argumentativo. De facto, um *argumento próprio* apresenta argumentos avançados pelo proponente que não se encontram explicitamente condicionados por *argumentos alheios* ou que não constituem formas de os contrapor, ao contrário do que se verifica com os *contra-argumentos*⁸⁶.

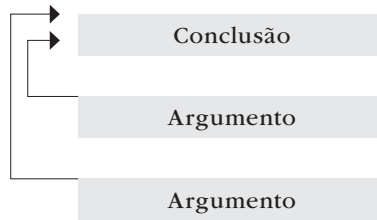
Deste modo, num texto argumentativo, poderemos distinguir diferentes blocos textuais de *argumentos* e diferentes blocos textuais com função de *conclusão*, tal como se ilustra nos esquemas que se seguem:



Esquema 2

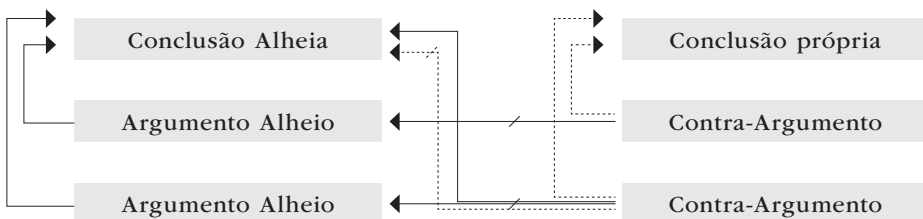
⁸⁶ É importante, todavia, esclarecer que os *contra-argumentos* são *argumentos próprios*, no sentido em que pertencem a um mesmo locutor. Por outro lado, os objetivos que estruturam ambos os blocos textuais passam por sustentar a *conclusão própria*.

Os blocos *conclusão* e *argumento* estabelecem entre si uma relação de sustentação, na medida em que a função dos blocos *argumentos* passa por sustentar diretamente o segmento que corresponde à *conclusão*. Estabelece-se, deste modo, uma relação de suporte (Apothéloz e Miéville in Rubattel 1989) entre *argumento* e *conclusão*, de acordo com o modelo que a seguir se apresenta:



Esquema 3

Nos textos que incluem *argumento(s) alheio(s)*, *conclusão alheia* e *contra-argumento(s)*, a relação entre blocos poderá ser relativamente mais complexa, visto que se mantém a relação de sustentação entre *argumento(s)* e *conclusão*, mas entre *contra-argumento(s)* e *argumento(s) alheio(s)* existirá uma relação distinta, dado que o bloco *contra-argumento(s)* se estrutura em relação ao bloco *argumento(s) alheio(s)* e/ou ao bloco *conclusão alheia*, estabelecendo uma relação de refutação, de restrição ou outra, como se esquematiza de seguida:



Esquema 4

3. Relações de coerência no texto argumentativo

138 Num texto argumentativo, os blocos textuais mantêm relações de coerência, típicas da organização argumentativa, que correspondem a uma orientação positiva ou a uma orientação negativa. No primeiro caso, estamos perante uma relação criada entre uma *conclusão própria* e um *argumento próprio* que a sustenta e, no segundo caso, esta processa-se entre uma *conclusão albeia* (ou *argumento albeio*) e um *contra-argumento* que a refuta.

A argumentação positiva está estreitamente relacionada com a ordem, progressiva ou regressiva, de acordo com a qual o texto argumentativo se organiza. A ordem regressiva corresponde a uma estrutura argumental na qual a *conclusão* surge no início do texto, sendo seguida do(s) *argumento(s)*, de acordo com o esquema seguinte:

Conclusão – pois/porque – Argumento

Esquema 5

Se atentarmos no tipo de relações de coerência que se podem instituir entre estes dois blocos textuais, concluiremos, na senda de Lopes (2009), que nos textos têm lugar dois tipos de relação: a justificação e a causa. Nesta última, os conteúdos proposicionais de dois segmentos textuais estabelecem entre si uma relação de causa-efeito entre duas situações temporalmente ordenadas. Por seu turno, a justificação processa-se num plano das relações pragmáticas de coerência entre dois segmentos textuais, no qual «the speaker presents one of the segments as an argument to support his (main) claim» (Lopes 2009).

A relação de coerência existente entre a *conclusão* e o *argumento* não se orienta no sentido de apresentar os eventos ou as situações do mundo real, ordenados temporalmente e ligados por um nexos de causalidade. Assistimos, antes, a uma relação de natureza justificativa, onde o bloco textual *argumento próprio* é avançado como uma justificação para a asserção expressa no bloco *conclusão própria*, segundo o esquema:

Eu afirmo X, e afirmo X, pois/porque Y

Esquema 6

É esta relação que nos permite concluir que (4) constitui um texto argumentativo, enquanto (5) não:

139

(4) A Joana está triste, pois está a chorar.

(5) A Joana caiu porque foi empurrada.

Em (5), estamos perante um único ato ilocutório assertivo, no interior do qual se apresenta uma relação causal, de acordo com a qual “ter sido empurrado” é a causa que tem como efeito “cair”. Já em (4), estamos perante dois atos ilocutórios assertivos: um primeiro no qual o locutor asseire “A Joana está triste”, enunciado apresentado como *conclusão*; um segundo que corresponde à asserção “(pois) está a chorar”, que se assume como a razão para concluir a tristeza da Joana.

É importante acrescentar, ainda, que a relação causal se distingue da relação justificativa do ponto de vista prosódico, dado que esta última exige uma pausa entre os dois constituintes (o ato ilocutório 1 e o ato ilocutório 2), o que não ocorre numa relação de causa.

A argumentação positiva pode também estruturar-se segundo uma ordem progressiva, que corresponde à seguinte organização dos blocos textuais:

Argumento(s) – logo/portanto – Conclusão

Esquema 7

Neste caso, entre *conclusão própria* e *argumento próprio* estabelece-se uma relação conclusiva, na qual o locutor apresenta o bloco textual *conclusão própria* como conclusão do que é afirmado no *argumento próprio*, de acordo com o seguinte esquema:

Eu afirmo X, logo/portanto (concluo/posso concluir) Y

Esquema 8

140 Uma vez que estamos perante uma construção argumentativa, a relação conclusiva que se estabelece, tal como se verificou para a ordem regressiva, processa-se no plano epistémico-ilocutório. O nexu criado vem sinalizar uma relação entre dois atos ilocutórios independentes, que, no seu conjunto, formam um ato ilocutório argumentativo complexo (Van Eemeren 1984).

As relações de coerência num texto argumentativo poderão ainda associar-se a uma argumentação de orientação negativa. A argumentação negativa assenta numa relação criada entre um *contra-argumento* e uma *conclusão alheia* e/ou um *argumento alheio*. Neste caso, o bloco textual *contra-argumento* estabelece com o bloco textual alheio uma relação de refutação. O raciocínio associado ao *contra-argumento* é o seguinte:

O proponente afirma X – Eu, oponente, afirmo não X
(e, desta maneira, sustento a minha conclusão própria)

Esquema 9

4. O jogo escolar

Um texto argumentativo nasce, como vimos, da divergência de opiniões, de onde resulta uma produção textual marcada pela intenção persuasiva. Fruto deste contexto, um texto argumentativo é constituído por blocos textuais que o distinguem de outros tipos de texto. São estes mesmos blocos textuais que constroem entre si relações estruturais e de coerência muito específicas. Quando colocamos o enfoque nos textos argumentativos produzidos em contexto escolar, facilmente concluimos que estes não evidenciam algumas das características apresentadas nas secções anteriores, indicadas como essenciais para a existência de um texto argumentativo.

Julgamos que tal fica a dever-se, em parte, às condições de produção que envolvem a argumentação oral formal desenvolvida em contexto escolar. De facto, em espaço escolar, um texto argumentativo oral formal não resulta, habitualmente, de uma situação real de produção, mas antes de uma simulação. Tal ocorre porque a divergência de opiniões, que dá origem à argumentação, não se desencadeia naturalmente na sala de aula, mas surge, antes, como consequência de uma solicitação do professor. Assim, a exigência de produção oral feita pela Escola é geradora de artificialidade, pelo que somente o desenvolvimento e a exploração de uma simulação associada, neste caso, à oralidade argumentativa, poderão conferir sentido à expressão oral e, assim, gerar textos mais próximos daqueles que a Escola pretende que os alunos produzam ao longo do seu percurso, ou seja, textos que se aproximem mais das características textuais consideradas definitórias do tipo de texto argumentativo.

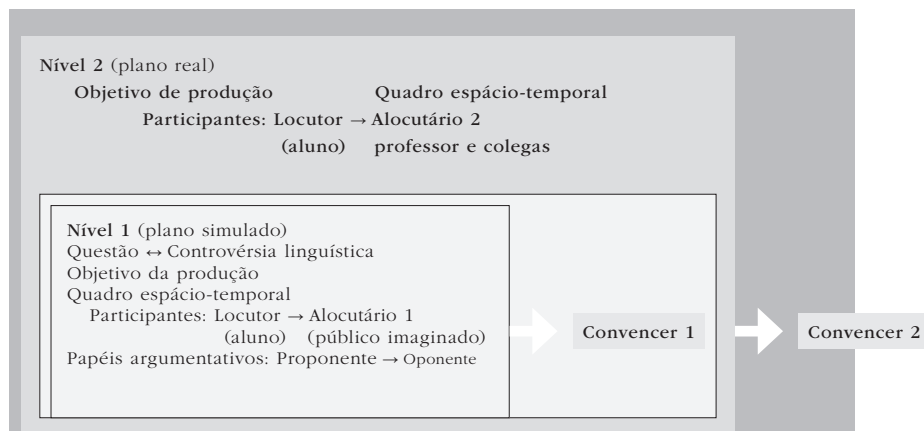
141

Ao processo que contempla a recriação consciente de um conjunto de elementos envolvidos numa argumentação oral formal daremos o nome de *jogo escolar*. O conceito em causa diz respeito a uma ação que deverá ser promovida, em contexto didático, de modo a desencadear a construção de um plano artificial que possa enquadrar a produção argumentativa, dotando-a, conseqüentemente, de sentido.

Deste modo, a produção de um texto argumentativo que contemple os elementos do *jogo escolar* desenvolve-se num plano simulado, onde o aluno, enquanto locutor, dirige um texto argumentativo ao alocutário 1, que corresponde ao público imaginado pelo aluno. Neste plano, o aluno procura, fazendo uso das diferentes componentes do jogo escolar, convocar um conjunto de *argumentos* que sejam válidos para confirmar ou refutar uma dada *conclusão*. Por meio deste movimento argumentativo, o locutor visa persuadir o alocutário 1, desenvolvendo um ato de convencer 1, que corresponde, também, a uma pretensão virtual, atendendo a que todo o contexto que envolve o texto argumentativo é simulado, construído como plausível. O mesmo texto argumentativo funciona, por outro lado, no plano real da produção verbal, envolvendo, como participantes, o aluno, enquanto locutor, e o professor e colegas, enquanto alocutários 2. Neste âmbito, o aluno, ao produzir um discurso argumentativo, estará, em última análise,

a realizar um ato de convencer 2, cujo conteúdo é um *argumento* em favor da *conclusão* “revelo capacidade de argumentação / de produção de textos argumentativos”, que surtirá efeito caso o professor e/ou os colegas admitam a capacidade/qualidade do aluno, avaliando de forma positiva a sua prestação. O esquema que se segue pretende ilustrar o contexto em que se deve desenvolver a produção argumentativa oral formal em contexto escolar:

142



Esquema 10

5. Textos argumentativos produzidos por alunos

A análise de textos argumentativos produzidos sem preparação prévia por alunos aponta para o facto de estas produções não contemplarem o elemento *controvérsia*, *conclusões alheias* ou *argumentos alheios*. Os textos em questão centram-se sobretudo na defesa da opinião pessoal, não recriando um contexto de inserção da produção argumentativa que contemple o eixo espaço-temporal, o número e natureza dos participantes na interação e o próprio objetivo da comunicação (elementos que o modelo de *jogo escolar* mostra serem fundamentais). No plano da estrutura argumental, os textos produzidos sem preparação revelam tendência para a ordem regressiva e para a estrutura argumental simples, assente nos seguintes blocos textuais: *questão*, *conclusão* e *argumento(s)*.

Para além destas características gerais, constatamos que os textos em análise exibem outros traços que importa destacar. Um deles é o facto de alguns apresentarem uma *conclusão dupla*, como se observa nos exemplos que se seguem:

(6) «por um lado sim casar é ter um companheiro [...] por outro lado não é preciso casar para ser feliz» A2[22/09/08] – 0:55

143

(7) «é assim eu tenho duas opiniões sobre isto» A9[22/09/08] – 2:11'

(8) «/// eu concordo e por outro lado também não concordo» B9[22/09/08] – 0:46'

A *conclusão dupla* corresponde, deste modo, a uma dualidade de resposta, que acaba por significar a introdução no texto de dois blocos textuais com função de *conclusão*, uma a favor de uma posição e outra de refutação dessa mesma posição. Estamos perante um processo de raciocínio em curso, onde se procura definir a opinião pessoal. Com efeito, o aluno, não conseguindo elaborar uma síntese da sua posição, desenvolve um raciocínio dialético, correspondente a um equacionar de posições opostas, ou pelo menos não coincidentes, que resulta numa forma de diálogo entre *conclusões* distintas, dentro do próprio texto. Esta situação ocorre, a nosso ver, porque a apresentação de um texto argumentativo exige um tempo de maturação prévio, no qual o aluno possa preparar, como afirmámos atrás, os eixos da produção discursiva e ainda definir a própria posição pessoal, pois «los adolescentes [...] pueden tener problemas para saber cuál es la tesis que quieren defender, sobre todo cuando se proponen temas de debate que no conocen lo suficiente o que tienen un carácter muy general.» (Cros in Vilà 2005, destaque nosso).

A apresentação de uma *conclusão dupla* leva, por outro lado, à mobilização de uma argumentação dupla, o que fica visível no excerto que se segue:

<i>Questão</i>	
A pena de morte devia ser abolida em todo o mundo	
<i>Conclusão1 e Conclusão 2</i>	
/ aa eu concordo e não concordo/ aa por um lado eu concordo por aa / por um lado eu não concordo	
<i>Argumento (Conclusão 1)</i>	<i>Argumento (Conclusão 2)</i>
porque / se nós a partir do momento que nascemos temos o direito à vida ee aproveitá-la e não desperdiçá-la /	[aa mas por outro lado não concordo] _{conclusão 2 (retoma)} porque ee quando há aquelas pessoas que estão mesmo a sofrer por exemplo acamadas / aa acho que essas / ee também aquelas que ee por exemplo andam por aí a matar essas também não merecem / não têm aa não merecem andar aqui

C5[22/09/08]- 0:36

Esquema 11

Nestes mesmos textos identificamos situações de incoerência entre a *conclusão* expressa inicialmente e a *conclusão* apresentada no final. Veja-se o texto de C3, no qual o aluno, inicialmente, assume não concordar com a *questão*, mas, como ao longo do texto equaciona duas posições distintas, acaba por apresentar uma opinião que é ambígua:

<i>Questão</i>	
Então a minha afirmação é casar é fundamental para a felicidade pessoal /	
<i>Conclusão1</i>	
eu não concordo nada acho que podemos ser felizes sem casar	
<i>Argumento (Conclusão 1)</i>	<i>Argumento (Conclusão 2)</i>
aliás acho que hoje em dia a maior parte das pessoas até se juntam mais em vez de casar acho que não é por assinar um papel que vamos ser mais ou menos felizes //	além disso também depende de da mentalidade das pessoas não é? / há pessoas que só são felizes em casar e ter filhos / ee e mais filhos e uma casa e um cão à porta /
<i>Conclusão (reformulada)</i>	
depende	

C3[22/09/08]- 0:37

Esquema 12

Noutras situações, fica claro que a *conclusão* não se encontra definida à partida, sendo construída à medida que se desenvolve o diálogo, interno ao texto, entre posições contrárias ou distintas. Esta situação está patente no texto de A14:

<i>Questão</i>	
Ser rico é uma questão de sorte e não de trabalho	
<i>Conclusão1 e Conclusão 2</i>	
aa eu concordo e discordo	
<i>Argumento (Conclusão 1)</i>	<i>Argumento (Conclusão 2)</i>
mas para uma pessoa ser rica precisa de ter sorte e ao mesmo tempo trabalhar muito porque ee sem sorte aa só com sorte e trabalho não se consegue uma pessoa ser muito rica uma pessoa pode-se esforçar muito pode não ser rica pode ser mediana uma pessoa que tenha dinheiro nem muito nem pouco se a pessoa também só tiver sorte também não vai conseguir ser rica pensar aa se eu jogar no euromilhões vou ser rico a pessoa assim também não vai ser rica	mas também há casos particulares de uma pessoa que nunca fez nada na vida e olha vou jogar no euromilhões / joga no euromilhões e calha-lhe aa sorte grande fica rico nunca fez nada na vida // e depois fica /// aa torna-se rico /// aa ///
<i>Conclusão1 e Conclusão 2- argumento</i>	
mas na minha opinião acho que para uma pessoa se tornar rica primeiro tem que estudar aa começar a estudar muito // para depois conseguir ter um bom emprego e ter um bocado de sorte na vida /	
<i>Coda</i>	
é só	

Esquema 13

Em síntese, nos textos argumentativos sem preparação prévia a estruturação textual pode orientar-se no sentido da reiteração da *conclusão* apresentada no início, ainda que esta seja dupla; noutras situações, apresenta-se mesmo a reformulação da posição inicial; ou, em certos casos, o aluno não chega a definir, de forma clara, uma *conclusão*.

No que respeita às relações de coerência estabelecidas no interior de textos argumentativos orais sem preparação prévia, fica clara a preferência dos alunos pela argumentação positiva. Julgamos que as razões justificativas desta situação são sobretudo de ordem contextual e pragmática. Teremos de recordar que os textos que consideramos são produzidos em ambiente escolar, o que condiciona algumas das suas características. Assim, verificamos que, nestes textos, a *conclusão* aparece, frequentemente, na sequência de uma questão apresentada pela própria professora. É assim natural que, em termos de gestão textual, o aluno se sinta compelido a iniciar o texto com uma “resposta” à questão colocada, apresentando a sua posição pessoal face ao tema proposto. Tal fica ilustrado pelo excerto que se segue:

(9) «[...] [tenho como tese principal oo oo a que o aborto deve só ser feito em casos urgentes]_{conclusão} [/// digo isto pois não não está eticamente correto uma pessoa matar uma pessoa aa porque ee /// porque tem de ser /// como sabem o embrião aa é uma pessoa em desenvolvimento por isso deve aa deve ser tratada como tal]_{argumento} [...]» B9 [04/11/08] – 2:10’

146

Não esqueçamos também que a colocação da *conclusão* em início de texto corresponde à organização preferencial dos blocos textuais no domínio da oralidade.

Se, por um lado, é um facto que a ordem regressiva condiciona, no âmbito de um texto argumentativo, o aparecimento de uma relação justificação, julgamos, por outro lado, que a opção por esta relação de coerência se fica a dever também ao contexto escolar de produção textual em que ela se enquadra. O aluno assume o seguinte raciocínio: «apresentei a minha posição, agora tenho de justificá-la». Esta organização acaba por corresponder ao modelo de texto de opinião que consideramos o modelo escolar por excelência. Trata-se de um modelo textual adequado a um contexto em que o locutor se sente avaliado, tendo, portanto, de provar as suas capacidades.

6. Conclusões

Os resultados da investigação-ação que desenvolvemos mostraram que existem claras diferenças entre os textos produzidos sem preparação prévia por alunos e o modelo prototípico de texto argumentativo. Em nossa opinião, este desfasamento deve-se em primeira análise à artificialidade introduzida pelo ambiente escolar. Uma vez que, em espaço de sala de aula, os alunos não são capazes de recriar o contexto de produção de um texto argumentativo, o ensino da argumentação oral formal poderá tirar vantagens do modelo que corresponde ao *jogo escolar*. Defendemos, deste modo, que a Escola, através da sua ação didática, deve promover junto dos seus alunos a consciência de que a produção argumentativa oral se inscreve num plano simulado

que terá de assentar na ativação de um conjunto de eixos estruturantes da produção verbal, que são suscetíveis de introduzir coerência no produto verbal. Assim, a produção argumentativa não deverá, antes de mais, assentar no recurso ao improviso. Por outro lado, na preparação do texto argumentativo, dever-se-á selecionar uma questão problemática adequada (que se centre, inicialmente, em temas concretos e objetivos, próximos da realidade dos alunos) e ter em consideração o domínio da persuasão, enquanto objetivo último do texto argumentativo. Será importante, também, promover a consciência do confronto linguístico subjacente a qualquer texto argumentativo. Este domínio poderá conduzir ao desenvolvimento da capacidade de definição da *conclusão própria* e da seleção de *argumentos* pertinentes e ainda à diversificação dos tipos de *argumentos* a utilizar. Noutro plano, a noção de planificação não se deverá esgotar na organização das ideias a apresentar, mas antes alargar-se ao trabalho dos domínios estruturantes da produção verbal, de acordo com as regras do *jogo escolar*.

Por outro lado, consideramos que será partindo da realidade das produções textuais dos alunos que se poderão identificar as áreas específicas do texto argumentativo a trabalhar em sala de aula. Assim, em termos concretos, destacamos o facto de os alunos mobilizarem preferencialmente nos seus textos os blocos textuais *conclusão* e *argumento próprio*. Recordemos, porém, que o mesmo já não se verifica relativamente a outros blocos textuais, como os *argumentos albeios* ou a *conclusão albeia*, o que significa que a Escola deverá trabalhar explicitamente os modelos estruturais associados à contra-argumentação, associando-os ao estudo das relações de coerência.

Não obstante a importância que assumem todos os elementos associados à dinâmica do *jogo escolar*, convirá que a Escola não reduza a sua ação ao tratamento da competência linguística, contemplando também a competência comunicativa (cf. Kerbrat-Orecchioni 1998 [1990]: 51), pois a aprendizagem da argumentação oral formal só fará sentido se contribuir para a formação de cidadãos mais conscientes, críticos e interventivos.

Referências

148

- Adam, Jean-Michel (2001). *Les textes: types et prototypes. Récit, description, argumentation, explication et dialogue*. 4^e ed.. Paris: Nathan.
- Aristóteles (2006). *Retórica*. 3.^a ed., Tradução e notas de Manuel Alexandre Júnior, Paulo Farmhouse Alberto e Abel do Nascimento Pena. Lisboa: Centro de Filosofia da Universidade de Lisboa e Imprensa Nacional Casa da Moeda.
- Kerbrat-Orecchioni, Catherine (1998 [1990]). *Les interactions verbales. Approche conversationnelle et structure des conversations*. Tome 1, 3^e édition. Paris: Armand Colin.
- Lopes, Ana Cristina Macário (2009). Justification: a coherence relation. *Pragmatics*, n.º 19, 2, 241-252.
- Lopes, Ana Cristina Macário (1997). A argumentação: uma área de investigação pluridisciplinar. *Revista Portuguesa de Filologia* (separata), vol. XXI. Coimbra: Faculdade de Letras da Universidade de Coimbra, 157-175.
- Plantin, Christian (2005). *L'argumentation*. Paris: PUF.
- Plantin, Christian (1996). *L'argumentation*. Paris: Editions du Seuil.
- Perelman, Chaïm e Olbrechts-Tyteca, Lucie (2008 [1958]) *Traité de l'argumentation*. 6^e édition. Bruxelles: Editions de l'Université de Bruxelles.
- Rubattel, Christian (ed.) (1989). *Modèles du discours: recherches en Suisse Romande*. Berne, Francfort-s. Main, New York, Paris: Lang.
- Van Eemeren, Frans H. e Grootendorst, Rob (1984). *Speech Acts in Argumentative Discussions. A Theoretical Model for the Analysis of Discussions Directed towards Solving Conflicts of Opinion*. Dordrecht: Foris Publications.
- Vilà i Santasusana, Montserrat *et al.* (coord.) (2005). *El discurso oral formal. Contenidos de aprendizaje y secuencias didácticas*. Barcelona: Graó.

REFUTAÇÃO E COERÊNCIA PRAGMÁTICA

1. Introdução

Neste trabalho, pretendemos evidenciar a natureza pragmática da ‘refutação’, entendida como a relação discursiva que se estabelece entre um enunciado cuja função é rejeitar um outro e o enunciado que, por sua vez, é alvo de refutação⁸⁷. Este aspeto tem sido objeto de particular atenção nos estudos que se debruçam sobre a negação metalinguística, apontada com um dos meios privilegiados para a realização deste ato discursivo. Um dos argumentos subjacentes à defesa desta posição diz respeito aos aspetos que podem estar na base de uma refutação e, mais precisamente, ao facto de esta não poder ter no seu escopo o conteúdo proposicional do enunciado que é refutado (Horn 1985, 1989, 2002). No entanto, a não validade do que é asserido num determinado enunciado parece, intuitivamente, uma das razões privilegiadas para a sua rejeição. Assim, neste trabalho procuraremos demonstrar que a refutação pode ter efetivamente por base o conteúdo proposicional de um determinado enunciado, mas que tal não é incompatível com a defesa de que esta é, antes de mais, uma relação que se estabelece ao nível pragmático e não ao nível semântico.

⁸⁷ Este trabalho insere-se no meu projeto de doutoramento, financiado pela Fundação para a Ciência e Tecnologia (SFRH/BD/35809/2007).

2. A relação discursiva ‘refutação’

150 Nos estudos existentes sobre a refutação⁸⁸, esta é entendida, regra geral, como a relação que se estabelece entre um enunciado cuja função é rejeitar um outro e o enunciado que, por sua vez, é rejeitado ou refutado. A refutação será assim, por definição, uma relação de ordem pragmática: o que está em causa é, antes de mais, uma relação entre atos ilocutórios e não uma relação entre conteúdos proposicionais.

Vejam-se as intervenções de B nas sequências seguintes:

(1) A: The king of France is *bald*.

B: The king of France isn't *bald*, (because) there is no king of France.

(Horn, 1989: 362)

(2) A: * Tu é que *dissestes* isso?

B: Não é: *dissestes!* É: *disseste!*

(3) A: Este livro é *super fixe*, professora.

B: Este livro é *muito bom*. “Super fixe” não é uma expressão para utilizar na aula.

Em todos estes exemplos, a intervenção do locutor B tem um por função refutar o enunciado de A: em (1), um exemplo adotado por Horn a partir de um célebre artigo de Russell (1905)⁸⁹, a refutação tem por base a não verificação do conteúdo pressuposicional associado ao sintagma nominal

⁸⁸ Adotamos aqui o termo utilizado em Moeschler (1982). Nos trabalhos de matriz anglo-saxónica, esta relação é geralmente designada de ‘denial’.

⁸⁹ Numa reação à proposta de Frege que, no célebre artigo «On sense and reference» (1970 [1892]), defende que as pressuposições não fazem parte do conteúdo proposicional dos enunciados em que ocorrem, Russell, no igualmente célebre artigo «On denoting» (1905), recorre ao enunciado ‘The present king of France is bald’ (p. 490) para demonstrar que o que Frege designa de conteúdo pressuposicional faz igualmente parte do que é asserido. Assim, segundo Russell, um enunciado como o anterior, na ausência de tal entidade, será necessariamente falso. No entanto, quando o mesmo se apresenta sob uma forma negativa, tal já não acontecerá necessariamente: « (...) the present King of France is not bald” is false if it means “There is an entity which is now King of France and is not bald”, but is true if it means “It is false that there is an entity which is now King of France and is bald”» (p.490). A proposta de Russell esteve na origem da distinção entre os conceitos de negação interna

‘The King of France’, i.e., a pressuposição de que tal entidade existe; em (2), o que está em causa não é o que é dito ou implicado no enunciado alvo de refutação, mas sim a forma como algo é dito – no exemplo em questão, a incorreta flexão do verbo *dizer* na segunda pessoa do singular do Pretérito Perfeito Simples do Indicativo; finalmente, em (3), é também, em sentido lato, o modo como algo é dito que se encontra na base do ato refutativo – no exemplo em causa, a inadequação do termo *super fixe* ao registo de língua que seria adequado à situação enunciativa.

Embora a refutação possa assumir uma forma afirmativa (cf. (3)), um enunciado contendo uma negação metalinguística tem sido apontado como um dos meios privilegiados para a sua realização linguística (cf. (1) e (2)). Este tipo de negação, na terminologia de Ducrot (1972, 1973) / Horn (1985, 1989), caracterizar-se-á justamente por ter no seu escopo um enunciado e não uma proposição, i.e., por não operar ao nível proposicional, mas ao nível metalinguístico.⁹⁰

Dada a sua natureza, este tipo de sequências refutativas ocorrerá prototipicamente em textos dialogais, sendo particularmente frequente em contextos que favoreçam interações de carácter mais polémico, como sejam o Discurso Político ou, ainda que em termos diversos, o Discurso Didático.

2.1. Refutação e valor de verdade

Os trabalhos que se debruçam sobre os aspetos que podem estar na base de uma refutação partem, regra geral, da análise da negação metalinguística proposta em Horn (1989):

e negação externa, que corresponderão, numa formulação mais atual, aos conceitos de negação descritiva e negação metalinguística.

⁹⁰ Por não operar ao nível proposicional, a negação metalinguística tem características que a distinguem da negação descritiva a vários níveis. Para além de poder ter no seu escopo as pressuposições linguísticas (comportamento que esteve justamente na origem da distinção entre os dois tipos de negação – cf. nota anterior, é compatível com termos de polaridade positiva que, por definição, não são aceitáveis em contextos negativos, e nunca assume a forma de uma negação morfológica. Os enunciados em que ocorre este tipo de negação são também marcados, do ponto de vista prosódico, por uma entoação contrastiva.

While two distinct uses of sentential negation must indeed be admitted, the marked, non descriptive variety is not a truth-functional or semantic operator on propositions, but rather an instance of the phenomenon of METALINGUISTIC NEGATION – a device for objecting to a previous utterance on any grounds whatever, including the conventional or conversational implicata it potentially induces, its morphology, its style or register, or its phonetic realization.

(Horn, 1989:363)

Os exemplos seguintes ilustram alguns dos elementos referidos acima:

(4) A: O Pedro *conseguiu* ter positiva no teste.

B: O Pedro não *conseguiu* ter uma positiva no teste! O teste foi tão *fácil* que qualquer um teria positiva!

(5) A: É verdade que tivemos um prejuízo de *alguns* milhões, mas estamos a trabalhar para corrigir essa situação.

B: Não foi um prejuízo de *alguns milhões*, Sr. Ministro. Foi um prejuízo de *muitos milhões*!

(6) A: * Acho que o jornalista *entreviu* no momento certo.

B: O jornalista não *entreviu* – o jornalista *interveio*!

(7) A: Queres *um cimbalino*?

B: Não quero *um cimbalino*. Quero um *café*.

(8) (Esker too ah coo-pay luh vee-and?)

Non, je n'ai pas 'coo-pay luh vee-and' – j'ai coupé la viande. (Horn, 1989: 371)

Em (4), o locutor B refuta o enunciado de A com base na implicatura convencional que dele decorre – neste caso, a implicatura de que terá sido necessário esforço para ter uma boa nota no teste, desencadeada pela utilização da forma verbal *conseguiu*; em (5), a refutação tem por base uma

implicatura conversacional – neste caso, a *Q-implicatura* desencadeada pela utilização do predicado escalar *alguns*⁹¹; em (6), o que está em causa é o modo como algo é dito – no exemplo em questão, a incorreta flexão do verbo *intervir* na terceira pessoa do singular do Pretérito Perfeito Simples do Indicativo; em (7), é também, em sentido lato, o modo como algo é dito que se encontra na base do ato refutativo – no exemplo em causa, a utilização da palavra *cimbalino*, associada a um determinado dialeto da língua portuguesa e considerada menos padronizada do que o termo *café*; finalmente, em (8), o que está em causa, para além da errónea atribuição do género masculino a *viande*, é «the woeful English accent» (Horn 1989: 371) com que a frase em francês é pronunciada.

Tal como é evidenciado por Carston (1996), embora Horn afirme que um enunciado pode ser alvo de refutação ‘on any grounds whatever’, a verdade é que o seu valor de verdade nunca é referido como uma das razões que podem levar à sua refutação. Aliás, tal é claramente assumido por Horn que, num artigo publicado em (2002), afirma que «The target of “classical” MN [metalinguistic negation] is what is not asserted: material that is not part of explicit content and/or not communicated» (p. 79).⁹²

No entanto, não parece razoável afirmar que a falsidade da proposição associada a um determinado enunciado não possa ser uma das razões – porventura, até uma das mais pertinentes – para a sua refutação. Aliás, tal parece acontecer justamente nalguns exemplos apontados por Horn (1989: 437):

- (9) Negation is ambiguous not semantically but pragmatically.
- (10) I saw not Chris but Pat.
- (11) John was born not in Boston, but in Philadelphia.

⁹¹ Na proposta de Levinson (2000) aqui adotada, o termo *Q-implicatura* diz respeito às implicaturas conversacionais generalizadas que resultam da aplicação da primeira submáxima da Quantidade de Grice (1975 [1967]), podendo, *grosso modo*, ser parafraseáveis por *O que não é dito, não se verifica*. Assim, no exemplo em análise, a *Q-implicatura* gerada pela utilização do predicado escalar *alguns (milhões)* seria a de que o prejuízo em questão foi apenas de alguns (milhões) e não mais do que isso (na escala em questão).

⁹² Yoshimura (2002) partilha a posição de Horn, defendendo que, no escopo de uma refutação, estará «what is necessarily accompanied but not communicated unintentionally by the attributed utterance» (p. 129). Esta formulação parece, no entanto, pouco clara.

Esta questão foi utilizada como um dos argumentos para a defesa de que, ao contrário do que propõe Horn (1985, 1989), também no seu uso metalinguístico a negação é um operador vero-condicional (cf., e.o., van der Sandt, 1991; Carston, 1996; Geurts, 1998).

154 No entanto, tendo em conta que os defensores desta posição não negam que no escopo deste tipo negação possam estar aspetos de ordem mais formal, não diretamente relacionados com o que é dito ou implicado no enunciado alvo de refutação, esta posição parece muito difícil de sustentar.⁹³ Tal é evidenciado por Horn (1992) que, a este propósito, afirma:

(...) any such attempt to propositionalize not only upper-bounding implicata but the stylistic, connotative, and mechanical aspects of utterances that fall within the scope of marked negation would seem self-defeating, representing a kind of category mistake: an 'echoic use' is not the sort of beast to which a truth-functional operator applies.⁹⁴

(Horn, 1992: 172)

Concordando com estas observações e com o modo como, em termos gerais, a distinção entre a negação descritiva e a negação metalinguística é equacionada em Horn (1985, 1989), consideramos, no entanto, que a

⁹³ As dificuldades levantadas pela defesa de uma posição deste tipo estão bem patentes no modo como Geurts (1998) analisa um exemplo em que estão em causa questões de ordem fonética:

«Horn, presumably, would object to this [the existence of only one type of negation operator, which is truth-functional] on the grounds that a truth-functional operator could never apply, for example, to phonetic information. But there is really no deep mystery about this. Consider (30):

(30) A: Your wife reminds me of Frankentst[iʷ]n.

B: She doesn't remind you of Frankentst[iʷ]n: she reminds you of Frankentst[áʷ]n.

The propositional content of A's utterance is that B's wife reminds him of Frankenstein, and his pronunciation of the monster's name doesn't enter into this. But by uttering his line in this particular way, A commits himself to the truth of at least one further proposition, to wit that the name is pronounced Frankentst[iʷ]n. Therefore, this proposition is part of the information conveyed by A's utterance, and because B echoes this utterance it lands within the scope of an ordinary truth-functional, negation operator.» (1998: 284-285).

⁹⁴ Carston (1996) defende que a característica principal da negação metalinguística é o facto de os elementos que podem estar no seu escopo serem 'echoically used': «The correct generalization about the metalinguistic cases is that the material in the scope of the negation operator, or some of it at least, is echoically used (...). A representation is used echoically when it reports what someone else has said or thought and expresses an attitude to it.» (1996: 317).

exclusão do conteúdo proposicional do conjunto de elementos que podem levar à refutação de um determinado enunciado é dificilmente defensável. Contudo, e ao contrário do que defendem van der Sandt (1991), Carston (1996) ou Geurts (1998), consideramos que tal não implica que a negação presente em contextos deste tipo possa ser considerada um operador vero-condicional, semelhante ao que ocorre no uso descritivo da negação. De facto, e este parece ser o aspeto essencial, mesmo em contextos deste tipo a negação continuará a ter no seu escopo um enunciado e não uma proposição, tendo por função a refutação de uma determinada asserção e não a descrição de um estado de coisas.

3. Refutação e coerência pragmática

Entre os fatores que concorrem para a construção de uma representação mental (coerente) dos textos/discursos, é hoje consensual que a coerência relacional – propriedade que garante a existência de um nexos (ou interligação consequente) entre os vários segmentos textuais⁹⁵ – desempenha um papel fundamental, daí decorrendo a existência de um grande número de propostas de análise das relações sobre as quais esta se edifica.

Pese embora a multiplicidade de critérios que têm sido convocados na consideração de tais relações, a distinção entre os domínios de significação envolvidos na conexão realizada parece reunir atualmente um largo consenso na comunidade linguística (cf. Sanders 1997). De facto, haverá relações que operam no plano ideacional da significação, relacionando as situações ou estados de coisas do mundo sociofísico para o qual remete o conteúdo proposicional dos segmentos textuais conectados⁹⁶ e contribuindo, assim,

⁹⁵ Na literatura especializada, tais nexos são correntemente designados de conexões/relações discursivas, relações retóricas ou relações de coerência.

⁹⁶ Um exemplo paradigmático deste tipo de conexões será a relação de *causa*:

(i) O Rui faltou às aulas porque está doente.

Neste exemplo, adotado de Lopes (2009: 3), a conexão explicitada por *porque* envolve justamente uma relação causal entre duas situações do mundo sociofísico, sendo parafraseável por *O facto de o Rui estar doente é a causa do facto de ele ter faltado às aulas*. Esta relação tem também sido designada de *causal de re*.

para a coerência (mais estritamente) semântica das sequências em que ocorrem. Por outro lado, haverá conexões que operam no plano interpessoal da significação, envolvendo o valor ilocutório dos segmentos conectados⁹⁷ e contribuindo, assim, para a coerência (mais estritamente) pragmática das sequências em que ocorrem.⁹⁸

Como é evidenciado pela própria definição da refutação, esta será uma relação discursiva que envolverá, antes de mais, o plano interpessoal da significação: na conexão que se estabelece entre o enunciado refutativo e o enunciado alvo de refutação, o que está em causa é uma relação entre atos ilocutórios e não uma relação entre conteúdos proposicionais.

Este aspeto torna-se particularmente evidente se tivermos em conta que, na base de uma refutação, podem estar elementos que nada têm a ver com o que é asserido no enunciado que é rejeitado. No entanto, e mesmo nos casos em que a refutação tem por base o valor de verdade de tal enunciado, a conexão relevante continua a estabelecer-se no plano ilocutório e não no plano semântico. Aliás, se assim não fosse, uma sequência como a seguinte seria necessariamente contraditória (e, por conseguinte, incoerente), o que na realidade não se verifica:

(12) A: O Pedro é o presidente da empresa.

B: O Pedro não é o presidente da empresa! (O presidente é o João.)

⁹⁷ Um exemplo paradigmático deste tipo de conexões será, por sua vez, a relação de *justificação*:

(i) Está gente em casa, porque as luzes estão acesas.

Neste exemplo, igualmente adotado de Lopes (2009: 2), a conexão explicitada por *porque* envolve não uma relação entre duas situações do mundo sociofísico, mas antes uma relação entre uma asserção com valor de premissa e uma asserção com valor de conclusão, sendo parafraseável por *O facto de as luzes estarem acesas leva-me a concluir/a afirmar que está alguém em casa*. Esta relação tem também sido designada de *causal de dicto*.

⁹⁸ Os planos ideacional e interpessoal da significação não são, obviamente, independentes. A divisão entre relações de ordem semântica e relações de ordem pragmática apenas pretende evidenciar o domínio de significação mais relevante para uma determinada conexão discursiva.

4. Síntese

Como procurámos demonstrar, entre as múltiplas razões que podem levar à refutação de um determinado enunciado, o valor de verdade do seu conteúdo proposicional será certamente um aspeto importante. No entanto, tal não é incompatível com a defesa de que esta é, antes de mais, uma relação que se estabelece ao nível pragmático e não ao nível semântico. De facto, independentemente dos aspetos que podem estar na base da sua ocorrência, esta é uma relação entre atos ilocutórios e não entre conteúdos proposicionais.

157

5. Referências

- Carston, Robyn (1996). Metalinguistic negation and echoic use. *Journal of Pragmatics* 35, 309-330.
- Ducrot, Oswald (1972). *Dire et ne pas dire*. Paris: Hermann.
- Ducrot, Oswald (1973). *La preuve et le dire*. Paris: Mame.
- Frege, Friederich Gottlob (1970 [1892]). On sense and reference. In: Peter Geach e Mark Black (eds.). *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell, 56-78.
- Geurts, Bart (1998). The mechanisms of denial. *Language* 74, 274-307.
- Horn, Laurence R. (1985) Metalinguistic Negation and Pragmatic Ambiguity. *Language* 61:1, 121-174.
- Horn, Laurence R. (1989). *A Natural History of Negation*. Chicago: The University of Chicago Press.
- Horn, Laurence R. (1992). The said and the unsaid. *Semantics and Linguistic Theory (SALT)* II. Ithaca: Cornell University, 163-192.
- Horn, Laurence R. (2002). Assertoric inertia and NPI licensing. In *Papers from the Chicago Linguistic Society* 38(2), 55-82.
- Kehler, Andrew (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford: CSLI Publications.
- Levinson, Stephen C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge: MIT Press.
- Lopes, Ana C. M. Lopes (2009) Justification: A coherence relation. *Pragmatics*, 19:2, 241-252.
- Mann, W. C. & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8:3, 243-281.
- Moeschler, Jacques (1982). *Dire et Contredire – Pragmatique de la négation et acte de réfutation dans la conversation*. Berne: Peter Lang.
- Russell, Bertrand (1905). On Denoting. *Mind*, 14. Oxford: Oxford University Press, 479-493.
- Sanders, Ted (1997). Semantic and pragmatic sources of coherence: on the categorization of coherence relations in context. *Discourse Processes* 24, 119-147.
- Schwenter, Scott (2000). Viewpoints and polysemy: linking adversative and causal meanings of discourse markers. In: Elizabeth Couper-Kuhlen & Bernd Kortmann (eds.). *Cause, Condition, Concession, Contrast: Cognitive and Discourse Perspectives*. Berlin: Mouton de Gruyter, 257-281.
- Van der Sandt, Rob (1991). Denial. *Papers from the Chicago Linguistic Society* 27(2), 331-344.
- Yoshimura, Akiko (2002). A cognitive-pragmatic approach to metalinguistic negation. *Proceedings of the Sophia Symposium on Negation*. Sophia: Sophia Linguistic Institute for International Communication, 113-132.

(Página deixada propositadamente em branco)

SECÇÃO 6:
PROCESSAMENTO AUTOMÁTICO
DA LÍNGUA PORTUGUESA

(Página deixada propositadamente em branco)

ONTO.PT: CONSTRUAO AUTOMATICA DE UMA ONTOLOGIA LEXICAL PARA O PORTUGUES

1. Introduao

As bases de conhecimento lexical de grande cobertura tem um papel cada vez mais importante no desenvolvimento de ferramentas computacionais que necessitam de interpretar informaao transmitida atraves de linguagem natural. Para o ingles, a WordNet de Princeton, WN.Pr (Fellbaum 1998), e amplamente utilizada na realizaao de tarefas de processamento de linguagem natural (PLN), desde a determinaao de semelhanas (Agirre *et al.* 2009), passando pela desambiguaao do sentido das palavras (Resnik 1995) ou pela resposta automatica a perguntas (Pasca e Harabagiu 2001).

E indiscutivel que a existencia de uma *wordnet* potencia o desenvolvimento de ferramentas de PLN para a sua lngua, da a expansao desta forma de recurso para varias lnguas. Ainda assim, ha que referir que a existencia de uma *wordnet* no implica o fim da investigaao na extracao automatica de conhecimento lxico-semantico. Mais uma vez, para o ingles, ha trabalhos que tem, por exemplo, o objectivo de ampliar a WN.Pr (veja-se (Hearst 1998)).

Neste artigo e apresentado o Onto.PT (Gonalo Oliveira e Gomes 2010), um projecto que tem o objectivo de criar uma alternativa publica dentro do paradigma das ontologias lexicais para o portugues. Esta ontologia sera construida de forma automatica e estruturada de forma semelhante a uma *wordnet*.

Muito resumidamente, numa primeira fase de construao, sao extraidas instancias de relaoes semanticas, e numa segunda procuram-se aglomerados

de palavras relacionadas por sinonímia de maneira a formar conceitos. Por fim, procuram-se os conceitos mais adequados a cada argumento, uma palavra, das demais relações extraídas. Apesar de, na primeira versão do Onto.PT, terem apenas sido explorados dicionários e *thesauri*, o processo de construção é suficientemente flexível para, no futuro, poder incorporar conhecimento de outros recursos, tais como a Wikipédia.

162

Para além de apresentarmos, brevemente, a metodologia seguida na construção do Onto.PT, este artigo foca-se em vários exemplos, não só para complementar a explicação de cada uma das fases de construção, mas também para ilustrar alguns dos resultados mais interessantes a que, até ao momento, chegamos.

Além de uma nova rede lexical e de um novo *thesaurus* electrónico público para o português, que acreditamos serem os maiores do seu tipo, as principais contribuições deste projecto são um novo recurso lexical para o português, e uma metodologia flexível que poderá ser adaptada à construção de recursos semelhantes para outras línguas, e que proporciona ao Onto.PT um grande potencial de crescimento.

Iniciamos este artigo com o enquadramento deste trabalho, onde referimos alguns trabalhos relacionados e recursos semelhantes. Depois de explicitarmos a abordagem adoptada na construção do Onto.PT, apresentamos alguns resultados obtidos, dando exemplos de relações extraídas e *synsets* descobertos.

2. Enquadramento

A necessidade de realizar tarefas de PLN onde, para além de reconhecer as palavras e as suas interações, é crucial interpretar o significado do texto, levou à criação de recursos léxico-semânticos de ampla cobertura, tais como as ontologias lexicais, recursos que têm ao mesmo tempo propriedades de um léxico e de uma ontologia (Hirst 2004, Prévot *et al.* 2010).

Do ponto de vista da engenharia informática, uma ontologia é um sistema formal baseado em conceitos, regra geral de um domínio, e relações entre eles. Por outro lado, um léxico é um recurso linguístico que contém

conhecimento sobre uma língua e inclui, por isso, itens lexicais, como as palavras. No entanto, como, na realidade, as relações semânticas não se estabelecem entre palavras, mas entre significados (sentidos de palavras), e como os significados são inerentemente conceptuais, as bases de conhecimento lexical podem ser vistas como ontologias lexicais, onde conceitos da linguagem natural se encontram descritos através de palavras.

Como a construção de uma ontologia é uma tarefa tipicamente complexa que requer muito tempo, vários autores desde cedo estudaram formas de automatizar este processo. Tendo em conta a sua estrutura, os dicionários electrónicos foram dos primeiros alvos a ser explorados na construção automática de ontologias lexicais. Após analisar as regularidades presentes nas definições, Amsler (1981) chegou à conclusão que não seria complicado desenvolver procedimentos automáticos que, tirando partido de padrões lexicais ou léxico-sintácticos empregues, permitissem a extracção de relações de hiperonímia (Chodorow *et al.* 1985) ou mesmo a construção de bases de conhecimento lexical, como a MindNet (Richardson *et al.*, 1998), que inclui um vasto conjunto de relações semânticas.

Apesar de tudo, foram apontados problemas à utilização de um dicionário para esta tarefa (Ide e Veronis 1995). Para além de nem sempre estarem disponíveis para fins de investigação, verificou-se que dicionários diferentes apresentam normalmente a informação de forma diferente (por exemplo, não cobrem exactamente da mesma forma as mesmas partes do léxico) e remover, por vezes, até se complementam. Assim, a utilização de vários dicionários é vista como uma forma de ultrapassar alguns dos problemas identificados.

Com o estabelecimento da WN.Pr como uma base de conhecimento lexical amplamente utilizada no inglês, a investigação na extracção automática deste tipo de conhecimento a partir de dicionários diminuiu consideravelmente. Ainda assim, os dicionários continuam a ser utilizados na construção de ontologias (Nichols *et al.* 2005), onde se inclui o projecto PAPEL (Gonçalo Oliveira *et al.* 2010), que constitui uma rede lexical para o português.

A WN.Pr é um recurso criado manualmente por especialistas e baseado em *synsets* – grupos de palavras sinónimas que traduzem conceitos da linguagem natural. Cada *synset* tem uma definição, semelhante às entradas de um dicionário, sendo possível estabelecer várias relações semânticas

(e.g. hiperonímia, parte-de) entre *synsets*. Além de ser uma alternativa interessante à representação computacional de palavras e sentidos, a WN.Pr tornou-se num recurso amplamente utilizado devido ao seu carácter público. No entanto, sendo uma base de conhecimento lexical, não contempla conhecimento específico de determinados domínios, o que levou a que investigadores a tenham enriquecido com conhecimento extraído de outros recursos, como corpora (veja-se (Hearst 1998), (Lin e Pantel 2002)) ou enciclopédias, onde se inclui a Wikipédia (ver (Medelyan *et al.* 2009)).

O sucesso da WN.Pr levou a que fosse adaptada a outras línguas, incluindo as que se inserem nos projectos EuroWordNet (Vossen 1997) e MultiWordNet (Pianta *et al.* 2002). No entanto, a criação de uma *wordnet* requer muito tempo e envolve um grande volume de trabalho manual. Isto levou a que autores propusessem a tradução, de forma automática, da WN.Pr (Melo e Weikum 2008) para outras línguas. No entanto, e apesar desta abordagem provavelmente se adequar a algumas tarefas, línguas diferentes representam diferentes realidades socio-culturais: não cobrem exactamente a mesma parte do léxico e, por vezes, conceitos diferentes são lexicalizados de forma diferente (Hirst 2004). Assim, acreditamos que a construção da *wordnet* para uma língua deve ser desenvolvida de raiz para essa língua.

Para a língua portuguesa existem actualmente duas *wordnets* proprietárias, a WordNet.PT⁹⁹ (Marrafa 2002) e a MultiWordNet.PT (MWN.PT)¹⁰⁰. Ambas (ou partes de ambas) estão disponíveis para pesquisa através da rede. Além disso, a MWN.PT, que se encontra alinhada com a WN.Pr, pode ser adquirida. Ambas sofrem também pelo seu método de construção manual, que os limita em termos de crescimento e cobertura (Santos *et al.* 2010). No contexto das ontologias lexicais, existem mais alguns recursos para a o português que devem ser mencionados:

⁹⁹ <http://www.clul.ul.pt/clg/eng/wordnetpt/index.html>

¹⁰⁰ <http://mwnpt.di.fc.ul.pt/>

- Dois *thesauri*: TeP¹⁰¹ (Maziero *et al.* 2008), para a variante brasileira, que acreditamos ser a base de *synsets* para a *wordnet* brasileira (WordNet.Br (Dias-da-Silva *et al.* 2002)); e o OpenThesaurus.PT¹⁰², um *thesaurus* colaborativo utilizado no processador de texto do OpenOffice¹⁰³ para sugestão de sinónimos;
- Wikcionário¹⁰⁴, um dicionário colaborativo onde, além de definições, é possível adicionar informação acerca de relações semânticas para cada entrada. No entanto, a versão portuguesa deste recurso é ainda pequena e limitada;
- PAPEL¹⁰⁵ (Gonçalo Oliveira *et al.* 2010), uma rede lexical pública que contém instâncias de vários tipos de relações semânticas, extraídas de forma automática a partir da edição de 2005 do Dicionário da Língua Portuguesa (DLP 2005). As principais diferenças entre o PAPEL e uma *wordnet* são a construção automática do PAPEL e a sua estrutura baseada em palavras, ao contrário de *synsets*.

Mais informação sobre alguns destes recursos pode ser encontrada em (Santos *et al.* 2010). Uma comparação focada nos verbos encontra-se em (Teixeira *et al.* 2010). Ambos os trabalhos de comparação verificaram que, apesar de se tratarem assumidamente de recursos léxico-semânticos de grande cobertura, o seu conteúdo se complementa. Logo, a utilização não de um, mas de vários destes recursos, seria positiva na construção de um recurso único, com maior cobertura. Este problema parece ser comum a outras línguas, tais como a inglesa, onde, como referido anteriormente, há vários trabalhos com o objectivo de enriquecer a WN.Pr.

¹⁰¹ <http://www.nilc.icmc.usp.br/tep2/>

¹⁰² <http://openthesaurus.caixamagica.pt/>

¹⁰³ <http://www.openoffice.org/>

¹⁰⁴ <http://pt.wiktionary.org/>

¹⁰⁵ <http://www.linguateca.pt/PAPEL/>

3. Abordagem

166 O Onto.PT é uma ontologia lexical para o português, estruturada de forma semelhante a uma *wordnet*, que pretende ser uma nova alternativa no campo dos recursos lexicais para a nossa língua. É criada de forma automática a partir de recursos textuais, seguindo uma abordagem flexível que permite a integração de conhecimento presente em vários recursos, tais como os apresentados anteriormente. A abordagem para a construção do Onto.PT consiste num procedimento automático com três fases, que se descrevem de seguida.

3.1. Extracção de relações

Nesta fase é necessário construir, manualmente, um conjunto de gramáticas com padrões textuais indicadores de um conjunto pré-definido de relações semânticas. O texto é depois processado por um analisador sintáctico que utiliza as gramáticas na extracção automática de instâncias de relações, representadas através de triplos $t=(a,R,b)$, onde a e b são palavras e R é o nome de uma relação estabelecida entre um sentido de a e um sentido de b .

Para a primeira versão do Onto.PT, apenas foram processadas definições de dicionários. O procedimento de extracção utilizado é inspirado na construção do PAPEL (ver (Gonçalo Oliveira *et al.* 2010)). Sendo assim, a partir de definições como:

candeia s.f. utensílio doméstico rústico usado para iluminação, com pávio abastecido a óleo
espiga s.f. parte das gramíneas que contém os grãos
inquietar v.t. causar ansiedade
severo adj. grave , crítico

É possível tirar partido dos padrões textuais sublinhados para extrair, por exemplo, os seguintes triplos:

utensílio HIPERONIMO_DE *candeia*
iluminação FINALIDADE_DE *candeia*
espiga PARTE_DE *gramínea*
grão PARTE_DE *espiga*
inquietar CAUSADOR_DE *ansiedade*
grave SINONIMO_DE *severo*
crítico SINONIMO_DE *severo*

3.2. Descoberta de synsets

Ao olhar apenas para as relações de sinonímia, verifica-se que estas tendem a formar aglomerados (vulgo *clusters*) de palavras. Nesta fase, aglomerados são identificados de forma automática e aproximados a *synsets*, de acordo com um procedimento semelhante ao apresentado em (Gonçalo Oliveira e Gomes 2011a). O resultado é um *thesaurus*, onde cada conceito é representado por um conjunto de palavras sinónimas, à imagem de uma *wordnet*.

As figuras 1 e 2 apresentam dois exemplos de redes de palavras, formadas por relações de sinonímia extraídas automaticamente. Cada ligação entre duas palavras indica que foi extraída uma relação de sinonímia entre ambas, e fundos com diferentes tonalidades de cinzento identificam aglomerados. Na figura 1 existe uma ambiguidade relacionada com o substantivo 'histórico', que tanto se pode referir a uma cronologia, como a alguém que estuda história, um historiador. A ambiguidade acaba por ser desfeita e, devido aos parâmetros utilizados, 'histórico' fica apenas com o significado de cronologia.

Por outro lado, a figura 2 apresenta um exemplo mais complexo, numa rede maior, onde existem várias ambiguidades. No fim, é possível identificar três conceitos que, apesar de próximos, são diferentes: (A) alguém que muda de local em busca de melhor sorte; (B) alguém que foge do seu país por estar a ser perseguido; e ainda (C) alguém que foi expulso do seu país. Também na figura 2, verifica-se que algumas palavras mantêm alguma ambiguidade e pertencem a dois ou três aglomerados. Isto vai ao encontro

da realidade das linguagens naturais, onde a mesma palavra pode efectivamente ter mais de um significado.

3.3. Integração das relações

168

Por fim, procuram-se associar os argumentos das restantes relações extraídas, aos synsets descobertos anteriormente, de acordo com um dos métodos apresentados em (Gonçalo Oliveira e Gomes 2011b).

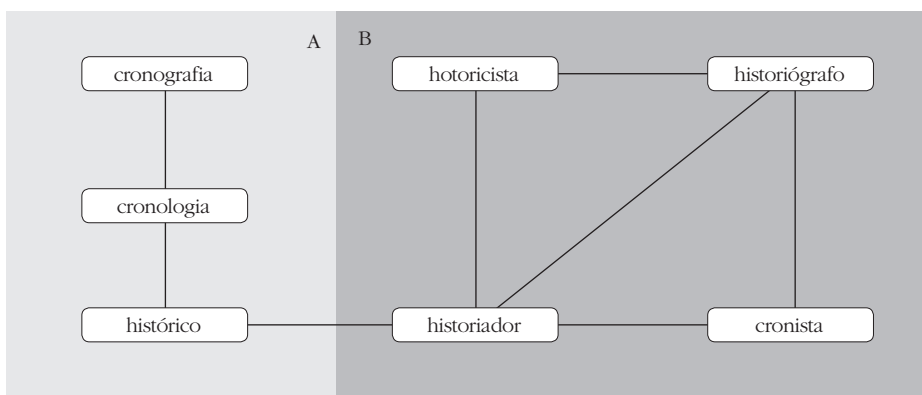


Figura 1: Rede de sinonímia em torno da palavra “histórico” e aglomerados identificados

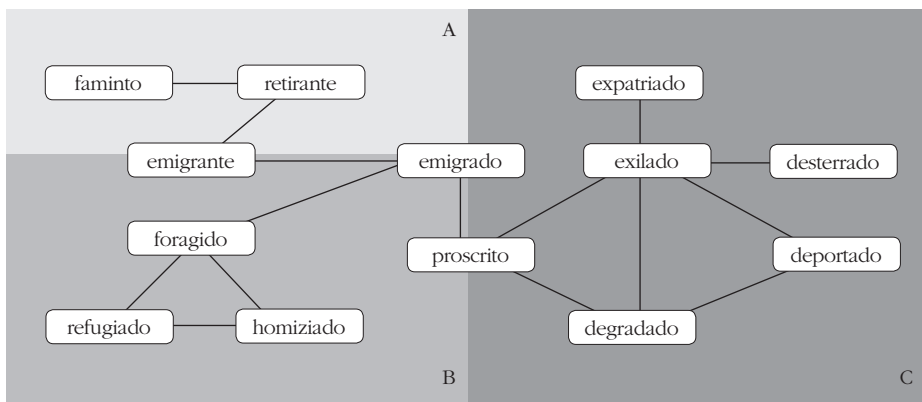


Figura 2: Rede de sinonímia em torno da palavra “emigrado” e aglomerados identificados

Por outras palavras, os triplos que relacionam palavras passam a relacionar conceitos, ainda que descritos por palavras. Um dos métodos que utilizamos na realização desta tarefa procura o par de *synsets* mais semelhante, dentro de todos os pares possíveis de *synsets* (S_a, S_b), em que S_a contém o termo a , e S_b contém o termo b . A semelhança é calculada com base nas vizinhanças dos termos dos *synsets* numa rede formada por todos os triplos extraídos. A tabela 1 apresenta alguns exemplos de relações entre termos incluídos no PAPEL e a sua correspondência após a integração num *thesaurus*, o TeP.

Triplos de termos	Triplos de synsets
<i>aparelho</i> Hiperónimo-de <i>televisor</i>	<i>apresto, utensílio, petrechos, instrumento, apetrechos, aparelho</i> Hiperónimo-de <i>tv, televisão, televisor, têvê</i>
<i>extensão</i> Hiperónimo-de <i>território</i>	<i>superfície, dimensão, extensão; espaço, área</i> Hiperónimo-de <i>território, área</i>
<i>ângulo</i> Parte-de <i>triângulo</i>	<i>ângulo, face, lado</i> Parte-de <i>triângulo, trilateral</i>
<i>técnica</i> Membro-de <i>marketing</i>	<i>arte, técnica</i> Membro-de <i>marketing</i>
<i>edição</i> Finalidade-de <i>programa</i>	<i>edição, lançamento</i> Finalidade-de <i>programa, aplicativo</i>

Tabela 1. Exemplos da integração de relações no thesaurus.

4. Resultados obtidos

Nesta secção são apresentados os resultados mais recentemente obtidos na criação do Onto.PT. Até ao momento, foram explorados três dicionários: o Dicionário da Língua Portuguesa (DLP, através do PAPEL 2.0); o Dicionário Aberto (DA, (Simões e Farinha 2010)), um dicionário público do português, com data de 1913, recentemente convertido para um formato digital; e também a versão portuguesa do Wikcionário. Foram ainda utilizados dois

thesauri, mais propriamente o TeP e o OpenThesaurus.PT. Na construção aqui referida, os *synsets* foram transformados em pares de sinonímia e, antes da descoberta de *synsets*, acrescentados às relações de sinonímia extraídas dos dicionários¹⁰⁶.

170 Iniciamos esta secção com uma análise ao vocabulário utilizado nas definições dos dicionários, crucial para a criação das gramáticas utilizadas na extracção de relações semânticas. Depois, apresentamos as relações extraídas, dados sobre os *synsets* descobertos acompanhados de exemplos e, por fim, dados sobre a versão actual do Onto.PT.

4.1. Análise de vocabulário

Antes de construir as gramáticas utilizadas na extracção de relações semânticas, é necessário analisar o vocabulário utilizado nas definições de dicionários para, assim, identificar mais facilmente os padrões que podem e devem ser utilizados na tarefa anteriormente referida. A tabela 2 mostra alguns dos padrões mais frequentes nas definições dos dicionários explorados, a sua quantidade em cada dicionário e ainda a relação semântica que normalmente transmitem.

Além de padrões como os da tabela 2, as gramáticas tiram ainda partido de outras duas construções, nomeadamente:

- Definições com apenas uma palavra, ou uma enumeração de palavras, são utilizadas na extracção de relações de sinonímia entre as palavras da definição e a palavra definida.
- Uma grande parte das definições de substantivos iniciam-se por um hiperónimo da palavra definida, eventualmente modificado por um adjectivo, a menos que essa palavra seja uma "cabeça vazia", ou seja, uma palavra sem conteúdo como, por exemplo, 'tipo' ou 'parte' (ver (Chodorow *et al.* 1985)).

¹⁰⁶ Além do procedimento aqui descrito, temos experimentado outras formas de integrar *synsets* existentes no Onto.PT. (Gonçalo Oliveira e Gomes 2011c).

Padrão	Categoria	Frequência			Relação semântica
		DLP	DA	Wikcionário	
<i>o mesmo que</i>	Substantivo	0	10.627	960	Sinonímia
<i>a[ç]to ou efeito de</i>	Substantivo	3.851	2.501	553	Causa
<i>aquele que</i>	Substantivo	1.148	3.357	476	Hiperonímia
<i>conjunto de</i>	Substantivo	1.004	316	293	Membro
<i>espécie de</i>	Substantivo	798	2.846	202	Hiperonímia
<i>género de</i>	Substantivo	29	4.148	47	Hiperonímia
<i>variedade de</i>	Substantivo	455	621	42	Hiperonímia
<i>[a] parte do/da</i>	Substantivo	445	433	96	Parte
<i>natural ou habitante de/do/da</i>	Substantivo	528	0	0	Local/Origem
<i>habitante ou natural de</i>	Substantivo	6	0	211	Local/Origem
<i>instrumento[,] para</i>	Substantivo	94	284	17	Finalidade
<i>... produzid[o/a] por/pelo/pela</i>	Substantivo	155	146	31	Produtor
<i>o mesmo que</i>	Verbo	0	166	84	Sinonímia
<i>Fazer</i>	Verbo	1.680	1.294	328	Causa
<i>Tornar</i>	Verbo	1.359	1.672	277	Causa
<i>Ter</i>	Verbo	467	519	154	Propriedade
<i>o mesmo que</i>	Adjectivo	0	2.685	179	Sinonímia
<i>relativo a/á/ao</i>	Adjectivo	1.236	5.554	946	Propriedade
<i>que se</i>	Adjectivo	1.602	1.599	443	Propriedade
<i>que tem</i>	Adjectivo	2.698	4.291	437	Parte/Propriedade
<i>diz-se de</i>	Adjectivo	2.066	738	272	Propriedade
<i>relativo ou pertencente</i>	Adjectivo	1.647	9	50	Membro/ Propriedade
<i>habitante ou natural de</i>	Adjectivo	0	0	143	Local/Origem
<i>de modo</i>	Advérbio	398	2.261	89	Maneira
<i>de maneira</i>	Advérbio	49	9	28	Maneira
<i>de forma</i>	Advérbio	30	3	18	Maneira

Tabela 2: Padrões frequentes e produtivos em definições de dicionários.

Ao realizarmos esta análise, verificamos que, com pequenas adaptações, as gramáticas utilizadas na construção do PAPEL 2.0¹⁰⁷ poderiam ser reutilizadas, já que grande parte dos padrões utilizados no DLP são também empregues nos outros dois dicionários.

¹⁰⁷ Disponíveis livremente a partir de <http://www.linguateca.pt/PAPEL/>

4.2. Relações extraídas

Ao reutilizarmos as gramáticas do PAPEL, no Onto.PT acabamos por extrair o mesmo tipo de relações presentes no primeiro recurso. Na tabela 3, apresentamos os vários tipos de relação extraídos, que separamos ainda em sub-relações, tendo em conta a categoria gramatical dos argumentos. Ainda na tabela 3, apresentamos as quantidades destas sub-relações extraídas e ainda exemplos para cada uma delas.

172

Relação	Args.	Quantidade				Exemplos
		PAPEL	DA	Wikc.	Total	
Sinonímia	n,n	37.452	24.327	13.569	64.131	<i>alegria,satisfação</i>
	v,v	21.465	11.031	4.076	30.380	<i>esticar,estender</i>
	adj,adj	19.073	10.114	6.640	29.676	<i>racional,filosófico</i>
	adv,adv	1.171	1.187	289	2.292	<i>imediatamente,já</i>
Hiperonímia	n,n	62.591	54.059	17.845	108.900	<i>sentimento,afecto</i>
Parte-de	n,n	2.805	1.280	718	4.485	<i>núcleo,átomo</i>
	n,adj	3.721	3.165	558	6.058	<i>vício,vicioso</i>
	adj,n	17	45	26	87	<i>sujeito,oração</i>
Membro-de	n,n	5.929	1.259	1.153	7.889	<i>aluno,escola</i>
	n,adj	34	25	23	77	<i>coisa,coletivo</i>
	adj,n	883	84	121	1.071	<i>rural,campo</i>
Causa-de	n,n	1.013	172	306	1.440	<i>vírus,doença</i>
	n,adj	17	7	4	23	<i>paixão,passional</i>
	adj,n	498	142	163	735	<i>horível,horror</i>
	n,v	39	17	5	59	<i>fogo,fundir</i>
Produtor-de	v,n	6.399	5.487	1.139	8.812	<i>mover,movimento</i>
	n,n	898	571	310	1.652	<i>oliveira,azeitona</i>
	n,adj	35	25	8	65	<i>fermentação,fermentado</i>
Finalidade-de	adj,n	359	230	37	520	<i>fonador,som</i>
	n,n	2.886	2.302	1.349	6.194	<i>sustentação,mastro</i>
	n,adj	63	48	9	111	<i>habitação,babitável</i>
	v,n	5.192	2.186	1.479	8.314	<i>calcular,cálculo</i>
Local-de	v,adj	260	197	9	403	<i>comprimir,compressivo</i>
	n,n	849	391	728	1.858	<i>Galiza,galego</i>
Maneira-de	adv,n	1.113	1.535	156	2.482	<i>ociosamente,indolência</i>
	adv,adj	N/A	1.594	112	1.664	<i>virtualmente,virtual</i>
Maneira	adv,n	117	148	24	260	<i>prontamente,demora</i>
sem	adv,v	11	4	5	19	<i>seguido,parar</i>
Propriedade-de	adj,n	6.518	3.460	1.636	10.226	<i>daltónico,daltonismo</i>
	adj,v	17.543	9.701	3.239	27.916	<i>geoquímico,geoquímica</i>

Tabela 3: Relações semânticas extraídas

4.3. Synsets descobertos

Ao observar as redes estabelecidas não só pelas relações de sinonímia acima descritas, mas também por pares de sinonímia obtidos a partir dos *thesauri* TeP e OpenThesaurus.PT, verificou-se que estas redes eram constituídas por uma grande sub-rede de palavras ligadas e por várias pequenas sub-redes. A tabela 4 apresenta alguns números relativamente ao tamanho das sub-redes para cada categoria gramatical, nomeadamente, o número total de palavras da categoria, o total de sub-redes, o tamanho da maior e segunda maior sub-redes e ainda o número de sub-redes com apenas duas palavras.

173

Olhando para estes números, verificamos que, se não existisse ambiguidade, todas as palavras de cada sub-rede seriam sinónimas entre si, porque é possível encontrar um caminho de relações de sinonímia entre qualquer par de palavras na mesma sub-rede. No entanto, é necessário descobrir *synsets* dentro das sub-redes e assim desfazer ou tratar algumas ambiguidades.

A descoberta automática de *synsets* resulta num *thesaurus* cujas propriedades se apresentam na tabela 5, para cada categoria gramatical. As propriedades incluem, para as palavras, o número total de palavras, palavras com mais de um sentido (Ambigs.), o número médio de sentidos por palavra (med(sents)) e o número de sentidos da palavra mais ambígua (max(sents)). No que diz respeito aos *synsets*, é apresentado o número total, o número médio de palavras por *synset* (med(pals)), o número de *synsets* com duas palavras (*pals*=2), com mais de 25 palavras (*pals*>25) e ainda o número de palavras no maior *synset* (max(pals)).

Categoria	Palavras	Sub-redes			
		Total	Tamanho maior	Tamanho 2ª maior	Apenas 2 palavras
n	45.214	5.597	30.922	48	4.016
v	11.975	345	11.186	6	277
adj	21.840	2.928	14.587	28	2.155
adv	2.504	215	1.894	10	135

Tabela 4: Dados sobre as sub-redes de sinonímia extraídas.

Tabela 5: Propriedades do thesaurus

POS	Palavras				Synsets				
	Total	Ambigs	med (sents)	max (sents)	Total	med (pals)	pals=2	pals>25	max (pals)
n	45.213	17.717	1,75	8	14.057	5,63	4.178	48	46
v	11,904	6.612	1,96	8	3.549	6,58	571	46	55
adj	21.837	8.695	1,73	8	6.550	5,78	2.303	61	48
adv	2.504	876	1,61	7	651	6,19	135	1	27

De forma a verificar como o *thesaurus* gerado automaticamente lida com a polissemia, procuramos por diferentes *synsets* de palavras claramente polissêmicas, que tentamos separar pelos diferentes sentidos que essas palavras transmitem. Embora nem todos os sentidos estivessem representados para todas as palavras e, por vezes, houvesse *synsets* que misturavam mais do que um sentido, há bons exemplos, como os representados na tabela 6. Aí, temos a palavra: ‘etiqueta’, que pode ter o sentido de uma legenda/rótulo ou de um protocolo que deve ser seguido, e ‘pastel’, com dois dos seus possíveis sentidos figurados, nomeadamente uma pessoa indolente e também dinheiro.

Palavra	Sentido	Synsets
<i>etiqueta</i>	legenda	<i>etiqueta, signal, letreiro, rótulo, dístico, tabuleta, matrícula, tãtuleiro, cartel, inscrição, epítáfio, legenda</i>
	protocolo	<i>etiqueta, praxe, práxis, protocolo, cerimónia, pragmática, cerimónia, formalidade, requijife, costumeira, usança, premática</i>
<i>pastel</i>	indolente	<i>pastel, zorrão, marralhão, indolente, mandriona, boleima</i>
	dinheiro	<i>pastel, guínes, moeda, jimbo, bagalboça, cobre, boro, dieiro, baguines, parolo, marcaureles, cacau, pataco, matambira, massaroca, gimbo, metal, cunques, bagalbo, níquel, fanfa, bilbestres, pecúnia, jan-da-cruz, cum-quibus, mussuruco, pilim, dinheiro, zerzulbo, numo, chelpa, calique, teca, pecunia, patacaria, carcanhol, pecuniária</i>

Tabela 6: Synsets de palavras polissêmicas.

O *synset* relativo a dinheiro é um dos maiores do *thesaurus* obtido e, de acordo com os parâmetros utilizados, poderá ainda ser maior, como é o caso do *synset* equivalente apresentado em (Gonçalo Oliveira e Gomes

2011a). Neste *synset* encontramos várias formas de nos referirmos a dinheiro, como por exemplo: formas coloquiais (e.g. pastel, carcanhol, pilim), formas populares (e.g. massaroca, cacau), formas da variante moçambicana do português (e.g. mussuruco, matambira), formas figurativas (metal), gírias (e.g. baguines, gimbo) ou formas mais antigas (dieiro).

Além deste *synset*, apresentamos de seguida, a título de curiosidade, os maiores *synsets* obtidos para substantivos e para verbos, juntamente com o significado que estes transmitem. Observa-se que os maiores *synsets* de verbos têm mais palavras que os *synsets* de substantivos.

175

Substantivos:

confusão (46 palavras): *baixaria, furdúncio, matalotagem, fuzuê, confusão, complicação, feijoada, mistura, encrenca, remexida, amalgamação, salada_russa, saricoté, brigas, estrilbo, escangalho, abstrusão, trapalbada, siricutico, vira-teimão, cambulha, cu-de-boi, valverde, kanvuanza, javardice, embrolho, canvanza, indistincção, estricote, caravançarai, imbróglío, desmanho, ensalsada, enredia, vuvu, mexedura, cancaburra, ula, fula-fula, timaca, misturada, pastelada, lelê, abstrusidade, assarapantamento, encrequilha*

embriaguez (38 palavras): *pisorga, porre, berzundela, zurca, zuca, trapisonda, torta, piela, bebedeira, perua, perunca, bicancra, gateira, taçada, carraspana, samatra, pizorga, tachada, ganso, tortelia, caroça, turca, pifão, borracheira, carrega, pifo, zerenamora, zola, marta, parrascana, gardinhola, tropecina, bezana, ema, cardina, lavanco, tiorga, bêbeda*

prostituta (38 palavras): *pécora, galinha-polaca, meretriz, faniqueira, barregã, mulher-dama, mulber-do-fandango, mulber-de-mã-nota, dadeira, mulber-perdida, zoupeira, arruadeira, hervoeira, messalina, mundana, prostituta, michê, tronga, desproveito, prostituída, culatrão, perca, zoína, puta, rascoa, marafona, bagaxa, dama-da-noite, fuampa, lúrpia, michela, mulber-da-vida, perdida, biraia, rameira, samarrão, pelejo, cróia*

Verbos:

176

ridicularizar (55 palavras): *embromar, ridicularizar, motejar, trotar, gozar, palbetar, chasquear, malbar, ridiculizar, sotaquear, toirear, chacotear, ironizar, pantear, cachetar, gingar, facetear, gracejar, galbofear, palbetear, desfrutar, debicar, chalacear, mexer, desfruir, zombetear, debochar, achincalbar, pagodear, derriçar, trotear, apepinar, tourear, vasconcear, chalaçar, chocarrear, sorrir, bexigar, mofar, empulbar, cafangar, apodar, troçar, brincar, caçoar, pilberiar, satirizar, rir, bigodear, zingrar, mangar, galbofar, escarnecer, escarnir, zombar*

guitar (53 palavras): *toar, atroar, urrar, exclamar, vozeirar, rebramar, estrilar, troar, tonitruar, estrondar, esbravejar, gritar, trovejar, ulular, uivar, esbravear, bradejar, tempestear, retroar, explodir, bramir, mugir, deblaterar, estrugir, trovoar, vozear, tronar, expluir, esgoelar, ressonar, bravejar, rimbombar, barulhar, conclamar, tempestuar, goelar, barregar, fragorar, ribombar, ressoar, estourar, resoar, vociferar, berrar, retumbar, estrondear, rebombar, bradar, esbravecer, latir, estoirar, berregar, bramar*

enfeitar (50 palavras): *colorear, incrementar, adernar, emoldurar, alfaiar, floretear, formosear, paramentar, assazonar, decorar, tecer, ataviar, aparelhar, enfeitar, enflorear, ornar, sobredoirar, sazonar, aformosear, recamar, amoldurar, aparamentar, brilbantar, sobredourar, embrincar, assazoar, ornamentar, jaezar, emoldar, enflorar, colorar, colorizar, moldurar, aparatar, embelezar, adereçar, alindar, embonecar, engalantar, espenicar, aformosar, honestar, florear, abonecar, adornar, ajaezar, formosentar, galantear, exornar, aformosentar*

O recurso final

Após a integração das relações do PAPEL e das relações extraídas do DA e Wikcionário, obtivemos uma ontologia lexical cujas quantidades de relações, agora entre *synsets*, se apresentam na tabela 7. Apesar de, nesta tabela, não terem sido contabilizadas relações inversas, é sempre possível inferir este tipo de relações, ou seja, quando temos uma relação A hiperónimo-de B, é sempre possível inferir que B hipónimo-de A.

Quando se pretende integrar uma palavra que o *thesaurus* não contém, é necessário criar um novo *synset* apenas com essa palavra. Assim, na tabela 7, optamos por separar as relações de acordo com aquelas que se estabelecem entre dois *synsets* com apenas uma palavra ($1 \rightarrow 1$), *synset* com uma palavra e *synset* com várias palavras ($1 \rightarrow n$, $n \rightarrow 1$) e, ainda, dois *synsets* com mais de uma palavra ($n \rightarrow n$). Verifica-se que a maior parte das relações liga um *synset* com uma palavra a um *synset* com mais palavras.

177

Relação	Args	Instâncias			
		$1 \rightarrow n$	$1 \rightarrow n ; n \rightarrow 1$	$n \rightarrow n$	Total
Hiperónimo-de	n,n	2.789	45.637	40.161	88.584
Parte-de	n,n	654	1.910	1.654	4.218
	n,adj	663	2.614	2.017	5.294
	adj,n	23	41	15	79
Membro-de	n,n	776	3.244	2.326	6.346
	n,adj	12	43	13	68
	adj,n	265	366	182	813
Causador-de	n,n	201	580	573	1.354
	n,adj	2	7	12	21
	adj,n	102	221	305	628
	n,v	2	11	41	54
	v,n	388	1.920	4.338	6.646
Produtor-de	n,n	282	627	574	1.473
	n,adj	12	28	24	64
	adj,n	30	224	201	455
Finalidade-de	n,n	620	2.498	2.745	5.863
	n,adj	18	50	33	101
	v,n	1.531	3.617	2.720	7.868
	v,adj	27	183	156	366
Local-origem	n,n	674	520	151	1.345
Maneira-de	adv,n	135	1.073	939	2.147
	adv,adj	47	732	619	1.398
Maneira	adv,n	1	95	137	233
sem	adv,v	1	10	7	18
Propriedade-de	adj,n	1.994	4.181	2.564	8.739
	adj,v	7.761	15.184	2.778	25.723

Tabela 7: Relações entre *synsets* na versão actual do Onto.PT

Por fim, para dar uma ideia da estrutura e do que é possível encontrar no Onto.PT, mostramos, na figura 3, os resultados da pesquisa pela palavra 'hospital', na versão actual do nosso recurso.

178

- S: (adj) **hospital**, bem-querente, bem-intencionado, caridoso, benevolente, benévolo
- S: (adj) **hospital**, caridoso, esmoler, benfazejo, bem-fazejo, bem-fazente
- S: (s) **hospital**, espiritual, nosocômio
 - o temParte
 - S: (s) enfermaria
 - parteDe
 - hiperonimoDe
 - S: (s) ambulatório
 - hiponimoDe
 - meioPara
 - S: (s) paciente, enfermeiro, doente
 - o referidoPorAlgoComPropriedade
 - S: (adj) hospitalar, hospitalário
 - S: (adj) nosocômico, nosocomial
 - o hiperonimoDe
 - S: (s) gafaria, leprosaría
 - S: (s) leprocómio
 - S: (s) lazareto
 - S: (s) hospital psiquiátrico, manicómio, hospício, rilhafales, casa-de-orates, manicómio
 - o hiponimoDe
 - S: (s) edifício, edifícamento, edificação
 - S: (s) instituto, instituição, fundação, instauração, infra-estrutura, estabeleza, implantação, implante, estabelecimento, inauguração
 - S: (s) telhado, mobiliário, teito, tecto, habitação, dinastia, meisom, casa, cosque, mesão

Figura 3: Relações e synsets a partir da entrada 'hospital'.

5. Notas conclusivas

Neste artigo foi apresentada a abordagem adoptada na construção automática de uma ontologia lexical para o português, o Onto.PT, juntamente com um conjunto de exemplos que ilustram tanto as fases de construção como os resultados mais recentemente obtidos.

Um ponto forte deste trabalho é a dimensão dos recursos gerados, onde, ao agregar um conjunto de recursos livres para o português, se inclui a maior rede lexical e também o maior *thesaurus* electrónico para a nossa língua. Além disso, dada a flexibilidade da abordagem seguida, será possível aumentar ainda mais o Onto.PT, tirando partido de outros recursos textuais.

Por outro lado, há ainda um longo caminho a percorrer no que diz respeito à qualidade do recurso e à organização do conhecimento nele

contido. Acreditamos que, brevemente, o Onto.PT seja uma alternativa livre aos recursos lexicais actualmente existentes para o português, e que possa contribuir para um maior desenvolvimento de ferramentas e aplicações dentro do processamento computacional da nossa língua.

Referências

179

- Agirre, Eneko, Oier Lopez De Lacalle, e Aitor Soroa (2009). Knowledge-based WSD on specific domains: performing better than generic supervised WSD. *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)*. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1501-1506.
- Amsler, Robert A. (1981). A taxonomy for English nouns and verbs. *Proceedings of the 19th Annual Meeting of The Association for Computational Linguistics*. Morristown, NJ: ACL Press, 133-138.
- Chodorow, Martin S., Roy J. Byrd e George E. Heidorn (1985). Extracting semantic hierarchies from a large on-line dictionary. *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*. Morristown, NJ: ACL Press, 299-304.
- De Melo, Gerard e Gerhard Weikum (2008). On the Utility of Automatically Generated Wordnets. *Proceedings of the 4th Global WordNet Conf. (GWC)*. Szeged, Hungary: University of Szeged, 147-161.
- Dias-da-Silva, Bento Carlos, Mirna F. de Oliveira, e Helio R. de Moraes (2002). Groundwork for the Development of the Brazilian Portuguese WordNet. *Proceedings of Advances in Natural Language Processing, 3rd International Conference (PorTAL)*, Faro, Portugal. LNCS 2389, Springer. 189-196.
- DLP (2005). *Dicionário PRO da Língua Portuguesa*. Porto: Porto Editora.
- Fellbaum, Christiane (ed.) (1998). *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. Cambridge, Massachusetts: The MIT Press.
- Gonçalo Oliveira, Hugo e Paulo Gomes (2011a). Automatic Discovery of Fuzzy Synsets from Dictionary Definitions. *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*. Barcelona, Spain: AAAI Press, 1801-1806.
- Gonçalo Oliveira, Hugo e Paulo Gomes (2011b). Ontologising Relational Triples into a Portuguese Thesaurus. *Local Proceedings of the 15th Portuguese Conference on Artificial Intelligence (EPIA 2011)*. Lisbon, Portugal: APPIA.
- Gonçalo Oliveira, Hugo e Paulo Gomes (2011c.) Automatically enriching a Thesaurus with information from Dictionaries. *Progress in Artificial Intelligence. Proceedings of the 15th Portuguese Conference on Artificial Intelligence (EPIA 2011)*, LNCS 7026. Springer, 462-475.
- Gonçalo Oliveira, Hugo e Paulo Gomes (2010). Onto.PT: Automatic Construction of a Lexical Ontology for Portuguese. *Proceedings of the 5th European Starting AI Researcher Symposium (STAIRS 2010)*, Lisbon, Portugal. IOS Press, 199-211.
- Gonçalo Oliveira, Hugo, Diana Santos e Paulo Gomes (2010). Extração de relações semânticas entre palavras a partir de um dicionário: o PAPEL e sua avaliação. *Linguamática* 2:1, 77-93.
- Hearst, Marti A. (1998). Automated Discovery of WordNet Relations. In Fellbaum, Christiane (ed.) (1998). *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. Cambridge, Massachusetts: The MIT Press, 131-151.

- Hirst, Graeme (2004). Ontology and the Lexicon. In: S. Staab & R. Studer (eds.) *Handbook on Ontologies*. Springer, 209-230.
- Ide, Nancy e Jean Veronis (1995). Knowledge Extraction from Machine-Readable Dictionaries: An Evaluation. In: *Machine Translation and the Lexicon, LNAI*. Springer, 19-34.
- Lin, Dekang e Patrick Pantel (2002). Concept Discovery from Text. *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, 577-583.
- Marrafa, Palmira (2002). Portuguese WordNet: general architecture and internal semantic relations. *DELTA* 18: 131-146.
- Maziero, Erick G., Thiago A. S. Pardo, Ariani Di Felippo, e Bento C. Dias-da-Silva. (2008). A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Eletrônico para o Português do Brasil. *Proceedings of the VI Workshop em Tecnologia da Informação e da Linguagem Humana (TIL)*, Vila Velha, Espírito Santo, 390-392.
- Medelyan, Olena, David Milne, Catherine Legg, e Ian H. Witten (2009). Mining Meaning from Wikipedia. *International Journal of Human-Computer Studies* 67:9, 716-754.
- Nichols Eric, Francis Bond e Dan Flickinger (2005). Robust Ontology Acquisition from Machine-Readable Dictionaries. *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)*, Edinburgh, Scotland, UK. Professional Book Center, 1111-1116.
- Pasca, Marius e Sanda M. Harabagiu (2001). The Informative Role of WordNet in Open-Domain Question Answering. *Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations*. Pittsburgh, USA, 138-143.
- Pianta, Emanuele, Luisa Bentivogli, e Christian Girardi (2002). Multiwordnet: developing an aligned multilingual database. *1st International Conference on Global WordNet*. Mysore, India, 2002.
- Prévot Chu-Ren Huang Laurent, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, e Alessandro Oltramari (2010). Ontology and the Lexicon: a multi-disciplinary perspective (introduction). In: Chu-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, e Laurent Prevot (eds.). *Ontology and the Lexicon: A Natural Language Processing Perspective*. Cambridge: Cambridge University Press, 3-24.
- Resnik, Philip (1995). Disambiguating Noun Groupings with Respect to WordNet Senses. *Proceedings of the 3rd Workshop on Very Large Corpora*. Cambridge, MA: The MIT Press, 54-68.
- Richardson, Stephen D., William B. Dolan e Lucy Vanderwende (1998). MindNet: Acquiring and Structuring Semantic Information from Text. *Proceedings of the 17th International Conference on Computational Linguistics (COLING)*. ACL, Stroudsburg, PA, USA. 1098-1102.
- Santos, Diana, Anabela Barreiro, Cláudia Freitas, Hugo Gonçalo Oliveira, José Carlos Medeiros, Luís Costa, Paulo Gomes, e Rosário Silva (2010). Relações semânticas em português: comparando o TeP, o MWN.PT, o Port4NooJ e o PAPEL. In: A. M. Brito, F. Silva, J. Veloso, e A. Fiéis (eds.). *Textos seleccionados. XXV Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: Colibri/APL, 681-700.
- Simões Alberto e Rita Farinha (2011). Dicionário Aberto: Um novo recurso para PLN. *Vice-Versa* 16, 159-171.
- Teixeira, Jorge, Luís Sarmento e Eugénio Oliveira (2010). Comparing Verb Synonym Resources for Portuguese. *Proceedings of the 9th International Conference (PROPOR 2010)*, LNAI 6001. Springer, 100-109.
- Vossen, Piek (1997). EuroWordNet: a multilingual database for information retrieval. *Proceedings of the DELOS workshop on Cross-Language Information Retrieval*. European Research Consortium For Informatics and Mathematics. Zurich, 85-94.

Sara Candeias¹, Fernando Perdigão^{1,2}

¹ Instituto de Telecomunicações, Polo de Coimbra, DEE-UC

² Departamento de Engenharia Eletrotécnica e de Computadores da UC

INTEGRAÇÃO LINGUÍSTICA EM SISTEMAS DE CONVERSÃO DE GRAFEMA PARA FONE(MA)

1. Introdução

Sendo uma das áreas de estudo da linguística teórica, a fonética-fonologia tem como objetivo de estudo a descrição e descodificação da fala de uma língua. O estudo da fala de uma língua implica obrigatoriamente conhecer o conjunto de elementos sonoros que a formam, assim como as possibilidades de combinação desses elementos. É do setor fonético o conhecimento de que, para falar, fazemos uso de consoantes e de vogais; de que as vogais podem ser orais, nasais ou até nasaladas; de que, em português europeu (PE), a lateral <l> a fechar sílaba é pronunciada como velarizada (como em <mel>); de que as vogais posteriores fechadas são reduzidas em posição final de palavra antecidas de consoante bilabial (como em <rumo> ou <campo>); ou de que um simples <e> em contexto inicial de vocábulo e de sílaba fechada por <s> (como em <esperança>) pode ser pronunciado como [ə], como [i], ou até nem ser pronunciado ([∅]) (cf., a propósito desta multirrealização, os estudos de (Andrade 1994) e de (Candeias *et al.* 2011)). É do setor fonológico distinguir as características dos sons que diferenciam significados numa língua, sendo por isso possível afirmar, a título de exemplo, que as vibrantes múltipla e simples são, em PE, entidades fonológicas distintas em contexto intervocálico, uma vez que opõem significados (como é o que acontece no binómio <carro> vs. <caro>).

Ora, uma vez que o módulo de conversão de Grafema em Fone(ma) tem sido, à parte algumas flutuações de estruturas, um bloco essencial da arquitetura geral de sistemas de tecnologias da fala, e que, para o seu

desenvolvimento, aflui naturalmente o conhecimento linguístico, tornou-se premente a necessidade de se estreitar a relação entre a ciência da língua e a engenharia do processamento linguístico. Assim, partir do conhecimento linguístico em direção ao seu processamento e integração em sistemas automáticos quer de reconhecimento (ASR, do inglês *Automatic Speech Recognition*) quer de síntese de voz (TTS, do inglês *Text-to-Speech*)¹⁰⁸ trouxe a necessidade de se estender o estudo da língua a um novo paradigma: o da engenharia do processamento linguístico ou da linguística tecnológica.

Encontramo-nos, obviamente, perante uma área complexa onde saberes clássicos entram em confluência e que tem crescido a par da revolução tecnológica a que temos vindo a assistir. Vivemos numa atualidade que não concebe um quotidiano sem acesso à internet ou sem o recurso ao uso de computador ou de telemóvel. Cada vez com mais frequência, somos atendidos por uma voz artificial que nos dá instruções quando telefonamos para um serviço público ou comandamos com a nossa própria voz aplicações no telemóvel ou no computador. Tem sido a consciência de que as tecnologias linguísticas se destinam a facilitar o diálogo entre Homem-Máquina que tem impulsionado o desenvolvimento nesta área, ampliando soluções tecnológicas com interface de voz adaptadas à língua portuguesa, falada em todo o mundo por 240 827 milhares de pessoas¹⁰⁹.

Este artigo vem no sentido de descrever como o conhecimento fornecido pela linguística teórica tem sido (e pode ser) processado e integrado em sistemas automáticos de processamento do português, nomeadamente no módulo de conversão de Grafema-Fone(ma). Em simultâneo, o problema da conversão é caracterizado e, na secção 2, algumas questões relacionadas com opções terminológicas e formais no que diz respeito à tarefa da conversão são levantadas. A secção 3 descreve as abordagens que têm sido adotadas no desenvolvimento de conversores, incluindo as que resultam de

¹⁰⁸ Um sistema de síntese de fala (TTS) é um modelo computacional que permite gerar fala automaticamente a partir de uma entrada de texto. Ao invés, um sistema de reconhecimento de fala (ASR) identifica as unidades dessa língua a partir de um sinal de voz, construindo estruturas vocabulares.

¹⁰⁹ Número obtido no sítio do Observatório da Língua Portuguesa, <http://observatorio-lp.sapo.pt/pt/dados-estatisticos/falantes-de-portugues>.

opções específicas tomadas pelos autores. Conclusões e desafios para trabalhos futuros são apresentados na última secção, a secção 4.

2. Conversão Grafema-Fone(ma)

O módulo denominado Grafema-Fone(ma) ou G2P (do inglês *Grapheme-to-Phone(me)*¹¹⁰) tem por objetivo interpretar a grafia de uma dada palavra¹¹¹ e convertê-la numa sequência de símbolos sonoros, representativos de como essa palavra é pronunciada. Ainda que com alguma flutuação no que diz respeito às arquiteturas dos sistemas quer de ASR quer de TTS, o processo de conversão tem sido, regra geral, um bloco integrante desses sistemas, fazendo parte de uma tarefa mais ampla, vulgarmente conhecida como pré-processamento de texto (ou *front-end*), e à qual cabe o processo de decodificação e de normalização de elementos presentes nos textos (tais como são os numerais, as siglas, os estrangeirismos, as abreviaturas, os acrónimos e as siglas, entre outros), para além do fracionamento do texto em frases e em vocábulos.

183

Antes de se passar à descrição das abordagens existentes no processo de conversão do grafema na forma simbólica representativa da sua pronúncia, algumas questões devem ser levantadas, as quais se relacionam com as opções terminológicas e formais adotadas. Por merecerem uma ponderação mais cuidada, estas questões são, de seguida, exploradas.

2.1. Fonema ou Fone?

Como é do conhecimento geral no âmbito da linguística teórica, e retomando conceitos já abordados na introdução deste texto, a fonologia de uma língua é a fonética interpretada dessa língua, firmando-se o fonema como uma unidade distintiva que se reconhece por uma relação de contraste.

¹¹⁰ Por vezes, o conversor é também designado por L2S, do inglês *Letter to Sound*.

¹¹¹ Por palavra entende-se, aqui, todo o átomo do léxico português, contendo, pelo menos, um grafema ou dígito.

Dito de outro modo, o fonema de uma língua, sem valor distintivo próprio, é possuidor de um valor distintivo capital dentro da língua. O fonema é assim admitido como a contrapartida linguística, formal e abstrata, da realidade sonora e, podendo ser realizado por um fone ou mais (alofones), é muitas vezes interpretado como uma classe de alofones que se revelam num campo de distribuição específico (sobre a matéria, cf., entre outros, (Barbosa 1994) e (Mateus e Andrade 2000)) e a introdução teórica expressa em (Candeias 2007).

A clarificação terminológica dos termos fone e fonema não tem sido regular nos estudos que têm vindo a ser efetuados no âmbito do G2P, sendo frequente encontrar uma alternância não justificada entre eles (cf. (Caseiro *et al.* 2002) e (Barros e Weiss 2006), como exemplo). Admitindo as razões teóricas ligadas ao binómio fone/fonema, sempre que o resultado do conversor dá conta de fenómenos de pronúncias alternativas, é porque exprime o nível da superfície, atuando no âmbito da fonética (cf. conversor descrito em (Oliveira *et al.* 1992)). Quando o resultado do procedimento de conversão adotado admite valências do contexto mais ou menos alargado no âmbito do vocábulo, opera-se a um nível de maior abstração, atuando no âmbito da fonologia. Nestes casos, mais comuns e geralmente centralizados na pronúncia estandardizada do PE, a unidade para a qual o grafema é convertido é considerada como uma escolha significativa de entre todas as outras unidades que o sistema da língua coloca ao dispor. Estudos como o de (Veiga *et al.* 2011) atendem a essa clareza terminológica, apresentado sistemas que convertem o grafema numa entidade mais abstrata, o fonema. Atuando a este nível, abre-se a possibilidade de poder vir a ser integrada no sistema a capacidade de cada fonema ser realizado por um feixe de alofones, representativos de ambientes acústicos decorrentes de efeitos de coarticulação, de *iod* ou de *sandbi*, e de casos de multipronúnciação. Um caso comum de alofonia em PE é a possibilidade de pronúnciação em [e]¹¹², em [ei]¹¹³ ou em [ɔ], quando a vogal <e> se localiza em posição tónica, seguida de consoante palatal <j> (como em <igreja>) ou <lh> (como em <abelha>). A multirrealização do *schwa* é

¹¹² Notação em SAMPA. Sobre o assunto de notação, cf. secção seguinte.

¹¹³ É opção dos autores não notar glides. Sobre o assunto, cf. 2.3.

outro caso de alofonia a demarcar, em especial porque potencia significativas taxas de erro no reconhecimento de fones (Candeias *et al.* 2011). Ambos os exemplos de alofonia mencionados, evidenciando a coexistência de mais do que um sistema linguístico, tornar-se-iam obrigatoriamente lacunares em termos de representatividade da realidade da língua portuguesa, caso observassem apenas o nível de superfície. Uma perspectiva de representação fonética, nestes casos, anularia a possibilidade da existência de mais do que uma pronúncia. Por outro lado, a multirrealização possível do fonema dependente do contexto é um parâmetro que pode ser automaticamente extraído usando modelos acústicos dessas unidades (Veiga *et al.* 2010).

185

2.2. Alfabeto Fonético

Outra questão subjacente à conversão G2P diz respeito à seleção do sistema de notação, seja ela fonética ou fonológica. Neste ponto, tendo-se já esclarecido o campo de ação relacionado quer com a fonética quer com a fonologia, não é pertinente aqui continuar essa distinção, pelo que o que se afirmar em relação ao alfabeto fonético pode igualmente afirmar-se relativamente ao alfabeto fonológico. No mesmo sentido, a ‘matéria sonora’ compreende metaforicamente segmentos fonéticos e fonológicos.

Ainda que o português seja uma língua cuja ortografia tem uma base fonética-fonológica (Mateus 2006), ocorrências existem em que a correspondência entre grafia e respetiva representação sonora não é invariavelmente unívoca, devida a fatores diversos, muitas vezes advenientes do contexto grafemático (caso dos grafemas <r> e <u> em <português>, <rosa> ou <arguido>) ou dependentes do estatuto morfológico ou sintático (cf. 2.4). São várias as situações nas quais essa correspondência não é unívoca: situações em que a um grafema pode corresponder quer a sons distintos, quer a mais do que um som (Tabelas 1 e 2) ou em que a mais do que um grafema corresponde apenas um som (caso dos dígrafos, por exemplo - Tabela 3). Com vista a uma implementação computacional, todos estes casos têm que ser tratados, seja por forma de regras, seja por anotadores estatísticos (cf. secção 3), numa tentativa de tornar a relação entre o grafema e o som inequívoca.

Grafemas	Fonemas	Exemplos
	ó	cama → /kóma/
a	a	mala → /ma/
	ó~*	canto → /kó-tu/
á		pás → /paS/
à	a	às → /aS/
ã	ó~	anão → /nó-u~/
â	ó~*	atlântico → /dó-tiku/
	ó	câmara → /kómó/
e	ó	desenho → /d@só/
	e	aquele → /kél@/
	@	refere → /R@fEr@/
	E	pele → /pEl@/
	i	areal → /órial/
	e~*	vento → /ve-tu/
	i~*	visões → /vizo-iS/
	ói*	extra → /óiStró/
	ó~*	votem → /votó-i~/
é	E	café → /kófE/
	ó-i-	contém → /kóntó-i~/
ê	ó	amêijoa → /ómóiZuó/
	e	você → /vóse/
	e~	pêndulo → /pe~dulu/
	ói	êxito → /óizitu/
	ó~i-ó~i~*	têm → /tó-i-ó~i~/
i, í	i	cima → /simó/, saí → /só/
	i~	cinco → /si~ku/, límpida → /li~pidó/
o	o	corpo → /kópu/
	O	copo → /kOpu/
	o~	conto → /ko-tu/
	u~	visão → /vizó-u~/
	u	porque → /pórk@/
ó	O	cópia → /kOpíó/
ô	o	avô → /vó/
	o~	cônsul → /ko~sul/
õ	o~	põe → /pó-i~/
	o~i~	põem → /pó-i-ó~i~/
u, ú	u	apura → /ópú/
	u~	túnel → /tunEl/
	∅	unto → /u~tu/, anúncio → /ónu~siu/
	∅	quente → /k t@/

Tabela 1 – Possíveis sons (fonemas) associados a cada grafema – caso de vogais. Atendendo a uma relação linear de ‘um para um’ entre grafemas e fonemas, situações em que o fonema é determinado pelo contexto estão indicadas por *. ∅ representa zero, isto é, uma não correspondência fonológica. A coluna à direita mostra exemplos de notação fonológica.

Grafemas	Fonemas	Exemplos
c	k s ks Ø	capa → /kap6/ cedo → /sedu/ ficcional → /fiksiunal actuar → /6tuar/ (<i>pré-AO</i>)
g	g Z	gatu → /gatu/ girafa → /Ziraf6/
l	l L	lado → /ladu/ malha → /maL6/
m	m u~ [*] i~ [*] Ø	mal → /mal/ apelam → /6pEl6~u~/ votem → /vOt6~i~/ campo → /k6~pu/
n	n i~ [*] J~ [*] Ø	ana → /6n6/ tens → /t6~i~S/ manha → /m6J6/ canto → /k6~tu/
p	p Ø	par → /par/, apto → /aptu/ óptica → /Otik6/ (<i>pré-AO</i>)
r	r R	caro → /karu/ carro → /kaRu/
s	s S z Z	massa → /mas6/ às → /aS/ casa → /kaz6/ abismo → /6biZmu/
x	s S z Z ks	máximo → /masimu/ xadrez → /S6dreS/ exame → /ez6m@/ ex-bar → /6iZbar/ fixo → /fiksu/
z	S z Z	arroz → /6RoS/ azar → /6zar/ felizmente → /f@liZme~t@/

Tabela 2 – Continuação da Tabela 1 – caso de consoantes.
A expressão (*pré-AO*) diz respeito à grafia adotada previamente
ao estabelecimento do Acordo Ortográfico de 1990.

Relativamente ao modo de representação da matéria sonora, no módulo do conversor, o alfabeto SAMPA (*Speech Assessment Methods Phonetic Alphabet*) e a sua extensão X-SAMPA (*The Extended Speech Assessment Methods Phonetic Alphabet*) têm sido usados, com mais frequência, em detrimento do Alfabético Fonético Internacional (AFI ou IPA – *International Phonetic Alphabet*). Esta opção assenta numa atitude de otimização dos recursos, uma vez que o

SAMPA recorre aos caracteres disponíveis no teclado dos computadores (Wells 1997). Neste artigo, o alfabeto usado é o SAMPA, salvo uma exceção de pormenor onde se recorre igualmente ao IPA pela necessidade de comparação e de esclarecimento das representações sonoras. O SAMPA disponibiliza um alfabeto dirigido para a língua portuguesa – *SAMPA for Portuguese*

Dígrafos	Fonemas	Exemplo
cc	s	perfeccionismo → /p@rfEsiuniSmu/ (<i>pré-AO</i>)
	ks	ficcional → /fiksiunal/
cç	s	intersecção → /i~t@rsEs6~w~/ (<i>pré-AO</i>)
	ks	ficção → /fiks6~w~/
rr	R	carro → /kaRu/
ss	s	massa → /mas6/
lh	L	molho → /moLu/
nh	J	unha → /uJ6/
ch	S	chave → /Sav@/
ou	o	dourada → /dorad6/
an, am	6~	canto → /k6~tu/
		angolano → /6~gul6nu/
		campo → /k6~pu/
en, em	e~	sente → /se~t@/
		ensaio → /e~saiu/
		sempre → /se~pr@/
in, im	i~	cinco → /si~ku/
		limbo → /li~bu/
on, om	o~	conto → /ko~tu/
		pomba → /po~b6/
un, um	u~	assunto → /asu~tu/
		tumba → /tu~b6/
ân, âm	6~	pântano → /p6~t6nu/
		iâmbico → /i6~biku/
ên, êm	e~	ênfase → /e~f6z@/
		êmbolo → /e~bulu/
ín, ím	i~	índio → /i~diu/
		límpido → /li~pidu/
ôn, ôm	o~	cônsul → /ko~sul/
		cômputo → /ko~putu/
ún, úm	u~	denúncia → /denu~si6/
		cúmplice → /ku~plis@/

Tabela 3 – Lista de dígrafos associados a fonemas possíveis e exemplos convencionados. Os primeiros 7 símbolos representam consoantes; os restantes 11 representam seqüências específicas de vogal ditongada e de vogais nasais. De observar que as duas primeiras representações de dígrafos podem constituir-se como uma estrutura alternativa do português grafado atendendo às normas listadas pelo Acordo Ortográfico de 1990.

(SAMPA-PT) (disponível em <http://www.phon.ucl.ac.uk/home/sampa/portug.htm>), acerca do qual merece ser tecida uma apreciação. Na verdade, a observação atenta dos alfabetos fonéticos SAMPA-PT e X-SAMPA dá conta de alguma indefinição de regularidade, exemplificada na atribuição de mais do que um símbolo para o mesmo som. A título de exemplo, o símbolo [r] no SAMPA-PT parece ter como correspondente no X-SAMPA o símbolo [4] (IPA: [r]), simbolizando o [r] no X-SAMPA a vibrante alveolar múltipla (IPA: [r]).

189

2.3. Notação

Independentemente do alfabeto usado na conversão, há que ponderar sobre a forma de notação no sentido de otimizar o conversor. Por vezes, poderá ser necessário afinar a notação até um nível que poderá admitir o processo de *iod* e de *sandhi* (exemplificadas em <extra> [6iStr6] e em <a água> [6 i agu6]), numa tentativa de aproximação aos efeitos sonoros provenientes de coarticulação, ou a não consideração da velarização da lateral (presente em <sal>), caso se considere que o contexto fonético é suficiente para categorizar a unidade. Outra situação em termos de notação e que não tem sido consensual é a que se relaciona com a questão dos hiatos, uma vez que as vogais <e>, <i>, <o> e <u> podem ser notadas como parte dos chamados ditongos crescentes (ou falsos ditongos), presentes em sequências como em <Leonor>, <relógio>, <póvoa> e <água>. A decisão da notação de glides, em especial nas sequências que formam os ditos ditongos crescentes, tem vindo a ser ponderada pela necessidade de caracterização do ritmo de fala, e, ainda que a questão tenha sido já levantada na literatura da área (cf., a título de exemplo, o estudo de (Rua *et al.* 2005)), não existe, tanto quanto é do nosso conhecimento, um estudo acústico, sustentado num corpus expressivo do PE falado, que suporte a distinção clara entre as vogais [i] - [j] e [u] - [w]. Uma vez que o contexto fonético é esclarecedor da estrutura da sequência, a nossa opção metodológica recai em não considerar pertinente a identificação de ditongos crescentes nem a notação das vogais ditas decrescentes do ditongo como glides. Admitimos, no entanto, que não é um formato de representação comum, e que pode eventualmente suscitar alguma discussão em termos da teoria fonológica.

2.4. Alternâncias Vocálicas

190 Um dos problemas com o qual o desenvolvimento de um conversor de G2P se tem vindo a deparar diz respeito à situação em que uma vogal, envolvida pelo mesmo contexto, pode apresentar uma alternância de pronúncia. Em PE, referimo-nos especificamente à alternância entre os pares /e/-/E/ e /o/-/O/. Casos de homógrafos heterófonos, nos quais as vogais <e> e <o> podem ser pronunciadas como /e/ e /E/ ou /o/ e /O/, não têm sido de fácil resolução, como já se mostrou no estudo de (Braga e Marques 2007). <selo> ou <namoro>, em estruturas frásicas tais como <Eu **selo** a carta com um **selo** comemorativo.> e <O **namoro** dos dois vem já de longa data, ao passo que eu **namoro** apenas há um mês.>, são disso exemplos. Poder-se-ia implementar um módulo que identificasse a categoria gramatical dos homógrafos, estratégia que desambiguaria ocorrências como <selo> e <namoro> nos exemplos, associando-se à categoria gramatical a pronúncia esperada. No entanto, exemplos como <lobo> em estruturas frásicas tais como <Ele caiu na boca do **lobo**.> e <O **lobo** frontal relaciona-se com a inibição de comportamentos.> mostram que homógrafos com a mesma categoria gramatical (ao caso, a do nome) podem ser igualmente heterófonos (de pronúncias /lobu/ e /lObu/, respetivamente, nos exemplos). Outra situação de alternância vocálica a referir ocorre em vocábulos que, do ponto de vista diacrónico, sofreram metafonía, alterando a pronúncia da vogal tónica por influência da vogal átona final, num processo de assimilação regressiva. Assim, a vogal <o> é pronunciada como /o/ em <miolo> e em <forno> e como /O/ em <miolos> e em <fornos>, independentemente de, na palavra latina primitiva, se pronunciar com o chamado ‘o breve’.

3. Abordagens Teóricas

Várias têm sido as abordagens teóricas adotadas no sentido de se encontrar uma solução para a tarefa da conversão G2P. No que diz respeito ao PE, destacamos as seguintes:

- por regras inferidas a partir dos dados (Teixeira *et al.* 2006);
- por máquinas de estados finitos (Caseiro *et al.* 2002);
- por modelos de máxima entropia (Barros e Weiss, 2006);
- por regras linguísticas (Braga *et al.* 2006), (Oliveira *et al.* 1992), (Teixeira 2004), (Candeias e Perdigão 2008); e
- por modelos híbridos (Veiga *et al.* 2011).

Uma outra técnica que poderá ser mencionada, e que é aplicada especialmente em línguas que não apresentam uma clara correspondência entre grafema e fone(ma), como é o caso do inglês, é a que resulta de uma abordagem por meio de um dicionário com léxico e respetiva notação fonético-fonológica. Funcionando como uma tabela de recurso finita, esta técnica falha sempre que surgem novos vocábulos para conversão.

3.1. Regras Linguísticas

Regras para marcação da tonicidade do vocábulo (Candeias e Perdigão, 2008; Almeida e Simões, 2001; Braga e Resende, 2007; Viana e d'Andrade, 1985) e para a identificação da unidade silábica (Almeida e Simões 2001; Braga e Resende 2007; Teixeira *et al.* 2000; Teixeira *et al.* 2006), foram usadas na implementação de algoritmos para a conversão G2P. A relevância da implementação de regras linguísticas deve-se, em grande parte, à razoável regularidade fonética e fonológica bem como à ortografia de base fonética-fonológica que a língua portuguesa, na sua vertente europeia apresenta, a par das línguas românicas na sua generalidade. Resolvendo-se os casos de falta de univocidade com algumas estratégias, como a da configuração de uma nova entidade, denominada grafonema, identificativa da associação unívoca entre grafema e respetivo símbolo fonético/fonológico (Bisani e Ney 2008), são as características da estrutura da língua que explicam, em parte, o êxito da aplicação de regras linguísticas, facto que os trabalhos aqui referenciados exemplificam.

3.2. Modelos Mistos

192

Contrastando com as abordagens baseadas em regras suportadas por um conhecimento linguístico da língua, que se pretende exaustivo, as abordagens estatísticas baseiam-se no pressuposto de que a pronúncia de um vocábulo pode ser prevista, por analogia, a partir de exemplos de sequências suficientes. Se, por um lado, uma das vantagens apontada pelos modelos estatísticos consiste em não implicar uma verificação constante da interdependência das regras, em especial quando surge uma sequência de grafemas que sai fora das regularidades até então admitidas, por outro, tem-se verificado que a conversão G2P proveniente desses modelos não capta um contexto suficientemente amplo de forma a impedir que a estrutura fonológica do PE seja violada. A título de exemplo, o contexto vizinho do grafema <a> na sequência <vaga>, não é suficiente para determinar a pronúncia a atribuir na medida em que em <vagarosa> ou em <divagar>, a pronúncia é /6/ (/v6g6rOz6/ e /div6gar/, respetivamente, e não /a/ como em /vag6/). Assim, a reunião de abordagens estatísticas e por regras linguísticas foi feita com êxito em (Veiga *et al.* 2011), onde algoritmos baseados em regras para a acentuação vocálica, reconhecendo o núcleo da sílaba tónica de cada vocábulo, foram imbuídos nos algoritmos dos modelos probabilísticos do conversor G2P¹¹⁴.

4. Conclusões e Desafios

Um dos módulos principais que integra a arquitetura geral de sistemas de processamento da fala do português diz respeito à tarefa de encontrar a pronúncia de um dado vocábulo na sua forma escrita. Não sendo uma tarefa de fácil resolução, várias têm sido as abordagens adotadas no sentido de se encontrar uma solução eficaz para o processo de mapear grafemas e

¹¹⁴O sucesso do modelo misto que congrega abordagens estatísticas e por regras fonológicas é possível de ser comprovado pela experimentação do conversor em <http://www.co.it.pt/~labfala/g2p/>.

fone(ma)s. Modelos baseados em regras, ou em probabilidades de sequências de grafonemas são duas estratégias adotadas comuns.

Neste artigo, foi nosso objetivo mostrar como o conhecimento fornecido pela linguística teórica tem sido, e pode ser, integrado em sistemas automáticos de processamento da língua portuguesa, nomeadamente no módulo de conversão de G2P. Nesse sentido, foram apresentadas soluções desenvolvidas no seio do grupo que os autores integram¹¹⁵ e que incluem a identificação de vogal tónica, a divisão silábica e a conversão de grafemas em fone(ma)s.

Simultaneamente, foi igualmente nossa intenção identificar novos rumos de orientação para o estudo da língua portuguesa, apontando-se a engenharia do processamento linguístico ou a linguística tecnológica como uma linha de investigação onde investir.

5. Referências

- Almeida, J. e A. Simões (2001). Text to Speech: a rewriting system approach. XVII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN), Jaén: Universidad de Jaén, 247-255.
- Andrade, A. (1994). Reflexões sobre o ‘e mudo’ em Português Europeu. Actas do Congresso Internacional sobre o Português. Lisboa: Colibri, vol II, 303-344.
- Barbosa, J. Morais (1994). Introdução ao Estudo da Fonologia e Morfologia do Português. Coimbra: Almedina.
- Barros, M. e C. Weiss (2006). Maximum Entropy Motivated Grapheme-to-Phoneme, Stress and Syllable Boundary Prediction for Portuguese Text-to-Speech. IV Jornadas en Tecnologías del Habla. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla, 177-182.
- Bisani, M. e H. Ney (2008). Joint-Sequence Models for Grapheme-to-Phoneme Conversion. Speech Communication, Amsterdam: Elsevier, vol. 50(5), 434–451.
- Braga, D. e M. A. Marques (2007). Desambiguação homógrafos para Sistemas de conversão Texto-Fala em Português. Diacrítica 21.1. Braga: CEHUM/Universidade do Minho, 25-50.
- Braga, D., L. Coelho e F. Resende (2006). A Rule-Based Grapheme-to-Phone Converter for TTS Systems in European Portuguese. VI International Telecommunications Symposium. Fortaleza-CE, Brazil: IEEE Signal Processing Society, 328-333.
- Braga, D. e F. Resende (2007). Módulos de Processamento de Texto Baseados em Regras para Sistemas de Conversão Texto-Fala em Português Europeu. XXII Encontro da Associação Portuguesa de Linguística. Lisboa: Colibri, 141-155.

¹¹⁵ Laboratório de Processamento de Sinal (<http://lsi.co.it.pt/spl/>), do Instituto de Telecomunicações, polo de Coimbra.

- Candeias, S. e F. Perdigão (2008). Conversor de grafemas para fones baseado em regras para português. Perspectivas sobre a Linguateca. Actas do Encontro Linguateca: 10 anos. Linguateca, Lisboa, cap.14.
- Candeias, S., A. Veiga, C. Lopes e F. Perdigão (2011). A realização do schwa no Português Europeu. Comunicação apresentada no II Workshop On Portuguese Description-JDP (Satellite Workshop of STIL 2011), October 24-26, 2011. Cuiabá, UFMG, Brazil.
- Candeias, S. e F. Perdigão (2011). Investigating new syllables prototypes for the Portuguese language. Proceedings of the International Congress of Phonetic Sciences (ICPhS XVII), Hong Kong, 388-391. (<http://www.icphs2011.hk/resources/OnlineProceedings/RegularSession/Candeias/Candeias.pdf>)
- Candeias, Sara (2007). Sistema fonológico da Beira Interior e algumas considerações sintáctico-semânticas. PhD Thesis, Universidade de Aveiro (<http://ria.ua.pt/handle/10773/2871>)
- Caseiro, D., I. Trancoso, I., L. Oliveira e C. Viana (2002). Grapheme-to-Phone Using Finite-State Transducers. Proceedings of the IEEE Workshop on Speech Synthesis. Santa Monica, USA: IEEE Signal Processing Society, 215-218.
- Mateus, M. e E. d'Andrade, E. (2000). The Phonology of Portuguese. Oxford University Press.
- Mateus, M. (2006). Sobre a natureza fonológica da ortografia portuguesa. In: Questões de Fonética e Fonologia: uma Homenagem a Luís Carlos Cagliari. Bahia: Universidade Estadual do Sudoeste da Bahia, 159-180.
- Oliveira, L., M. Viana e I. Trancoso (1992). A Rule-Based Text-to-Speech System for Portuguese. Proceedings of The International Conference on Acoustics, Speech, and Signal Processing (ICASSP). San Francisco, USA: IEEE Signal Processing Society, vol. 2, 73-76.
- Oliveira, L. (1996). Síntese de Fala a Partir de Texto. PhD Thesis, Instituto Superior Técnico – Universidade Técnica de Lisboa.
- Rua, C., A. Teixeira e L. Moutinho (2005). Ditongos Orais no Português Europeu. In: Actas do XX Encontro da Associação Portuguesa de Linguística. Lisboa: APL/Colibri, 803-816.
- Teixeira, A., C. Oliveira e L. Moutinho (2006). On the Use of Machine Learning and Syllable Information in European Portuguese Grapheme-Phone Conversion. Proceedings of The International Conference on Computational Processing of Portuguese (PROPOR). Springer, 212-215.
- Teixeira, J. P. (2004). A Prosody Model to TTS Systems. PhD Thesis, Faculdade de Engenharia da Universidade do Porto.
- Teixeira, J. P., P. Gouveia e D. Freitas (2000). Divisão Silábica Automática do Texto Escrito e Falado. Proceedings of The International Conference on Computational Processing of Portuguese (PROPOR). Atibaia, SP, Brasil, vol. 1, 88-98.
- Veiga, A., S. Candeias e F. Perdigão (2011). Generating a Pronunciation Dictionary for European Portuguese Using a Joint-Sequence Model with Embedded Stress Assignment. Proceedings of the Brazilian Symposium in Information and Human Language Technology (STIL). Cuiabá, Mato Grosso, Brazil, 144 - 153.
- Veiga, A., S. Candeias, L. Sá, e F. Perdigão (2010). Using Coarticulation Rules in Automatic Phonetic Transcription. Proceedings of PROPOR 2010. Porto Alegre. (http://www.inf.pucrs.br/~propor2010/proceedings/regular_papers/VeigaEtAl.pdf)
- Viana, M. C. e E. d'Andrade (1985). *CORSO I: um conversor de texto ortográfico em código fonético para o português*. Relatórios do grupo de fonética e fonologia n. 6. Lisboa: CLUL.

(Página deixada propositadamente em branco)

Série Investigação

•

Imprensa da Universidade de Coimbra

Coimbra University Press

2011

Obra publicada
com a Coordenação
Científica



CENTRO DE ESTUDOS
DE LINGUÍSTICA
GERAL E APLICADA DA
UNIVERSIDADE DE COIMBRA

• U



C •

