

Présentation

Le volume 3 de la revue *Linguística* offre une sélection des communications présentées au colloque international *Change and Variation in Romance*, ainsi que les textes des deux conférenciers invités, les professeurs Gianpaolo Salvi (Université Loránd Eötvös, Budapest, Hongrie) et Rosa Virgínia Mattos e Silva (Université Fédérale de Bahia, Brésil). Le colloque a eu lieu à la Faculté des Lettres de l'Université de Porto, du 13 au 15 décembre 2007, et a été organisé par la Section de Linguistique du Département de Langues Romanes de la Faculté des Lettres et par le Centre de Linguistique de l'Université de Porto.

Le présent volume reflète la question fondamentale débattue lors du colloque: dans quelle mesure les recherches sur l'histoire et la variation des langues romanes nous apportent-elles des éléments pour la compréhension du changement linguistique en général et du rapport entre changement et variation en particulier.

L'intérêt pour les langues romanes se justifie encore pleinement. Comme l'affirme Martin Harris (1988: 24), «(...) quoique aucune des langues romanes ne puisse prétendre rivaliser avec l'anglais, la langue internationale depuis la deuxième moitié du XX^e siècle, on ne peut pas oublier que, parmi la douzaine de langues les plus parlées dans le monde, figurent trois langues romanes, présentant chacune un continuum de variantes qui vont des normes métropolitaines aux normes régionales et à des formes plus ou moins créolisées. Ce que l'Empire romain a fait pour le latin, l'expansion coloniale et la facilité contemporaine de voyager et de communiquer l'ont fait pour les langues romanes.» (ma traduction)¹

D'autre part, les études sur le changement et la variation constituent toujours un défi majeur pour les théories linguistiques au début du XXI^e siècle. Il ne s'agit pas de réintroduire le type de recherches que l'on faisait au XIX^e siècle mais de renouveler les études diachroniques et de variation à la lumière des théories nées aux siècles derniers.

¹ Harris, M. (1988) «The Romance Languages» in Martin, H. & N. Vincent (orgs.) *The Romance Languages*, Croon Holm, Londres & Sydney, p. 24.

Avec la découverte, à la fin du XVIII^e siècle, des similitudes entre langues éloignées dans l'espace et dans le temps – le sanscrit, le grec, le latin, le vieux norvégien, le vieil allemand – par William Jones, Bopp, Rask, et plus tard Grimm et Verner, la Linguistique Comparée était née. Mais la conception sous-jacente à ces recherches reposait sur l'idée que les langues sont des mécanismes semblables aux êtres vivants, une pensée très influencée par les idées de Darwin. Les notions de temps et de contexte culturel vont peu à peu transformer les recherches comparatistes en recherches historiques, qui vont marquer la fin du XIX^e siècle, notamment avec l'école des *Jung-Grammatiker*. Sous l'influence de ces recherches historiques naît la philologie, qui est surtout l'étude de la langue dans les textes et qui a une incidence particulière dans les domaines germanique et roman.

Saussure et les structuralismes européen et américain viendront ébranler les recherches historiques, en valorisant la conception de la langue en tant que système organisé et en privilégiant la synchronie sur la diachronie.

En outre, les premières propositions de Chomsky, selon lesquelles l'objet de la Linguistique est la description de la compétence d'un locuteur idéal, non soumis à des contraintes de mémoire et d'attention, ont valorisé pendant longtemps l'idée d'une stabilité des langues très éloignée de la réalité.

Au début de la deuxième moitié du XX^e siècle, on en arrive donc, en gros, à la situation suivante: d'une part, l'héritage d'un paradigme historico-comparatif qui souligne l'importance de la variation culturelle et de l'histoire, qui analyse surtout le lexique et où il n'y a pas de place pour la grammaire, et, d'autre part, un paradigme synchronique qui défend l'unité biologique du langage, sans préoccupation historique, et qui fait de la notion de grammaire le centre de toute recherche linguistique.²

Étant donné qu'il n'est plus possible de nier ni le caractère biologique du langage humain, ni l'importance de la notion de grammaire, ce qui continue à diviser les linguistes est le poids du social

² Cf. Longobardi, G. (2003) «On Parameters and Parameter Theory» in Stark, E. & Wandruszka, U. (eds.) *Syntaxtheorien. Modelle, Methoden, Motive*, G. Narr, Tübingen, 2003, pp. 273-290.

et du culturel en ce qui concerne l'explication du changement et de la variation.

Ainsi, l'on a, du côté du social, Labov, en particulier ses notions de «changement en progrès» et de «changement en temps apparent», qui ont lié, de façon cruciale, variation et changement, et l'approche de Thomason & Kaufman (1988), qui affirment que «c'est l'histoire sociolinguistique des sujets parlants et non la structure de leur langue qui détermine en premier le résultat du contact entre langues» (p. 35, ma traduction).

Du côté non social, il faut citer, entre autres, les théories de la grammaticalisation de Hopper & Tragott, mais aussi de Lehmann, de Heine et d'autres, développant une notion que l'on doit d'ailleurs à Meillet, ainsi que la Théorie des Principes et des Paramètres, proposée par Chomsky dans les années quatre-vingt et développée par Lightfoot, pour qui le changement est l'effet du rapport entre principes universaux et le rétablissement de paramètres en phase d'acquisition du langage selon les pistes de l'input («cue-based approach»).

Au début du XXI^e siècle, on a donc plusieurs approches théoriques sur le changement des langues ainsi que des hypothèses diversifiées sur les rapports entre changement et variation.

Les textes qui suivent constituent des réponses différentes à ces questions, à partir des langues romanes et de divers domaines linguistiques.

Les textes de Gianpaolo Salvi et de Rosa Virgínia Mattos e Silva, qui ouvrent le volume, posent des problèmes plus généraux. Salvi analyse un exemple de changement en italien, le parcours de la construction avec le pronom réfléchi *si*, qui en ancien italien n'avait que la valeur passive, contrairement à l'italien moderne, où il a la valeur passive et la valeur impersonnelle, comme dans d'autres langues romanes. À partir de cet exemple, Salvi montre que le changement peut être considéré comme le résultat de petits changements locaux qui éliminent des imperfections singulières et qui contribuent à la formation d'un système plus cohérent, mais pas nécessairement parfait.

En partant du projet brésilien «Programme pour l'histoire de la langue portugaise», Rosa Virgínia Mattos e Silva analyse les rapports entre l'histoire ou les histoires de la langue portugaise et les différentes

approches théoriques sur le changement linguistique, pour montrer que la complexité du changement justifie un choix théorique diversifié.

Le portugais est analysé au niveau syntaxique, morphologique, phonologique et lexical. Avelar et Cyrino étudient les constructions du portugais brésilien contenant des expressions locatives en position de sujet et se demandent à quel point une telle construction peut être le résultat de l'influence des langues bantoues sur la syntaxe de cette variante du portugais. Cardoso étudie les relatives de prédicat du type *mulher que foi de Lourenço Anes*, fréquentes en ancien portugais, mais ayant disparu de la grammaire du portugais contemporain, en explorant quelques conséquences théoriques de ce changement. Rio-Torto adopte un point de vue diachronique pour analyser la formation de mots en *-eir* en portugais et pour montrer que le parcours de ce suffixe révèle un processus de grammaticalisation croissante. Luís étudie des phénomènes de morphologie flexionnelle dans les variétés indo-portugaises parlées à Korlai, Daman et Diu, en les comparant à d'autres variétés du portugais parlées en Afrique et en Asie, pour essayer de comprendre quels facteurs justifient la diversité des formes flexionnelles analysées. Nunes étudie la terminologie du sucre au Brésil et essaie de comprendre les mécanismes de changement lexical à partir de la terminologie originelle du sucre à Madère et compte tenu de la variation géographique et sociale au Brésil. Helena Paiva analyse le phénomène du vocalisme atone à partir du témoin implicite et explicite des grammairiens du XVI^e siècle.

Le galicien est également présent dans ce volume grâce à Mariño Paz, qui étudie l'importance du purisme et de la normalisation comme facteurs décisifs du changement linguistique en partant de l'exemple des pluriels oxytons terminés en /l/ au singulier en galicien contemporain.

Quant au français, il est étudié ici à partir des connecteurs causaux, du lexique du sport et des prépositions locatives. Degand et Fagard analysent du point de vue synchronique et diachronique les connecteurs *car* et *parce que* en français, montrant ainsi que *parce que* exprime plutôt des relations causales plus objectives et *car* des relations causales plus subjectives et même pas nécessairement causales. Mélanie Bernard étudie l'emploi des anglicismes dans la langue du sport en français et les conditions sociales de leur utili-

sation, à partir d'une analyse du journal sportif *L'Équipe* pendant les Jeux Olympiques d'Athènes en 2004. Enfin, Štichauer analyse les prépositions *dans, sur, en* et *à* d'un point de vue diachronique, en explorant la relation entre syntaxe et conceptualisation (au sens cognitif) de l'espace.

À la sélection des textes du présent volume ont participé les chercheurs suivants, que je remercie vivement: Ana Maria Martins, Christianne Marchello-Nizia, Clarinda Azevedo Maia, Eulália Bonet, Gabriela Matos, Gianpaolo Salvi, Ivo Castro, Margarita Correia, Maria Antónia Mota, Marina Vigário, Rosa Virgínia Mattos e Silva, Sónia Frota.

Je veux aussi remercier la Faculté de Lettres de l'Université de Porto, la Fondation pour la Science et la Technologie (FCT) et le Centre de Linguistique de l'Université de Porto qui ont permis la publication de ce volume.

Ana Maria Brito

Porto, le 30 juin 2008.

La formazione della costruzione impersonale in italiano

Giampaolo Salvi

gps@ludens.elte.hu

Università Eötvös Loránd, Budapest (Ungheria)

RIASSUNTO: In italiano moderno possiamo identificare due diverse costruzioni in cui la coniugazione pronominale viene usata per la demozione del soggetto lessicale: la costruzione del *si* passivo e quella del *si* impersonale, che mostrano proprietà molto diverse. L'italiano antico aveva solo il *si* passivo, mentre la costruzione impersonale è il risultato di una serie di cambiamenti minori nelle regole e nel dominio di applicazione della costruzione passiva: questi cambiamenti sono cominciati nel periodo dell'italiano antico e si sono svolti durante almeno cinque secoli. Non essendo il risultato di un piano unitario, la nuova costruzione mostra (ancora) delle "imperfezioni". Nel caso esaminato, il cambiamento sintattico è consistito di cambiamenti locali che sono il risultato del caso (cioè di cause esterne) o di tentativi di correggere le imperfezioni insite nel sistema (cause interne).

PAROLE CHIAVE: coniugazione pronominale (riflessiva), costruzione passiva, impersonale, italiano, sintassi diacronica

ABSTRACT: In Modern Italian, one can identify two independent reflexive ("*si*") constructions which syntactically demote the lexical subject: a passive one and an impersonal one, with quite distinct properties. Old Italian only had passive *si* – the impersonal construction is the result of many small changes in the rules and the domain of application of the passive construction: these changes began in the Old Italian period and lasted for at least five centuries. But this new construction, not being the result of a unitary project, continues to show signs of being imperfectly put together. In this case study, syntactic change consists of (a series of) small, local changes which are the result of chance (external causes) or are brought about by the possible imperfections of the linguistic system (internal causes).

KEY-WORDS: diachronic syntax, impersonal, Italian (language), passive construction, reflexive (pronominal) conjugation

Come nelle altre lingue romanze, anche in italiano la coniugazione pronominale del verbo, accanto ad altri usi, può servire alla demozione sintattica del soggetto lessicale: il risultato è una costruzione in cui il soggetto lessicale non espresso ha un'interpretazione generica

o indeterminata. In italiano moderno questa possibilità si realizza in due costruzioni sintattiche distinte, tradizionalmente denominate *si passivo* e *si impersonale*. Mentre le strutture che appartengono al *si passivo* figurano fin dalle origini tra le strutture dell'italiano, quelle che appartengono al *si impersonale* sono apparse solo gradualmente nell'uso linguistico, e ancora nella lingua attuale rappresentano, come vedremo, un insieme meno coerente rispetto al *si passivo*.

Scopo di questo lavoro è di individuare alcune delle tappe e dei meccanismi che hanno portato alla formazione della costruzione del *si impersonale*. In particolare, dopo aver descritto brevemente le principali caratteristiche delle due costruzioni dell'italiano moderno (par. 1), mostreremo come in italiano antico esisteva soltanto la costruzione del *si passivo* (par. 2). Vedremo poi in che modo l'ambito della costruzione sia andato man mano ampliandosi fino ad arrivare, nella seconda metà dell'Ottocento, alla situazione moderna, e cercheremo di individuare i meccanismi attraverso i quali si sono imposte le singole innovazioni (par. 3). Concluderemo con alcune considerazioni generali relative ai meccanismi del cambiamento linguistico (par. 4).

Prima di addentrarci nell'analisi dei fatti linguistici, premettiamo alcune considerazioni relative ai dati utilizzati e alle analisi proposte.

Dati. Le costruzioni studiate mostrano nell'italiano moderno una grande variabilità a livello geografico e di registro. Per offrire un insieme di dati il più possibile coerente si è scelto di utilizzare un'unica varietà, che corrisponde all'uso colto dell'Italia settentrionale (o perlomeno alla sua varietà più diffusa, che è anche quella dell'autore), come descritto in Salvi & Vanelli (2004: 72-78). Per quanto riguarda l'italiano antico, la varietà descritta è la lingua parlata e scritta a Firenze nella seconda metà del Duecento e nei primi decenni del Trecento, oggetto del Progetto ItalAnt, diretto da Lorenzo Renzi e Giampaolo Salvi (cfr. la presentazione in Renzi 2000 e la descrizione della nostra costruzione in Salvi 2006: par. 3).

Mentre le descrizioni dell'italiano antico e dell'italiano moderno si basano su analisi dettagliate e nei limiti del possibile complete del materiale a disposizione, i dati relativi alle fasi intermedie sono stati raccolti con ricerche mirate all'individuazione di alcune innovazioni sintattiche, ricerche effettuate su testi informatizzati (disponibili in rete

nei siti <http://www.liberliber.it/home/index.php> e <http://www.bibliotecaitaliana.it/exist/bibit/>; citeremo esplicitamente soltanto quelle opere che hanno rilevanza dal punto di vista dell'introduzione di innovazioni, ma in realtà l'analisi è stata estesa a un numero molto maggiore di testi). Dato il tipo di raccolta, questi dati rappresentano solo una campionatura molto parziale rispetto ai dati disponibili e non permettono quindi una ricostruzione completa dei sistemi sincronici da cui sono stati prelevati (su questo v. anche le osservazioni all'inizio del par. 3). Le conclusioni del nostro esame diacronico vanno dunque valutate entro questi limiti e devono essere considerate solo una prima approssimazione, in attesa che una raccolta più sistematica di dati permetta una ricostruzione più dettagliata delle fasi intermedie. Un'ulteriore limitazione è data dal fatto che i testi studiati sono quasi esclusivamente opere letterarie, probabilmente rappresentative di un uso linguistico più controllato e meno aperto alle innovazioni (v. anche sotto 3.2).

Analisi. L'analisi presentata nel par. 1 accetta l'intuizione tradizionale che in italiano moderno l'uso della coniugazione pronominale per la demozione del soggetto lessicale si realizzi attraverso due costruzioni distinte. Mentre però l'analisi tradizionale fa appello al valore semantico delle denominazioni *passivo* e *impersonale*, l'analisi che proponiamo per l'italiano moderno si basa esclusivamente su proprietà sintattiche: chiamiamo *passive* tutte le strutture in cui la demozione del soggetto lessicale è accompagnata dalla promozione dell'oggetto diretto a soggetto (1.1), *impersonali* tutte quelle strutture in cui la demozione non è accompagnata da alcuna promozione (1.2). Come vedremo, però, le due costruzioni sono quasi in distribuzione complementare perché la costruzione impersonale non realizza tutte le sue potenzialità strutturali; sarebbe quindi concepibile anche un'analisi che consideri tutti gli usi come manifestazioni di un'unica costruzione. Dal punto di vista dei problemi affrontati in questo lavoro, tuttavia, questo non cambierebbe molto: anche all'interno di un'analisi unitaria bisognerebbe spiegare come la costruzione sia venuta man mano ampliando il suo ambito di applicazione a partire dall'italiano antico, dove aveva un ambito molto più ristretto.

Sempre per quanto riguarda l'uso dei due termini *passivo* e *impersonale*, le proprietà definitorie delle singole costruzioni possono cambiare in fasi diverse della storia di una lingua, in particolare alcune

strutture che oggi classificherebbero come impersonali in base alla definizione del capoverso precedente, saranno considerate come casi di costruzione passiva nel periodo dell'italiano antico (1.2); questa diversa classificazione dei dati è in sintonia, come vedremo, con le proprietà di costruzioni parallele presenti nella lingua (nel caso specifico la costruzione passiva perifrastica), proprietà che anch'esse sono cambiate nel corso del tempo. Per chiarezza useremo sempre la denominazione di *si impersonale* per identificare la costruzione dell'italiano moderno descritta in 1.2. Con senso più ampio useremo il termine *impersonale* anche per indicare strutture che non presentano un soggetto sintattico: per es. le strutture esemplificate in (18)-(19) saranno considerate varianti impersonali della costruzione passiva.

Studiando il progressivo ampliamento delle possibilità strutturali delle nostre costruzioni, faremo delle proposte concrete sui meccanismi che hanno contribuito all'introduzione di innovazioni nel sistema dell'italiano (proposte del resto in buona parte tradizionali). Dobbiamo però ricordare anche qui che, dato il carattere molto parziale dei dati raccolti, queste proposte vanno intese piuttosto come suggerimenti che dovranno essere verificati su dati più completi.

L'analisi sviluppata in questo lavoro riprende e in alcuni punti cerca di migliorare quella presentata in Salvi (2008). Altri dati sulla storia del *si* passivo e impersonale si trovano nell'importante studio di Wehr (1995), con le cui conclusioni concordiamo solo parzialmente.

1. Le costruzioni con *si* in italiano moderno

La coniugazione pronominale in it. mod. può esprimere la riflessività (*Piero si lava, Piero si lava le mani*) e la reciprocità (*I ragazzi si salutano, I ragazzi si mandano i saluti*) o contribuire a esprimere l'inaccusatività, sia in opposizione a una variante transitiva (*annerirsi* vs. *annerire*), sia senza (*arrabbiarsi*). Oltre a questi usi lessicali, la costruzione ha anche un uso grammaticale completamente regolare, in cui segnala la demozione sintattica del soggetto lessicale del verbo. Questo uso grammaticale si realizza in due costruzioni distinte: quella del *si passivo*, parallela alla normale costruzione passiva perifrastica (1.1), e quella del *si impersonale*, parallela alla normale costruzione attiva della frase (1.2).

1.1. *Si passivo*

La costruzione del *si* passivo presenta le caratteristiche che sono tipiche anche della costruzione passiva perifrastica: abbiamo demozione del soggetto lessicale (che non viene normalmente realizzato sintatticamente) e promozione dell'oggetto diretto a soggetto, come esemplificato in (1) e schematizzato in (2). L'attribuzione dello statuto di soggetto al SN *i diplomi* (masch. pl.) in (1b) si basa sul fatto che il SN è accordato in numero e genere con la perifrasi verbale *si sono consegnati* (masch. pl.), fenomeno che in italiano, in questa forma, è proprietà esclusiva dei SN soggetto (il cambio dell'ausiliare da *avere* (1a) a *essere* (1b) è una conseguenza automatica dell'uso del clitico riflessivo):

- (1) a. Il presidente *ha consegnato* i diplomi agli studenti
 b. *Si sono consegnati* i diplomi agli studenti
- (2) Soggetto (*il presidente*) → ∅
 Oggetto Diretto (*i diplomi*) → Soggetto
 Oggetto Indiretto (*gli studenti*) → Oggetto Indiretto

Inoltre, il SN può anche comparire nella posizione di soggetto preverbale:

- (3) I diplomi *si consegnano* agli studenti in occasione della festa nazionale

Come nel caso della normale costruzione passiva, l'eliminazione del soggetto lessicale è solo sintattica: l'attante corrispondente al soggetto lessicale resta presente semanticamente nella frase con un'interpretazione generica o indeterminata, a seconda del contesto.

Sempre come nel caso della normale costruzione passiva, in questa costruzione possono comparire solo verbi transitivi, i soli che dispongano di un oggetto diretto lessicale che possa essere promosso a soggetto. Ma diversamente dalla normale costruzione passiva, il *si* passivo è possibile solo se l'oggetto diretto è di 3. pers., per cui l'unico clitico riflessivo ammesso è *si*.

1.2. *Si impersonale*

Nella costruzione del *si* impersonale la demozione (ed eliminazione) del soggetto lessicale non è accompagnata dalla promozione di alcun argomento a soggetto, come esemplificato in (4) e schema-

tizzato in (5). Il verbo appare nella forma non marcata della 3. pers. sing. (il participio in (4a,b) è accordato con l'oggetto diretto, il clitico *li*; in altri casi il participio e il complemento predicativo possono apparire in forma pl., come negli ess. (10)-(11), un problema che non affronteremo qui—ma cfr. sotto 3.2):

- (4) a. Il presidente *li ha consegnati* agli studenti
 b. *Li si è consegnati* agli studenti
- (5) Soggetto (*il presidente*) → ∅
 Oggetto Diretto (*li*) → Oggetto Diretto
 Oggetto Indiretto (*gli studenti*) → Oggetto Indiretto

Anche in questa costruzione l'eliminazione del soggetto lessicale è solo sintattica: l'attante corrispondente al soggetto lessicale resta presente semanticamente nella frase con un'interpretazione generica o indeterminata, a seconda del contesto.

Diversamente dal *si* passivo, il *si* impersonale è possibile con ogni classe di verbi: transitivi (6), intransitivi, sia inergativi (7) che inaccusativi (semplici (8) o pronominali (9)), compresa la copula (10) e la costruzione passiva perifrastica (11) (per l'uso di *ci* al posto di *si* in (9b), v. il commento a (16) e sotto 3.3):

- (6) a. Piero *guarda* solo te
 b. *Si guarda* solo te
- (7) a. Piero *dorme*
 b. *Si dorme*
- (8) a. Piero *va*
 b. *Si va*
- (9) a. Piero *si addormenta*
 b. *Ci si addormenta*
- (10) a. Piero è intelligente
 b. *Si è intelligenti*
- (11) a. Piero *viene invitato* spesso
 b. *Si viene invitati* spesso

Come abbiamo visto, ambedue le costruzioni sono possibili con i verbi transitivi, e questo fatto è il principale argomento per considerare le due costruzioni come due costruzioni distinte, visto che, applicate alla stessa classe di verbi, danno strutture differenti. In realtà però l'intersezione nel dominio di applicazione delle due costruzioni è più ristretta: il *si* impersonale si può applicare ai verbi transitivi solo

se l'oggetto diretto è di 1. o 2. pers. (12a) o se è un pronome clitico di 3. pers. (12b), ma non se è un oggetto diretto di 3. pers. nominale o pronominale, ma non clitico (12c); il *si* passivo si può applicare ai verbi transitivi solo se l'oggetto diretto è di 3. pers., sia nominale che pronominale (13a,b), ma non nel caso di oggetti diretti di 1. o 2. pers. (13c) (possiamo assumere che il clitico corrispondente a un pronome soggetto in italiano sia il pronome nullo (\emptyset in (13b)), data la corrispondenza di valore semantico-pragmatico tra soggetto non espresso e pronomi clitici (cfr. Salvi & Vanelli 2004: 188-190); in (13b) dunque l'oggetto diretto clitico di *invitare* (*li*) è diventato soggetto clitico-nullo). L'intersezione è quindi limitata al caso di oggetti diretti clitici di 3. pers. (12b)/(13b):

- (12) a. Si invita *me* / *Mi* si invita
 b. *Li* si invita spesso
 c. *Si invita *i* *parenti* / solo *loro*
- (13) a. Si invitano *i* *parenti* / solo *loro*
 b. \emptyset si invitano anche per Pasqua
 c. **si*/*mi* invito *io*

In it. mod. il *si* che appare nella costruzione del *si* impersonale è un elemento diverso dal clitico riflessivo di 3. pers. *si*. I due morfemi vengono distinti in base a due proprietà sintattiche: a) quando cooccorre con un clitico accusativo di 3. pers., il *si* impersonale segue il clitico accusativo (14a), il clitico riflessivo invece lo precede (14b); b) il *si* impersonale cooccorre con il clitico *lo* che sostituisce un SA o un SN con funzione predicativa (15a), mentre questo non è possibile con il clitico riflessivo (né con nessun altro clitico–(15b)):

- (14) a. *Lo* *si* mangia
 b. *Se* *lo* mangia
- (15) a. Si sembra *imprudenti* / *Lo* si sembra
 b. Si è fatto *imprudente* / **Se* *lo* è fatto

Per quel che riguarda il *si* passivo, a causa delle sue proprietà non ricorre nelle due combinazioni che abbiamo appena trattato: a) essendo parte di una struttura passiva, non può cooccorrere con un clitico accusativo; b) il *lo* che sostituisce un SA/SN predicativo non può mai sostituire un predicato riferito a un oggetto diretto, neanche se l'oggetto diretto è diventato soggetto di una costruzione passiva (*Ritenevano Piero intelligente* / **Lo ritenevano Piero*, *Piero era ritenuto*

intelligente / **Piero lo era ritenuto*), non può quindi ricorrere neanche nella costruzione del *si* passivo. Non abbiamo dunque argomenti sintattici per classificare il morfema che compare nella costruzione del *si* passivo né con il *si* impersonale, né con il *si* riflessivo (ma in it. ant., dove, come vedremo (par. 2), abbiamo una sola costruzione, non abbiamo ragione di tenere distinto il morfema che entra nella costruzione del *si* passivo dal clitico riflessivo).

Nonostante in it. mod. esistano due morfemi *si* distinti, questi non possono essere combinati in un unico gruppo di clitici: il *si* riflessivo non può cooccorrere con il *si* impersonale (16a). Non è tuttavia escluso l'uso di verbi pronominali con il *si* impersonale, come abbiamo visto sopra in (9), solo che in questi casi al posto del clitico riflessivo di 3. pers. *si* nel gruppo di clitici deve comparire il clitico di 1. pers. pl. *ci* (16b). Questa soluzione è semanticamente giustificata dalla sovrapposizione funzionale tra pronomi generico e pronomi di 1. pers. pl., nei due sensi: la 1. pers. pl. può esprimere il valore generico (*Qui mangiamo molta polenta* = 'Qui la gente mangia molta polenta') e la costruzione con valore generico può servire a esprimere la 1. pers. pl. (francese *Nous on mange chez Mario*). Si noti in ogni caso che in altre varietà la combinazione dei due diversi *si* è ammessa, come per es. nel dialetto di Venezia (16c) (Lepschy 1989: cap. 9):

- (16) a. **Si si vede spesso*
 b. *Ci si vede spesso*
 c. *Veneziano*
 El se senta / Se se senta '(Egli) si siede / Ci si siede'

Alla luce della descrizione data in questo paragrafo possiamo vedere come in it. mod. esistano due costruzioni grammaticali distinte in cui la coniugazione pronominale serve alla demozione del soggetto lessicale; anzi, siccome il clitico che compare in una di queste costruzioni deve essere analizzato come un semplice omonimo del clitico riflessivo, la costruzione del *si* impersonale andrebbe forse piuttosto considerata una costruzione indipendente dalle altre strutture in cui si usa la coniugazione pronominale.

D'altra parte però la costruzione del *si* impersonale non può essere del tutto separata dalla costruzione del *si* passivo. In primo luogo il valore semantico delle due costruzioni coincide perfettamente: il soggetto

rimosso delle due costruzioni deve essere obbligatoriamente umano (mentre nella normale costruzione passiva perifrastica può anche essere non umano): sia in *Quando si distruggono i raccolti...* che in *Quando li si distrugge...*, l'agente non espresso può essere *l'esercito nemico*, ma non *la grandine* (mentre nel caso di *Quando vengono distrutti i raccolti...* non ci sono restrizioni). In secondo luogo la costruzione del *si* impersonale mostra delle limitazioni di applicazione che non sembrano giustificate dalla struttura generale della lingua: *i*) non è possibile con oggetti diretti di 3. pers. non clitici (12c), casi per cui si deve ricorrere alla costruzione del *si* passivo (13a); *ii*) il *si* impersonale non può cooccorrere con il *si* riflessivo: questa restrizione sarebbe naturale se si trattasse dello stesso morfema, perché in un gruppo di clitici ogni tipo può ricorrere una sola volta, ma abbiamo visto che in it. mod. i due *si* non possono essere considerati un'unica parola.

Queste "imperfezioni" nella grammatica della costruzione del *si* impersonale in it. mod. trovano una spiegazione, a livello diacronico, nel fatto che la costruzione è il frutto di una serie di innovazioni che hanno interessato la grammatica dell'italiano, che nella sua fase antica conosceva solo la costruzione del *si* passivo. Le innovazioni hanno allargato man mano l'ambito di applicazione della costruzione, ma questo non è avvenuto secondo un piano a largo raggio, ma attraverso innovazioni in gran parte locali, che hanno interessato di volta in volta solo parti della struttura, e in maniera non sempre coerente. La mancanza di un piano globale si riflette nel fatto che la costruzione del *si* impersonale non si è (per ora) completamente staccata da quella del *si* passivo e non ha sfruttato a pieno tutte le sue potenzialità.

2. Italiano antico

L'it. ant. disponeva di una costruzione del *si* passivo, perfettamente parallela alla costruzione passiva perifrastica (2.1), mentre quella del *si* impersonale non esisteva ancora (2.2).

2.1. *Si* passivo

In it. ant. la coniugazione pronominale poteva avere, alla 3. pers., valore passivo, come mostrano i seguenti esempi in cui il *si* passivo

alterna liberamente con la costruzione attiva (17a) o con la costruzione passiva perifrastica (17b), esprimendo lo stesso significato:

- (17) a. di ciò non *faccia* alcuno coscienza / non volemo che in dire questi pater nostri per alcuno *si faccia* coscienza (*Capitoli della Compagnia della Madonna d'Orsammichele* (1297), 57 e 66)
 b. Furono in questa concordia (...) che cierta quantità di pane e d'altre cose (...) *si dovessero vendere*, e *fuorono vendute* a cierte persone (*Ordinamenti della Compagnia di Santa Maria del Carmine*, 62.6)

Come nella normale costruzione passiva, il soggetto rimosso poteva essere espresso con un SP (*per alcuno* in (17a)).

La costruzione era possibile anche con verbi intransitivi inergativi (18). Esempi di questo tipo in it. mod. apparterrebbero alla costruzione del *si* impersonale (cfr. sopra (7)), ma in it. ant. i verbi inergativi potevano entrare anche nella normale costruzione passiva perifrastica (19) (mentre questo è impossibile in it. mod.), per cui esempi di questo tipo non contrastano con un'interpretazione passiva della costruzione e non possono essere considerati un argomento per l'esistenza della costruzione del *si* impersonale:

- (18) *Leggesi* della bontà del re giovane (*Novellino*, 18.3)
 (19) a. *fuè* del mese di settenbre *battalglato* fortemente con vij difici [7 macchine da guerra] (*Cronica fiorentina*, 125.6)
 b. ch'a llui *fosse dato* d'uno bastone (*Cronica fiorentina*, 118.33)

Gli esempi mostrano che l'it. ant. disponeva di una costruzione del *si* passivo, in gran parte coestensiva con la normale costruzione passiva perifrastica: come la costruzione passiva, si applicava ai verbi transitivi e inergativi e permetteva l'espressione del soggetto rimosso per mezzo di un SP. (In realtà la costruzione del *si* passivo aveva meno restrizioni della costruzione passiva, poiché si applicava (e si applica tuttora) anche a quei verbi stativi che non permettono la normale passivizzazione: *si hanno difficoltà* / **sono avute difficoltà*.)

2.2. *Si impersonale

In it. ant. la costruzione del *si* impersonale non esisteva: in tutti i casi in cui la coniugazione pronominale era usata per segnalare la demozione del soggetto lessicale, la costruzione aveva le caratteristiche tipiche di una costruzione passiva, in particolare: a) la demozione del

soggetto lessicale era accompagnata, se possibile, dalla promozione dell'oggetto diretto a soggetto, e *b*) si applicava solo ai verbi che ammettevano la costruzione passiva (2.1). L'applicazione di questi criteri ai dati dell'it. ant. non dà però apparentemente sempre risultati univoci, per cui è necessario considerare più nel dettaglio questi problemi.

a) Con i verbi transitivi la promozione dell'oggetto diretto a soggetto era obbligatoria, per cui ci aspettiamo di trovare dati paralleli a (13a,b) dell'it. mod., e di non trovare dati paralleli a quelli di (12). In effetti, abbiamo esempi come (17), paralleli a (13a), ed esempi come (20), paralleli a (13b), e non troviamo mai esempi del tipo di (12a,b). Troviamo, è vero, esempi come (21), che sembrano paralleli a (12c), ma in casi come questi il SN postverbale (*soldi XL* in (21)) non è l'oggetto diretto del verbo, nonostante la mancanza di accordo, ma il soggetto della frase: in it. ant. la mancanza di accordo era normale se il soggetto seguiva il verbo in una costruzione inaccusativa, come mostrano anche gli ess. (22) (che si tratti in effetti di soggetti, e non di oggetti diretti non promossi, è dimostrato dal fatto che in questi casi i SN postverbali non sono mai sostituibili con clitici accusativi):

- (20) l'astore [falco da caccia] (...) io consiglio che \emptyset non li [gli] si mandi.
(*Novellino*, 20.10-12)
- (21) s'ordinoe e fermò che per questa pasqua di Natale proxima, che ora dé venire, si dea soldi XL a' poveri, per l'amore di Dio, al modo usato
(*Ordinamenti della Compagnia di Santa Maria del Carmine*, 65.32)
- (22) a. In questo tempo, nel Mccx, si convertio alla fede di Cristo lxxm. uomini
(*Cronica fiorentina*, 115.20)
- b. Della buona volontà, di cui nasce le quattro virtù cardinali (Bono Giamboni, *Virtù e Vizi*, 2(titolo))
- c. sia inposto loro degnia penitentia (*Capitoli della Compagnia di S. Gilio*, 52.25)

b) La costruzione era possibile solo con i verbi transitivi e con gli inergativi, come la costruzione passiva perifrastica, mentre non si hanno esempi con inaccusativi, né semplici, come in (8), né pronominali, come in (9), e neppure con la copula, come in (10), o con la perifrasi passiva, come in (11). Questa situazione caratterizza il Duecento, ma verso la fine del secolo nelle opere di Dante la costruzione comincia a essere usata con gli inaccusativi semplici, in particolare con i verbi di movimento (ma non con la copula):

- (23) a. lo tenni li piedi in quella parte de la vita di là da la quale non *si puote ire* più per intendimento di ritornare (Dante, *Vita Nuova*, 14.8)
 b. non vuol che'n sua città per me *si vegna* (Dante, *Commedia*, 1.1.126)

Vorremmo suggerire che la struttura esemplificata in (23) sia un'innovazione introdotta da Dante stesso, probabilmente allo scopo di rendere possibile anche in italiano una costruzione che era normale in latino, dove la forma passiva poteva essere utilizzata anche con i verbi inaccusativi, con i quali assumeva un valore impersonale (*itur 'si va'*). In ogni caso la struttura utilizzata da Dante è chiaramente un caso di costruzione passiva, come mostra l'espressione dell'agente preposizionale in (23b) (*per me*). E un'analogia estensione avrà luogo anche nella costruzione passiva perifrastica durante il Trecento:

- (24) L'altra mattina seguente *fu andato* alla campana da casa Tornaquinci (F. Sacchetti, *Trecentonovelle*, 78.50)

Abbiamo qui, quindi, un'estensione del dominio di applicazione della costruzione del *si* passivo, non la nascita della costruzione del *si* impersonale. Un'estensione, vale la pena di aggiungere, piuttosto limitata: solo i verbi inaccusativi semplici, anzi probabilmente solo una sottoclasse di questi (presumibilmente quelli con soggetto agentivo), e soltanto nei tempi semplici (era possibile *Per me si viene*, ma non **Per me s'è venuto*). Si noti che un'estensione agli inaccusativi pronominali doveva essere formalmente impossibile, perché con questi verbi il clitico riflessivo era già presente con un'altra funzione. La mancata estensione ai tempi composti è più oscura, e sarà forse legata al fatto che in it. ant. la coniugazione pronominale non si era ancora del tutto affermata nei tempi composti (cioè a *egli s'inganna* corrispondeva *egli è ingannato*, piuttosto che *egli s'è ingannato*—cfr. Salvi in stampa); ed è forse da mettere in relazione con la restrizione attiva anche in it. mod. che proibisce l'uso del *si* impersonale con i verbi inaccusativi per descrivere eventi perfettivi non generici (*Oggi si è andati a pescare*, nelle varietà in cui è accettabile, non può significare 'Oggi qualcuno è andato a pescare', ma solo 'Oggi siamo andati a pescare', un'interpretazione su cui torneremo in 3.2; cfr. Cinque 1988).

3. Stadi intermedi

La formazione della costruzione del *si* impersonale è il frutto di una serie di innovazioni che si distribuiscono su un periodo di più secoli, a cominciare dalla fine del Duecento con l'innovazione che abbiamo discusso in 2.2 per arrivare alla seconda metà dell'Ottocento, in cui la costruzione ha ormai assunto la sua forma moderna. In realtà, inizialmente le innovazioni sono piuttosto rade e sembrano di tipo lessicale (v. 2.2 e 3.1), per accumularsi poi in un periodo relativamente breve tra la fine del Settecento e la metà dell'Ottocento, con cambiamenti strutturali più decisi (3.2-3). Benché il periodo delle innovazioni sia relativamente breve e sia quindi difficile dire quali fenomeni siano strettamente collegati e quali invece siano indipendenti, alcune linee di sviluppo possono essere individuate, anche grazie al confronto con altre varietà.

Come abbiamo premesso nelle note introduttive, il modo di raccolta dei materiali che abbiamo utilizzato per le fasi intermedie di sviluppo (e che abbraccia il periodo che va dal Cinquecento all'Ottocento), non permette una ricostruzione delle grammatiche sincroniche che si sono succedute nella storia di queste costruzioni. Ma naturalmente l'introduzione di una data innovazione può essere valutata in tutta la sua portata solo se inserita nel contesto della grammatica sincronica del suo periodo. Per fare un esempio molto semplice: il fatto che la costruzione con *si* si applicava ai verbi inergativi ha un significato diverso in un'epoca, come quella medievale, in cui anche la costruzione passiva perifrastica era applicabile a questi verbi, e in un'epoca più tarda, in cui la costruzione passiva fosse eventualmente limitata ai verbi transitivi-*se* per il primo periodo possiamo parlare di una costruzione del *si* passivo, questo non è più vero per il secondo periodo, dove la classe di verbi a cui si può applicare la costruzione del *si*, è molto più ampia di quella a cui si può applicare la costruzione passiva. Evidentemente la storia della costruzione del *si* impersonale non è fatta solo delle innovazioni positive nelle possibilità di questa costruzione, ma anche dei cambiamenti avvenuti in altre zone della grammatica. In questo lavoro, per le limitazioni dei materiali raccolti, non potremo tenere conto di questi aspetti. In particolare, non abbiamo potuto studiare un fenomeno che rappresenta un aspetto essenziale

della nostra ricostruzione, come si vedrà dalla discussione in 3.2-3, cioè l'uso della costruzione del *si* per esprimere la 1. pers. pl.

3.1. *Stadio I*

Le prime innovazioni nella grammatica della nostra costruzione appaiono nella prima metà del Seicento: nella *Istoria del Concilio Tridentino* di Paolo Sarpi (pubblicata nel 1619) e nel *Dialogo sopra i due massimi sistemi del mondo* di Galileo Galilei (pubblicato nel 1630) la costruzione del *si* ammette anche il verbo *essere* in frasi di localizzazione (25) (ma non ancora quando introduce SA o SN in funzione di predicato) e viene usata anche con i tempi composti dei verbi inaccusativi semplici (26) (in questi casi il participio, diversamente dall'it. mod., ha la forma sing. non marcata—v. sotto 3.2). Le due innovazioni sono forse collegate, poiché l'ausiliare dei tempi composti dei verbi inaccusativi è appunto *essere*:

(25) Luoghi possibili sono questi de' quali *si è in controversia* (G. Galilei, *Dialogo sopra i due massimi sistemi del mondo*)

(26) tuttavia *si è poi venuto* in certezza, tali mutazioni esservi a capello quali ricercava il sistema copernicano (G. Galilei, *Dialogo sopra i due massimi sistemi del mondo*)

La prima innovazione è un'estensione lessicale delle possibilità della costruzione: accanto ai verbi inaccusativi di movimento, appare anche il principale di verbi di stato, *essere*. La seconda innovazione colma apparentemente una lacuna nel paradigma: tutti i verbi che entrano nella costruzione possono ora comparirvi, senza restrizioni, anche nei tempi composti.

3.2 *Stadio IIa*

Un gruppo cospicuo di innovazioni compare in testi scritti da autori italiani settentrionali tra l'ultimo decennio del Settecento e i primi anni dell'Ottocento: nell'autobiografia di Vittorio Alfieri, scritta tra il 1790 e il 1803 (l'autore era nato nel 1749), e nel romanzo epistolare di Ugo Foscolo *Ultime lettere di Jacopo Ortis*, scritto tra il 1796 e il 1802 (l'autore era nato nel 1778).

La prima innovazione consiste nell'uso della nostra costruzione con le strutture copulative costituite da *essere* e un predicato agget-

tivale (27a). È interessante che in questo esempio la costruzione sia usata per esprimere la 1. pers. pl. (come mostra il contesto più ampio che abbiamo citato: *ci trovavano*, *ci era fatta la festa*). Esempi in cui la struttura abbia una interpretazione impersonale compaiono solo più tardi (27b) (in un'opera scritta nel 1821-23):

- (27) a. saliti poi innanzi ai birbi della Municipalità, *si era certi* di non poter più partire, d'andare anzi prigionieri, dove se ci trovavano nelle carceri il dí 2 settembre, cioè quindici giorni dopo, *ci era fatta la festa* insieme con tanti altri galantuomini che crudelmente vi furono trucidati (V. Alfieri, *Vita*, c. 22)
- b. quando *si è persuasi* d'una verità bisogna dirla (A. Manzoni, *Fermo e Lucia*, 1.3)

Riteniamo che si tratti di un fatto significativo, che ritroveremo almeno in un altro caso di innovazione (anche se non in tutti): i primi esempi di un'estensione nell'uso della nostra costruzione hanno il valore di 1. pers. pl., il significato impersonale compare solo successivamente. Possiamo pensare che questo sia un riflesso diretto del meccanismo che ha portato all'estensione nell'uso della costruzione: la costruzione del *si*, dato il suo valore generico, poteva essere usata per esprimere la 1. pers. pl.; a un certo punto ha cominciato a estendersi ad altre strutture in cui si poteva usare la normale forma verbale di 1. pers. pl.: nel nostro caso, in base all'equivalenza funzionale tra *siamo* e *si è*, *si è* ha cominciato a essere usata nel contesto di un predicato aggettivale (28):

- (28) Noi siamo certi -> Noi *si è* certi

Si noti che l'aggettivo è pl., come ci si aspetta nel caso di una 1. pers. pl. (mentre la forma verbale resta sing.), e questo è un fatto inatteso nel quadro della costruzione impersonale, come mostrano i casi dei tempi composti dei verbi inaccusativi, in cui un elemento aggettivale nel contesto del verbo *essere* appare al sing. (26) (e continua ad apparire nella forma sing. negli stessi autori che usano il pl. quando la costruzione del *si* ha valore di 1. pers. pl.–v. subito sotto). L'uso di queste strutture con valore impersonale sembra in un certo senso parassitario al loro uso come 1. pers. pl.: solo dopo che la

struttura *si* è aperta una breccia con un'estensione analogica nell'uso particolare della 1. pers. pl., *si* stabilizza in tutti gli usi.

Come abbiamo accennato, nei tempi composti dei verbi inaccusativi gli stessi autori continuano ad usare il participio al sing. nella costruzione con valore impersonale (29a). Ma anche in questo caso assistiamo a un fenomeno simile a quello precedente: se la costruzione è usata con valore di 1. pers. pl., compare il pl. (29b). E di nuovo, questa nuova struttura comparirà con valore impersonale solo molto più tardi (29c) (l'opera è del 1881, ma le condizioni d'uso della costruzione del *si* impersonale con i tempi composti dei verbi inaccusativi e con valore generico sono molto ristrette, per cui il ritardo nella documentazione è senz'altro determinato dalla rarità della struttura):

- (29) a. allorché finalmente *si* è venuto ad offenderli involontariamente (V. Alfieri, *Vita*, c. 13)
- b. Sei o sette giorni addietro *s'è iti* in pellegrinaggio. Io ho veduto la Natura più bella che mai. Teresa, suo padre, Odoardo, la piccola Isabellina ed io siamo andati a visitare la casa del Petrarca in Arquà. (U. Foscolo, *Ultime lettere di Jacopo Ortis*, parte I, lett. 11)
- c. Padron 'Ntoni, e Alessi, e Mena, tutto quello che buscavano colla pesca, col telaio, al lavatoio, e con tutti gli altri mestieri, potevano metterlo da parte, per quella famosa barca di san Pietro, colla quale *si* guadagnava di rompersi le braccia tutti i giorni per un rotolo di pesce, o per la casa del nespolo, nella quale *si sarebbe andati* a crepare allegramente di fame! tanto lui un soldo non l'avrebbe voluto; povero diavolo per povero diavolo, preferiva godersi un po' di riposo, finché era giovane, e non abbaiva la notte come il nonno. (G. Verga, *I Malavoglia*, c. 13)

La struttura esemplificata in (29b) sarà nata nelle stesse condizioni di quella esemplificata in (27a), in base all'equivalenza funzionale tra *siamo* e *si* è, secondo lo schema (30):

- (30) Noi siamo andati -> Noi *si* è andati

Una terza innovazione è l'uso di clitici accusativi di 1. (e, dobbiamo sopporre, di 2.) pers. nella costruzione del *si* (31), con valore impersonale:

- (31) Alcuni giorni dopo egli mandò ambasciata a voce, per sapere in che ore *mi si potrebbe trovare* (V. Alfieri, *Vita*, c. 12)

Qui non disponiamo di un numero sufficiente di esempi per documentare quello che potrebbe essere stato il meccanismo dell'innovazione, ma si potrebbe supporre anche qui che l'uso sia nato con l'interpretazione di 1. pers. pl., secondo lo schema in (32):

(32) Noi ti/vi diciamo -> Noi ti/vi si dice, Noi ti/vi vediamo -> Noi ti/vi si vede

Si noti però che questa analogia spiega in maniera diretta solo la combinazione con i clitici di 2. pers. (data l'agrammaticalità di **Noi mi diciamo*, **Noi mi vediamo*), per cui bisognerebbe supporre che la costruzione nasce con i clitici di 2. pers. (di cui non abbiamo trovato esempi a questa data, ma che dovevano sicuramente esserci), si estende prima all'uso impersonale e in seguito, all'interno dell'uso impersonale, si estenda anche ai clitici di 1. pers. Il fatto però che questa innovazione sia presente già così presto con il valore impersonale, può far dubitare che sia del tutto parallela alle due discusse più sopra (ma l'estrema scarsità di esempi raccolti non ci permette di fare molto affidamento su questa considerazione cronologica).

In ogni caso, nella costruzione del *si* fin dall'inizio si potevano usare clitici dativi non riflessivi (33) e, siccome i clitici di 1. e 2. pers. hanno un'unica forma per l'accusativo e per il dativo, questa combinazione può aver funzionato da modello formale per l'utilizzazione degli stessi clitici con funzione di accusativo. La presenza di un modello formale deve essere ritenuta essenziale, perché l'assenza di un modello simile per i clitici accusativi di 3. pers. ha impedito il loro uso in questa costruzione in altre varietà italiane (v. sotto par. 4), e anche nella varietà qui studiata l'introduzione di questi clitici nella costruzione del *si* ha seguito vie diverse (v. sotto):

(33) E nel forbire che fece, parve che degli occhi *mi si levasse* una crosta di sozzura puzzolente di cose terrene (Bono Giamboni, *Vizi e Virtudi*, 3.2)

Un fattore concomitante (nel caso delle prime due innovazioni) o essenziale (nel caso della terza) può essere stato il desiderio di rendere la costruzione impersonale francese con *on*, che, dato il suo carattere attivo, non presentava nessuna delle restrizioni tipiche della costruzione italiana del *si*, originariamente passiva. Il modello in questo caso avrebbe agito nel modo seguente: se al francese *On*

me dit corrisponde l'italiano *Mi si dice* (il tipo esemplificato in (33)), al francese *On me cherche* corrisponderà l'italiano *Mi si cerca*.

L'italiano non presentava invece nessun modello accettabile per il francese *On le dit*, per cui l'ultima delle innovazioni di questo periodo, l'uso nella costruzione del *si* impersonale dei clitici accusativi di 3. pers., avrà un'origine diversa. Questa fa la sua apparizione in lettere private all'inizio del secolo (34a-b) (Foscolo era nato nel 1778, Peticari nel 1779), mentre compare solo più tardi in opere letterarie (34c) (il romanzo è stato pubblicato nel 1845-46):

- (34) a. Ieri si credeva moribondo il generale Championnet, oggi *lo si predica* morto. (U. Foscolo, *Lettera del 10 gennaio 1800*)
 b. O si vuole immaginare il poeta coll'amante viva, o coll'amante morta. Se *lo si dipinge* mentr'era viva, è bisogno il mostrare due giovinetti: perché la Bice morì, che Dante aveva soli 24 anni. (G. Peticari, *Lettera del 19 ottobre 1820*)
 c. Così, cominciandosi a gridare – Viva il Palavicino – da quel gruppo d'uomini che gli si era serrato intorno, poco a poco il suo nome passò per tutte le bocche che gridavano in piazza, e, prese un così esteso giro, che *lo si udì* gridato in via Santa Margherita, e giù giù sino al Portone. (G. Rovani, *Manfredo Pallavicino*, c. 5)

La nuova struttura è frutto di un ipercorrettismo basato sull'uso toscano (colloquiale), dove il soggetto della costruzione del *si* passivo poteva essere espresso anche da un pronome clitico (in alcuni casi avrà contribuito anche il sostrato dialettale degli scrittori settentrionali, dove si trovavano le stesse strutture che nel toscano colloquiale). Nel caso della 3. pers. sing. femm. questo clitico era *la* (corrispondente alla forma libera *ella*), e precedeva il clitico riflessivo (35a). La forma corrispondente masch. era *e'* (35b), mentre con le forme pl. il verbo era naturalmente accordato al pl. (35c):

- (35) a. *La si* invita spesso
 b. *E' si* invita spesso
 c. *E' / Le si* invitano spesso

Ma la forma *la* che appare in (35a) era omofona con la forma del clitico accusativo di 3. pers. femm., e i parlanti non toscani, che probabilmente usavano già i clitici accusativi di 1. e 2. pers. nella costruzione del *si*, hanno interpretato il clitico nominativo di

(35a) come un clitico accusativo e ne hanno esteso l'uso alla forma masch., creando (36a), e alle forme pl., creando (36b) (con il verbo coerentemente al sing.):

- (36) a. *Lo si* invita spesso
 b. *Li / Le si* invita spesso

La caratteristica saliente di questa nuova struttura è che il clitico accusativo precede *si*, invece di seguirlo, come nelle combinazioni tradizionali di *si* con i clitici accusativi di 3. pers. (cfr. (14b)). Questo garantisce anche la correttezza della spiegazione: nella sequenza *la si* il clitico *la* precede *si* perché occupa la posizione dei clitici soggetto, non quella dei clitici accusativi di 3. pers. Si noti inoltre che la rianalisi ha avuto anche altre conseguenze strutturali. I clitici soggetto occupano una posizione strutturale diversa nella frase rispetto ai clitici obliqui: potevano per es. essere separati dagli altri clitici dalla negazione: *la non s'incomodi* (G. Giusti, *I discorsi che corrono*, atto 2, scena 5), ma dopo la rianalisi entrano a far parte a pieno diritto del gruppo dei clitici obliqui e seguono quindi la negazione: *non la si vede*. Assumono anzi, all'interno del gruppo, una posizione diversa da quella iniziale che ci si potrebbe aspettare: *me lo si diceva, glielo si dava*, ecc., probabilmente come risultato dell'azione incrociata dei modelli minimi preesistenti: *me lo + lo si > me lo si*, ecc. Il clitico accusativo di 3. pers. continua cioè a occupare la stessa posizione che occupava prima rispetto a tutti gli altri clitici, ma in più può essere seguito da *si* con valore impersonale.

3.3 *Stadio IIb*

Altre innovazioni fanno la loro comparsa in opere scritte intorno alla metà dell'Ottocento: nel saggio *Delle speranze d'Italia* di Cesare Balbo, pubblicato nel 1844 (l'autore era nato nel 1789), nei romanzi *Manfredo Pallavicino* (1845-46) e *Cento anni* (1856-65) di Giuseppe Rovani (l'autore era nato nel 1818) e nel romanzo *Le confessioni di un italiano*, scritto nel 1857-58, di Ippolito Nievo (l'autore era nato nel 1831).

Compagnano nella costruzione del *si* i verbi riflessivi, con la caratteristica combinazione *ci si*:

- (37) la guerra è come un giuoco, che quanto è più forte la perdita, tanto più *ci si ostina*, e si continua finchè ci è vita (G. Rovani, *Manfredo Pallavicino*, c. 3)

Nonostante non disponiamo in questo caso di esempi sufficienti e l'esempio più antico tra quelli raccolti presenti già l'interpretazione impersonale, la presenza del clitico *ci* di 1. pers. pl. rende probabile anche qui che il punto di partenza sia stato l'uso della costruzione del *si* con valore di 1. pers. pl., secondo lo schema (38), in base all'equivalenza *laviamo = si lava*:

- (38) Noi laviamo / Noi ci laviamo -> Noi si lava / Noi ci si lava

La combinazione di clitici *ci si* esisteva del resto già fin dalle origini con *ci* non riflessivo in funzione di dativo (39a) e si era venuta recentemente imponendo (v. 3.2) con *ci* non riflessivo in funzione di accusativo (39b). Siccome nei clitici di 1. e 2. pers. si neutralizza la distinzione tra forme riflessive e non riflessive, queste combinazioni potevano fungere da modello formale per l'introduzione del tipo esemplificato in (37)::

- (39) a. quello siamo tenuti di volere più sollicitamente, lo quale non *ci si dà* per lo gran pericolo che nn'è (*De Amore di Andrea Cappellano volgarizzato*, 1.18)
 b. Ma la funzione è protratta... e noi *ci si manda* a Malamocco (G. Rovani, *Valenzia Candiano*, c. 12)

Una seconda innovazione consiste nell'applicazione della costruzione del *si* alla costruzione passiva perifrastica:

- (40) a. galantuomini tutti ai quali è assai ben nota la parrocchia ove *si è stati battezzati*, ma se in sajo o in cappa ci corrà la morte, e quale de' quattro venti si porterà la polvere de' nostri dieci carcami, è quanto sta ancora nascosto in un fitto bujo... (G. Rovani, *Manfredo Pallavicino*, c. 10)
 b. non si scrive egli e dipinge, e promuove le culture tutte, e non si governa egli e non *si è governati*, per viver buoni, per la virtù? (C. Balbo, *Delle speranze d'Italia*, c. 12)

Nonostante l'uso di 1. pers. pl. (40a) e quello impersonale (40b) facciano la loro comparsa contemporaneamente, anche qui si può pensare all'analogia della 1. pers. pl., secondo lo schema (41), anche perché il participio è accordato al pl. Esisteva inoltre già il modello delle frasi copulative con un predicato aggettivale (v. 3.2):

(41) Noi vediamo / Noi siamo visti -> Noi si vede / Noi si è visti

Possiamo pensare che l'attestazione leggermente più tarda di queste due strutture sia casuale. In primo luogo, esse si inseriscono perfettamente nel gruppo di casi, trattati in 3.2, in cui l'uso di 1. pers. pl. sembra aver avuto un ruolo essenziale nello sviluppo delle strutture innovative. In secondo luogo, mentre nei casi visti in 3.2 la struttura con interpretazione di 1. pers. pl. appariva prima nella documentazione di quella con interpretazione impersonale, nei casi trattati qui, tra i pochi esempi rilevati, o questo divario non c'è, o non abbiamo trovato esempi con interpretazione di 1. pers. pl., e questo può far sospettare che anche in questo caso l'uso di 1. pers. pl. abbia preceduto quello impersonale, ma (per ora) semplicemente non ne abbiamo trovato esempi. In terzo luogo, se consideriamo la generazione a cui appartiene lo scrittore, e non la data dell'esempio, almeno il caso del passivo può essere raggruppato con quelli del paragrafo precedente, visto che Balbo (40b) appartiene piuttosto alla generazione di Foscolo che non a quella di Rovani.

Se consideriamo dunque casuale questo ritardo, possiamo pensare che le innovazioni che toccano la costruzione del *si* riflettano un fenomeno che si è svolto essenzialmente nella seconda metà del Settecento.

4. Conclusioni

La ricostruzione che abbiamo tentato nel par. 3 della formazione della costruzione del *si* impersonale, deve considerarsi sotto molti aspetti provvisoria: da una parte, i dati relativi al funzionamento della costruzione che abbiamo raccolto, sono molto parziali (come abbiamo messo più volte in luce); dall'altra, non abbiamo preso in considerazione vari aspetti "esterni" di questo processo che possiamo supporre siano stati rilevanti (per es. il sostrato dialettale degli scrittori nelle cui opere si manifestano le innovazioni; l'uso di varietà italiane diverse da quelle innovative, soprattutto di quella toscana, che doveva agire da modello per l'uso di altre regioni; la possibile interazione con altre lingue, in particolare il francese; ecc.).

Nonostante questo, il quadro che si è delineato è abbastanza articolato e molti suoi aspetti ci sembrano abbastanza sicuri per tentare di trarne alcune considerazioni generali relative al funzionamento del cambiamento diacronico (o al funzionamento delle lingue in generale).

È senz'altro possibile che le caratteristiche del cambiamento studiato siano più o meno strettamente collegate alle condizioni particolari in cui questo cambiamento è avvenuto: si tratta di un cambiamento avvenuto nella lingua letteraria di utenti che non avevano questa lingua come lingua materna (questo non è il caso naturalmente di Dante e Galileo). Il fatto stesso, però, che i cambiamenti siano avvenuti nonostante i forti impedimenti posti dalla tradizione, qualifica il processo come un processo diacronico spontaneo (e il risultato finale del cambiamento, nonostante le sue imperfezioni, è oggi parte integrante della lingua madre di una fetta considerevole della comunità linguistica italiana). Quanto al plurilinguismo degli utenti, questa deve essere considerata la situazione normale delle comunità linguistiche, anche nel senso che i parlanti possono contribuire al cambiamento di una varietà che non è la loro varietà primaria. Tutto questo naturalmente non significa che un cambiamento diacronico non possa seguire anche vie diverse—ci sono però cambiamenti che hanno le caratteristiche descritte qui.

Prima di tutto si può notare la *manca*za di un piano coordinato nello sviluppo della costruzione, nonostante il risultato finale sia abbastanza coerente: oggi abbiamo una costruzione che si applica a tutti i verbi dell'italiano che dispongano di un soggetto lessicale, ma l'estensione a partire dal gruppo iniziale più ristretto è avvenuta in tappe separate, per buona parte indipendenti l'una dall'altra. L'indipendenza dei singoli cambiamenti è provata dal fatto che in altre varietà eventualmente solo una parte di questi cambiamenti parziali ha avuto luogo: per es., nel dialetto di Firenze (Stefanini 1983), semplificando un po', la costruzione con interpretazione impersonale è allo stadio che aveva ai tempi di Galileo; con interpretazione di 1. pers. pl., la costruzione si è ampliata ulteriormente, ma neanche così ammette i clitici accusativi di 3. pers. Inoltre, come abbiamo notato (1.2), non essendo il risultato di un piano coerente, la costruzione del *si* impersonale continua a mostrare delle imperfezioni.

Le innovazioni sono in genere *locali*: non riguardano strutture astratte di portata generale, ma punti o tratti particolari della costruzione. In parte si tratta di estensioni *lessicali*, basate su comunità di tratti semantici (verbi inergativi (di azione) -> verbi inaccusativi di moto (di azione) [2.2]; verbi inaccusativi di moto -> verbi inaccusativi di stato [3.1]); in parte di analogie basate su elementi singoli (sostituibilità reciproca tra *siamo* e *si* è [3.2]; reinterpretazione ed estensione della combinazione *la si* [3.2]).

L'ambito delle singole innovazioni è fortemente condizionato dall'esistenza di *modelli formali* preesistenti (nasce *mi si* con *mi* accusativo perché c'era già *mi si* con *mi* dativo, ma non nasce *se lo 'lo si'* per mancanza di modelli adeguati [3.2]).

D'altra parte, l'analisi delle strutture che sta alla base delle innovazioni è *superficiale*, non fa riferimento a un sistema di regole (*la si* è analizzato come gruppo di clitici obliqui nonostante la presenza di *la non si* [3.2]), e per questo può creare strutture che non hanno modelli formali preesistenti.

Le innovazioni sono *miopi*: basandosi su modelli strettamente locali, possono creare strutture che sono disarmoniche rispetto alle strutture esistenti (la rianalisi di *la si* introduce una combinazione di clitici in contrasto con quella preesistente *se la* [3.2]; l'utilizzazione del tipo *si è vivi* con valore impersonale contrasta con il tipo, sempre impersonale, *si è giunto* (usati ambedue, per es., da Rovani) [3.2]).

È anche vero che in alcuni casi i cambiamenti possono tendere alla regolarizzazione di una asimmetria (l'uso dei tempi composti dei verbi inaccusativi, in parallelo alle altre classi di verbi [3.1]; in genere tutto l'uso della costruzione per la 1. pers. pl., che porta a estenderla a una buona parte dei casi in cui si può usare un verbo flesso alla 1. pers. pl.) o possono avere come conseguenza la creazione di una nuova regola generale: l'uso del pl. nei complementi predicativi, per es., diventa un segno distintivo dell'interpretazione indeterminata/generica, e dalla costruzione del *si si* si diffonde ad altre strutture, per es. al caso degli infiniti con soggetto non espresso generico (42a) o al caso dei complementi predicativi riferiti a un oggetto diretto non espresso generico (42b)—in it. ant. in questi casi si usava il sing. (43):

- (42) a. Per salvarsi, bisogna essere *umili*
 b. La sua vista rende *beati*
- (43) a. se l'uomo si vuole partire dal peccato (...), in prima conviene essere *umile* (...); poi conviene essere *sollicito*, e non *negligente* (*Ottimo Commento della Commedia, Purgatorio*, c. 2, p. 30, rr. 4-8)
 b. *faria beato* pur descripto [renderebbe *beati* anche solo a descriverlo] (Dante, *Purgatorio*, 2.44)

Per passare agli aspetti esterni del cambiamento, possiamo notare l'importanza fondamentale del *contatto linguistico*, nel caso specifico, trattandosi di lingua letteraria, l'interazione con altre lingue letterarie usate normalmente dagli utenti in alternativa con l'italiano (latino nel caso di Dante [2.2], francese nel caso di Alfieri [3.2] e Balbo; la conoscenza di una varietà toscana colloquiale o di una varietà settentrionale nel caso del tipo *la si* [3.2]). L'innovazione può essere attivata dal desiderio conscio o semi-conscio di esprimere con i mezzi di cui dispone l'italiano tutto quello che un'altra lingua riesce a esprimere con una costruzione diversa, ma in parte parallela; questo porta in genere a modificazioni che rispettano le restrizioni formali valide nella lingua (l'estensione ai verbi inaccusativi si limita agli inaccusativi semplici perché con i pronominali il clitico riflessivo era già presente con altra funzione [2.2]; l'introduzione di *si viene* non è accompagnata da quella di *si è venuto* perché la coniugazione pronominale non aveva una forma distinta per i tempi composti [2.2]; l'introduzione di *mi si vede* (su *mi si dice*) non è accompagnata da quella di **se lo vede*, per la quale non esisteva un modello formale [3.2]; la costruzione usata originariamente per la 1. pers. pl. viene utilizzata nella sua interezza, con tutte le sue proprietà formali, per il valore impersonale [3.2-3]). Se invece l'interferenza avviene in maniera inconscia, il risultato può violare le strutture preesistenti (rianalisi di *la si* [3.2]).

Come si vede, i cambiamenti possono essere dovuti a cause esterne ("casuali") oppure motivati dal sistema (correttivi e generalizzanti). Nel caso preso in esame, nessuno dei cambiamenti avvenuti è di portata molto generale—anche il cambiamento che ha avuto maggiori effetti, l'estensione basata sul significato di 1. pers. pl., non è un cambiamento nato per la costruzione impersonale, la costruzione impersonale lo ha, in un certo senso, ereditato (si noti che per molti

dei parlanti che usano la costruzione impersonale descritta in 1.2, il significato di 1. pers. pl. non è un'opzione accettabile: questo uso ha fatto da tramite per l'estensione delle possibilità strutturali della costruzione e poi è sparito).

Il risultato finale è imperfetto (1.2), e lo è per il modo in cui si è formato. Più in generale, ogni sistema linguistico contiene delle imperfezioni che sono il frutto di innovazioni locali non perfettamente inserite nel sistema e allo stesso tempo sono potenzialmente il motore di innovazioni che cercano di eliminare queste imperfezioni e rendere il sistema più coerente. L'imperfezione inerente al sistema è, assieme alla sua arbitrarietà, la condizione essenziale per il cambiamento linguistico: se il sistema fosse perfetto, non ci sarebbe ragione perché cambi in direzione di una maggior sistematicità; se non potesse essere altro che perfetto, non potrebbe accogliere quei cambiamenti "casuali" che ne rovinano la perfezione.

Bibliografia

- Cinque, G. 1988. On *si* constructions and the Theory of Arb. *Linguistic Inquiry*. **19**: 521-581.
- Lepschy, G. C. 1989. *Nuovi saggi di linguistica italiana*. Bologna: Il Mulino.
- Renzi, L. 2000. 'ItalAnt': come e perché una grammatica dell'italiano antico. *Lingua e stile*. **35**: 717-729.
- Salvi, G. 2006. La realizzazione sintattica della struttura argomentale. In: G. Salvi; L. Renzi (a cura di). *Grammatica dell'italiano antico*. In rete: <http://geocities.com/gpsalvi/>.
- Salvi, G. 2008. Imperfect systems and diachronic change. In: U. Detges; R. Waltereit (Eds.). *The Paradox of Grammatical Change. Perspectives from Romance*. Amsterdam: John Benjamins, 127-145.
- Salvi, G. In stampa. Morphosyntactic Persistence from Latin into Romance. In: A. Ledgeway; M. Maiden; J. Ch. Smith (Eds.). *The Cambridge History of the Romance Languages*. Cambridge: Cambridge University Press.
- Salvi, G.; Vanelli, L. 2004. *Nuova grammatica italiana*. Bologna: Il Mulino.
- Stefanini, R. 1983. Riflessivo, impersonale e passivo in italiano e in fiorentino. *Quaderni dell'Atlante Lessicale Toscano*. **1**:103-114.
- Wehr, B. 1995. *SE-Diathese im Italienischen*. Tübingen: Narr.

Teorias da mudança lingüística e a sua relação com a(s) história(s) da(s) língua(s)

Rosa Virgínia Mattos e Silva
mattosesilva@hotmail.com

Universidade Federal da Bahia, Instituto de Letras/CNPq (Brasil)

Introdução: o problema

Desde que iniciamos o nosso grupo de pesquisa, “Programa para a história da língua portuguesa”, nos inícios da década de noventa do século passado, uma questão se colocou para o grupo, no momento em que deveria se apresentar ao CNPq, a fim de solicitar um Projeto Integrado. A questão era a seguinte: que teoria adotar se já havia no grupo descritivistas, variacionistas e gerativistas. Com o correr do tempo, incluíram-se funcionalistas. Que decisão tomar quanto a essa questão? Decidimos, coletivamente, que não trabalharíamos para a reconstrução histórica da língua portuguesa com uma única teoria e seu respectivo método. Cada pesquisa/projeto dos componentes do grupo explicitaria a sua teoria e o seu método, partindo sempre, contudo, de uma base descritiva. Assim se fez! E se faz!

Em 1996, Ataliba Teixeira de Castilho, começou a elaborar o “Projeto para a história do português brasileiro”, PHPB.

Então voltou a ser colocado o mesmo problema ou questão: o da teoria e seu método. A solução que vem sendo seguida é a que adotamos em nosso grupo, o PROHPOR: há nas equipes regionais que compõem o Projeto descritivistas, gerativistas, variacionistas, funcionalistas e, mais recentemente, os que pesquisam sobre “Tradições discursivas”, de orientação alemã. Basta que se examinem os livros publicados, a partir dos seminários bianuais do PHPB, ao todo sete (São Paulo, Campos do Jordão, Campinas, Teresópolis, Ouro Preto, Salvador/Ilha de Itaparica e Londrina, este último a ser publicado), para que se verifique o que acima afirmamos.

Para estudar o passado de qualquer língua, faz-se necessário o uso de textos bem editados, uma vez que se trata de edições para estudos lingüísticos de tempos pretéritos. Assim a Filologia se fez presente no PHPB: no Seminário de Campos do Jordão, um grupo,

coordenado por Heitor Megale, estabeleceu as normas para as edições a serem utilizadas no *corpus* documental que seria a base das análises lingüísticas. Começou-se com textos impressos no século XIX, sobretudo os jornais que, no Brasil, se multiplicaram naquele século, uma vez que, só a partir de 1808, foi permitida pela metrópole a imprensa no Brasil. Dessa pesquisa em jornais, publicou-se o livro *E os preços eram commodos... Anúncios de jornais brasileiros. Século XIX*, sob a responsabilidade de Marymárcia Guedes e Rosane de Andrade Berlinck, corria o ano 2000. Depois foi-se para os Arquivos Brasileiros, públicos ou privados. A equipe da Bahia publicou, também em 2001, *Cartas Baianas Trecentistas*, sob a responsabilidade da pesquisadora Tânia Lobo e com a colaboração dos bolsistas de Iniciação Científica – Klebson Oliveira, Permínio Ferreira e Uilton Oliveira. Outros *corpora* vêm sendo constituídos, como base, principalmente, de dissertações e teses, como, por exemplo – *As cartas do Século XIX* para as freiras do Desterro, base da tese de Tânia Lobo (inédita), em que a autora distingue as cartas de brasileiros das dos portugueses; as *Cartas* de Afrânio Gonçalves de comerciantes portugueses no Brasil, que também editou outros documentos para confrontar com as *Cartas*; as atas e outros documentos da Sociedade Protetora dos Desvalidos editadas por Klebson Oliveira, editadas no seu Mestrado, e ampliadas, em número, na sua tese de Doutorado e assim por diante. Verificou-se, no correr da pesquisa, que era o século XIX o mais fértil nos Arquivos Brasileiros, diminuindo a incidência de documentos quando se recua no tempo. Como formigas obreiras, tem-se pesquisado Arquivos em vários pontos do Brasil.

Para encerrar esse item, resalto a importância da Filologia, no sentido estrito, de edições de textos confiáveis, para estudos lingüísticos do passado.

1. O que é a mudança lingüística? Quais as teorias contemporâneas sobre a mudança das línguas no tempo?

A mudança das línguas no tempo é o cerne da Lingüística Histórica. Já que os principais testemunhos para o passado lingüístico são os textos escritos – inscrições, manuscritos, textos impressos –, são apropriadas metáforas para definir a Lingüística Histórica, como a de

Roger Lass, “Ouvir o inaudível” (1997: p. 45), e a mais conhecida, de William Labov, “a arte de fazer o melhor uso dos maus dados” (1982: p. 20). Metáforas que, em parte, podem delimitar o que seja o trabalho nesse campo da Lingüística.

Ao longo de seu tempo histórico, as línguas mudam: há mudanças fônicas, mórficas, sintáticas e léxico-semânticas. Contudo, a história de uma língua, como a história dos homens, como disse Michel Foucault, “não é uma duração: é uma multiplicidade de tempos que se emaranham e se envolvem uns nos outros” (2000 [1972]: p. 293).

Se assim é, a linearidade temporal nas línguas deve ser revista e a “multiplicidade de tempos que se emaranham” deve ser levada em conta por aquele que faz Lingüística Histórica. Uma lei do tipo neogramático, em que X (no tempo A) $>$ Y (no tempo B), não se sustenta, porque entre A e B múltiplos fatores condicionantes podem ter efeitos inesperados sobre a mudança de $X > Y$.

Pode-se conceber a mudança lingüística no sentido estrito e no sentido lato. No primeiro caso, pode ser trabalhada em duas orientações: a) a Lingüística Histórica sócio-histórica; b) a Lingüística Diacrônica ou a-social. Quanto à Lingüística Histórica no sentido lato, trabalha com dados datados e localizados, como qualquer Lingüística que trabalhe com *corpus*, como a Dialectologia e a Sociolingüística laboviana, a Etnolingüística e mesmo a teoria da conversação, desde que use *corpora*. Inspirei-me para essa dicotomia – sentido lato e estrito – em Eugênio Coseriu que, em *Sincronia, Diacronia e História* diz que a “descrição e a história da língua situam-se no nível histórico da linguagem e constituem juntas a Lingüística Histórica” (1979: p. 236) e no clássico *Empirical Foundations for a Theory of Language Change* (1968), em que Uriel Weinreich, estudioso do contacto lingüístico, Marvin Herzog, dialetólogo, e William Labov se uniram para elaborar os famosos *Fundamentos Empíricos para uma Teoria da Mudança Lingüística*.

De todos os tipos de mudanças as mais estudadas foram as fônicas com os neo-gramáticos do século XIX, cujo texto teórico básico é o de Hermann Paul, *Prinzipien der Sprachgesichte*, publicado em 1880. Os conceitos básicos dessa teoria são as leis fonéticas e a analogia. Estabeleceram-se leis fonéticas como as de Verner, precedida da lei de Grimm. Sem dúvida, o postulado das leis fonéticas estabeleceu

um rigor nos estudos históricos. O primeiro teórico da analogia foi Wilhelm Scherer. Seria a analogia a única razão para impedir as leis fonéticas. Segundo Mattoso Câmara Jr.

“A mente humana, associando formas distintas por seus significados ou semelhanças de sons, foi vista como capaz de interferir no desenvolvimento natural de sons, contrariando a esmagadora força de uma lei fonética.” (1990 [1975]: p. 76).

Na realidade, é Hugo Schuchardt que abre caminhos para uma nova orientação – o estudo da variação lingüística no espaço e sua compreensão para as mudanças no tempo; além disso, é o mesmo Schuchardt que vai introduzir nas preocupações dos oitocentistas o fenômeno do contacto entre línguas. Seu livro seminal é publicado em 2ª edição em 1928, com o título de *Brevier: Ein Vademecum der allgemeinen Sprachwissenschaft*.

No estruturalismo diacrônico ainda são os fatos fônicos o centro. Iniciou-se com o Círculo Lingüístico de Praga e a primeira proposta foi apresentada por Roman Jakobson em 1931, no livro *Prinzipien der Historischen Phonologie*. Contudo, é André Martinet a figura estelar da fonologia diacrônica. Desenvolveu uma teoria mais abrangente que a de Jakobson e, em 1955, publica o *Économie des Changements Phonétiques*. Parte Martinet do “ponto fraco”, que relaciona com os conceitos de “função”, “estrutura” e “economia”. A “função” do fonema é distinguir signos. Associado a isso, explora a noção de “rendimento funcional”, em relação direta com a maior ou menor estabilidade do fonema, que está em relação direta com a maior ou menor estabilidade do fonema no sistema: quanto mais integrado o fonema nesses feixes de correlações e pares opositivos, mais estável é o fonema. Retoma ainda Martinet a antiga noção de “menor esforço”, reinterpretando-a como “economia”. São esses, para Martinet, fatores internos fundamentais para a compreensão da mudança fônica.

Dos estruturalistas americanos, o que mais se preocupou com a mudança foi Edward Sapir que desenvolve a teoria da “deriva” (‘drift’) no seu livro de 1921: *Language: an Introduction to the Study of Speech*. Contudo, essa teoria, tal como o estruturalismo diacrônico europeu, é criticada por sua natureza teleológica, ou seja: ao obedecer uma direção terá a mudança um fim previsível.

No gerativismo diacrônico, que se inicia com publicações como *Syntactic Structures* de Noam Chomsky (1957) e *The Sound Pattern of English* de Noam Chomsky e Morris Halle (1968), são lançados alguns pontos de reflexão sobre a mudança, centrados, sobretudo, no processo de transmissão da língua de geração para geração. A mudança é então entendida como mudança de regras na gramática.

Tanto os neogramáticos como os estruturalistas diacrônicos e Chomsky/Halle vêem a mudança como intrassistêmica. Contudo, ainda na Teoria da Variação e Mudança, iniciada na segunda metade do século XX, que é uma teoria extrassistêmica, o ponto forte é a mudança fônica. Com isso quero dizer que, desde a segunda metade do século XIX, é a mudança fônica o centro das teorias sobre a mudança lingüística. Só em 1979, com David Lightfoot, no livro *Principles of Diachronic Syntax*, é que o foco se desloca para a sintaxe, para a mudança sintática. Na década de oitenta, voltando Fernando Tarallo de seu doutoramento na Pensylvania, juntou-se a Mary Kato e construíram o que veio a ser designado de “Sociolingüística Paramétrica”. Em 1993, já em homenagem a Tarallo, falecido precocemente, publicou-se o livro *Português Brasileiro: uma viagem diacrônica*, organizado por Mary Kato e Ian Roberts.

Sociolingüistas e gerativistas, juntos, vêm estabelecendo que, no final do século XIX, várias mudanças encadeadas tinham acontecido no português brasileiro, predominando os seguintes fatos: a relativa cortadora, que se sobrepõe à canônica com preposições; a tendência para o preenchimento do sujeito e apagamento do objeto; a adjacência do objeto indireto ao verbo. Todos esses fatos levam ao enrijecimento da ordem no sentido de sujeito/verbo/objeto.

Mais recentemente, no Brasil, entram em cena os estudos funcionalistas que, no que se refere à mudança, centram-se nos estudos de “gramaticalizações”. Entre os pioneiros, no Brasil, dos estudos de gramaticalização estão Ataliba Castilho e Maria Luiza Braga, mas é Antoine Meillet, em 1912, que não só cria o termo gramaticalização como apresenta um estudo sobre gramaticalização, intitulado “L'évolution des formes grammaticales”.

Com a sociolingüística paramétrica se estabeleceu uma cronologia, que pode ser revista, para as mudanças sintáticas acima referidas: o final do século XIX definiria assim um limite na história do português

brasileiro. Tais mudanças vêm sendo estudadas em outros *corpora* que representem o passado do português brasileiro.

Assim, a meu ver, as teorias contemporâneas da mudança são a da variação e mudança laboviana, a gerativista e a funcionalista. A primeira se centrava em fenômenos fônicos e morfofônicos, como a concordância e o movimento dos clíticos; a segunda, na sintaxe/morfossintaxe e a terceira na morfossintaxe. Quanto às teorias fonológicas pós-estruturalismo, para além de Chomsky/Halle, já referidos, há, no que concerne ao português, o trabalho de Gladys Massini Cagliari que, com base na fonologia não-linear, analisou o ritmo e o acento na poesia trovadoresca, que estão no seu livro de 1999, *Do poético ao lingüístico no ritmo dos trovadores: três momentos na história do acento*. Os três momentos são: o acento em latim, no português arcaico e no português brasileiro. No momento, Dermeval da Hora pesquisa, no âmbito da fonologia não-linear, aspectos fônicos do século XIX do português brasileiro, com base nas *Atas da Sociedade Protetora dos Desvalidos*.

Quanto às mudanças léxico-semânticas, que se evidenciam na leitura de um texto de período recuado, vejam-se os quatro primeiros versos da famosa *Cantiga da Garvaia* de Pai(o) Soares de Taveirós, dos inícios do século XIII:

“No mundo non me sei parelha mentre me for como me vai, ca já moiro por vós e ai! mia senhor branca e vermelha”

Veja-se no verso 1: “non me sei parelha” equivalente a: não sei eu nada de semelhante; no verso 2: “mentre”, equivalente a “enquanto”; no 3º, “moiro” por “morro” antes da regularização do verbo “morrer” e no. 4º: “mia senhor branca e vermelha”, ou seja, “de pele branca e rosada”.

Uma questão que envolve a mudança é o que se designa de várias maneiras: causas, motivações, tendências, interpretações. Herbert Schendl, no seu livro *Historical Linguistics*, de 2001, considera que há três tipos gerais de explanações: (i) funcionais, como a de André Martinet; (ii) psicológicas, que se concentram nos processos cognitivos e psicolingüísticos, como propõe a teoria gerativista, e (iii) sociolingüísticas, que buscam razões no papel dos falantes como seres sociais (p. 67-68). Contudo, prefiro a posição de Roger Lass, que, no

seu livro de 1980, *On explaining Language Change*, defende que a Lingüística Histórica, como qualquer disciplina histórica, uma vez que trata de realidades humanas, não pode fazer inferências dedutivas, porque na história o contingente é mais forte que o necessário. Para esse autor – concordo com ele – nas disciplinas históricas não se pode “explicar” mas “interpretar”. Para ele a Lingüística Histórica é um “ars interpretandi”, isto é, uma “arte interpretativa”.

2. As relações entre as teorias da mudança e a história da(s) língua(s)

Começo citando B. D. Joseph e R. D. Janda na introdução do *Handbook of Historical Linguistics*:

“Não importa quão cuidadosamente tratemos com evidências documentadas sempre ficarão lacunas... Surpreendem freqüentemente as descontinuidades postas por aparentes lacunas... Na verdade não estamos diante de uma interrupção ainda que única” (p. 19: tradução minha)

Esses autores, a meu ver, querem dizer que as lacunas se devem aos remanescentes da documentação pretérita que têm de ser interpretados e nem sempre recobrem o curso completo da história de uma língua. Anteriormente mencionei que, para a história do português brasileiro, a documentação é farta e diversificada para o século XIX, mas não para os séculos anteriores.

Tratarei a seguir das histórias escritas sobre a língua portuguesa. Em geral, não se menciona Duarte Nunes de Leão e sua *Origem da língua portuguesa*. Erudito do século XVI, publica em 1606 a referida obra. Leão tinha a intuição ou saber das mudanças “que as línguas fazem por discurso de tempo”; contrapõe a língua portuguesa à língua espanhola, o que ocupa os capítulos II a VII. Nesse último, trata dos vários tipos de “corrupção” da língua latina na Espanha (leia-se “Hispania”). Dos capítulos VIII a XVIII, trata dos vocábulos, ou seja, do léxico; do capítulo XIX ao XXI apresenta uma morfossintaxe do português em relação ao latim, dos capítulos XXII ao XXVI volta aos “vocábulos”. É sem dúvida um livro pré-teórico, mas que apresenta “insights” que demonstram o conhecimento erudito do autor (Buescu, 1983, p. 327-329)

Dando um grande salto no tempo, a primeira e volumosa história da língua portuguesa é a de Serafim da Silva Neto. Elaborada ao longo da década de cinquenta do século passado, concentra-se nas origens românicas do português e, de maneira fragmentada, investe mais no português arcaico, que chama de medieval, sua especialidade. A sua *História da língua portuguesa* é compatível com a Romanística de seu tempo. Publicada primeiro em fascículos, torna-se livro em 1952. Na introdução (p. 15 a 52), busca teorizar sobre a mudança da língua, considera a teoria dos substratos, faz crítica às “leis fonéticas” e, por fim, no item 10, “História de uma língua” diz:

“A história de uma língua não é um esquema rigorosamente pré-estabelecido, não é um problema algébrico. Não se pode partir do latim e chegar diretamente aos dias de hoje, saltando por sobre vários séculos de palpitante vida... A evolução é complexa e melindrosa, relacionada com mil acidentes... uma atividade em perpétuo movimento” (1952: p. 52).

Silva Neto reconhece, em *Nota final*, “as deficiências e desigualdades e até êrros” do seu livro. O maior capítulo é sobre o *Latim Hispânico* (p. 161 a 309) e o referente ao *Português Medieval* (p. 397 a 426). Trata da *Expansão da língua* (p. 427 a 442); ao século XVI, dedica dez páginas; ao XVII, outras dez; ao XVIII, mais dez; e ao XIX, cinco páginas, sempre fundado na língua literária.

A *História e estrutura da língua portuguesa*, de Joaquim Mattoso Câmara Jr., é publicada em português em 1976. Antes em inglês, língua em que foi originalmente escrita para uma coleção americana sobre as línguas do mundo; é depois traduzida para o português, porque o texto original de Mattoso Câmara Jr. se perdeu. A meu ver, sua abordagem estruturalista é original, única e indispensável. Dedicase mais à fonologia diacrônica, uma vez que era um fonólogo. Não se limita a isso: trata também da morfologia diacrônica e apresenta dois capítulos sobre o léxico, em que mostra os efeitos da sócio-história sobre o léxico do português; destaca quatro áreas léxicas, em que apresenta em quadros a forma portuguesa, a do latim clássico e a do latim vulgar. Não ficam de fora os empréstimos, que divide em empréstimos decorrentes do convívio de duas ou mais populações em um mesmo território (como é o caso do árabe e das línguas germânicas na Península Ibérica) e em empréstimos culturais, em que duas ou mais línguas entram em contacto por via cul-

tural, como foi o caso do francês e do provençal, no período medieval. Não se esquece dos “grecismos” (lembra-se que Mattoso Câmara Jr., apesar de arquiteto, doutorou-se em Letras Clássicas). Sou uma declarada admiradora de Mattoso Câmara Jr., com quem convivo através de sua extensa obra, desde a década de sessenta e, mesmo antes, com os seus *Princípios de Lingüística Geral*, ainda no meu curso de Letras.

A história de Paul Teyssier, publicada em 1980, sintética porém mais abrangente que a de Silva Neto e a de Mattoso Câmara Jr., ficou limitada pela coleção (a “Que sais-je?”) para a qual foi concebida. Foi logo traduzida por Celso Cunha para a editora portuguesa Sá da Costa e reeditada no Brasil em 2004 pela Martins Fontes de São Paulo. No Prefácio diz Paul Teyssier, referindo-se à edição francesa:

“Numa coleção que tem de sujeitar-se a uma regra imperativa de 128 páginas de pequeno formato... obriga os autores a um esforço de síntese que favorece um rigor de pensamento e a clareza da expressão... A tradução portuguesa proporcionou-me a possibilidade de corrigir esses inconvenientes. Aproveitei... para acrescentar notas explicativas... a bibliografia foi enriquecida com títulos novos” (2004: p.1)

É constituída de cinco capítulos, sendo que o IV é dedicado ao português do Brasil e o V ao português na África e na Ásia. No capítulo I, trata “Do latim aos primeiros textos em galego-português (de 1200 a aproximadamente 1390)”; no II, aborda “O português europeu (do século XIV aos nossos dias)”. Discorre sempre sobre os fatos históricos no início de cada capítulo. É estruturalista na fonética e na fonologia (p. 29-35); reduz mais a morfologia e a sintaxe (p. 36-39); no vocabulário, trata dos empréstimos (francês ao provençal) e das palavras eruditas e semi-eruditas. Da página 48 à 81, trata da “Evolução fonética do português europeu do século XIV aos nossos dias”. Quanto à morfologia, a sintaxe e vocabulário até ao final do século XVI, dedica as páginas 81 à página 92. O capítulo IV, “O português do Brasil”, ocupa as páginas 93 à página 116, em que segue o mesmo esquema dos capítulos anteriores (fatos históricos, periodização, principais características do português do Brasil: diversidade geográfica e diversidade cultural, fonética e fonologia, morfologia e sintaxe; vocabulário e, por fim, a questão da língua no Brasil entre os escritores e entre os filólogos e lingüistas).

À página 98 apresenta um ponto de vista com que concordo plenamente. Sem ser sociolingüista, diz Teyssier:

As divisões “dialetais” no Brasil são menos geográficas que socioculturais. As diferenças na maneira de falar são maiores, num determinado lugar, entre um homem culto e o vizinho analfabeto que entre dois brasileiros do mesmo nível cultural originários de duas regiões distantes uma da outra.” (2004, p. 98)

Pode-se dizer que a *História da Língua Portuguesa* de Paul Teyssier é um pequeno grande livro.

Fernando Tarallo publicou em 1990, pela Ática, *Tempos lingüísticos: itinerário histórico da língua portuguesa*. Usando a metáfora do “túnel do tempo”, Tarallo contrapõe, nos cinco primeiros capítulos, a teoria estruturalista da mudança e a teoria sociolingüística laboviana. No capítulo “O início do túnel”, depois de resenhar a obra de Mattoso Câmara Jr., antes referida, trata do que designou “Português do Brasil versus Português europeu: as querelas”. Este capítulo discute a questão da existência de uma ou duas línguas em torno do português europeu e do português brasileiro (pp. 86-90). Nos capítulos 7, “Túnel fonológico I: as vogais” (pp. 93-105), 8, “Túnel fonológico II: as consoantes” (pp. 106-116), 9, “Túnel morfológico I: as perdas” (p. 117-131), 10, “Túnel morfológico II: os ganhos” (pp. 132-145), 11, “Túnel sintático I: fixando a ordem das palavras” (pp. 146-159), 12, “Túnel sintático II: conectando sentenças” (pp.160-172), 13, “Penetrando as paredes do túnel” (pp. 172-183) e 14, “O túnel feito texto” (pp. 184-206), o autor apresenta uma antologia de textos de 1316, 1433, 1442, 1480, 1496, 1502, 1510, 1515, 1527, 1537, 1597, 1603, 1606, 1715, 1724, 1751, 1752, 1769, 1790, 1847, 1878, 1917, 1918, 1937. Ao todo, são vinte e cinco textos. No capítulo 15, “Saindo do túnel e entrando no livro”, pp. 207-208, sumariza os capítulos de seu livro.

Note-se que a cada capítulo segue-se um resumo e exercícios. O resumo ele designa de “Pensando no (além do capítulo)”. No final do seu Prefácio, Mary Kato, sua colega e amiga na UNICAMP e na PUC/SP diz:

“Tarallo, ao escrever este livro, não dá o peixe ao leitor; ensina-o a pescar. As estratégias retóricas utilizadas... deverão garantir um processo de ensino/aprendizagem prazeroso e estimulante por parte de seu usuário, seja ele o aluno ou o professor.” (1990: p. 13)

O livro, em alguns pontos, merecia uma revisão, que o autor não pôde fazer, uma vez que faleceu logo depois. Nunca fui aluna de Tarallo, mas deveria ter sido um excelente mestre. O livro está esgotado, mas merecia uma segunda edição revista.

O *Curso de História da Língua Portuguesa* de Ivo Castro *et al.*, de 1991, foi moldado para os objetivos da coleção a que pertence, uma coleção didática para ensino à distância da Universidade Aberta. Na introdução e no Capítulo I teoriza sobre a mudança lingüística e a história da língua (p. 7 a 15). Sobre a história da língua diz:

“História da língua designa uma disciplina ou um modo de abordar os fenômenos evolutivos da língua, que tanto pode ser considerados parte integrante da lingüística histórica, como da história propriamente dita”. (1991:15)

O Capítulo 2, “Geografia da língua portuguesa”, trata da dialectologia da história do português e dos crioulos de base portuguesa (pp. 18-63). No capítulo 3, “Do latim ao português antigo”, demonstra o seu conhecimento da Romanística (pp. 66 a 161) e do português antigo (pp. 162-241). O capítulo 4 é dedicado ao português clássico (pp. 242 a 260). Nas páginas 183-184 discute a relação textual entre as duas versões remanescentes do *Testamento de Afonso II*, primeiro texto oficial e régio em galego-português. Sobre a *Notícia do Torto*, escrita entre 1212-1214, é um testemunho desordenado de anotações que, à página 225, Ivo Castro busca ordenar, o que, sem dúvida, facilita a sua leitura (pp. 224-230).

No capítulo 5, “O português clássico” (pp. 242 a 260), vale ressaltar a afirmativa do autor de que o português clássico se inicia no ano de 1536, porque simboliza três acontecimentos: a publicação da *Grammatica da lingoagem portuguesa* de Fernão de Oliveira; a representação do último Auto de Gil Vicente (*A floresta de enganos*); e a morte de Garcia de Rezende. Sem dúvida é o livro uma inestimável contribuição para a história da língua portuguesa.

Mais recentemente, Ivo Castro publica *Introdução à história do português: geografia da língua, português antigo*, cuja primeira tiragem é de 2004 e a segunda de 2005. Na síntese introdutória trata, teoricamente, de fenômenos de mudança e variação (variação social, geográfica e cronológica). Destaca que no território português se usa, além do português, outra língua, o mirandês, que goza o estatuto de

língua oficial regional (p. 2). O capítulo I, “Território e comunidade lingüística” (pp. 11-51), trata dos conceitos “língua portuguesa”, norma, crioulo, pidgin, da área lingüística do galego-português, do português extra-europeu e dos crioulos de base portuguesa. O capítulo II, “Origens do português no quadro românico” (pp. 53-81). O capítulo, “Formação de um espaço nacional para a língua portuguesa” (p. 68 a 82). Este capítulo é muito semelhante ao do *Curso* de 1991. Não trata, contudo, do português clássico, como o fez em 1991.

Em 2006, publica *Introdução à história do português*, segunda edição revista e ampliada. Qual a ampliação em relação ao livro de 2004/2005? A “Síntese introdutória”, o capítulo I e o II não se distinguem do livro anterior. No capítulo III, inclui a *Notícia de Fiadores*, datada de 1175; transcreve a *Notícia* e a crítica de António Emiliano, que prefere ver na *Notícia* “um documento notarial latino-romance” (p.123), e também a posição do filólogo galego José Antonio Souto Cabo, que

“desvalorizou a importância da *Notícia de Fiadores*, ao mesmo tempo que apresentava um outro documento como provavelmente mais antigo... Trata-se de um Pacto de Gomes Pais e Ramiro Pais, não datado, que Souto Cabo sabia entre 1173 e 1175”.

(Castro 2006: 123-124).

Conclui Ivo Castro:

“Se os esforços dos investigadores envolvidos continuam com o ritmo e o nível que têm revelado, talvez daqui surja a constituição de um *corpus antiqüíssimo galego-português*”.

(Castro 2006:125).

Finaliza o capítulo com transcrição das duas versões do *Testamento de Afonso II* de 1214, seguida de alguns comentários lingüísticos; o mesmo fez para a *Notícia do Torto*, tal como no *Curso* de 1991. Acrescenta um capítulo sobre o português médio (pp.149-184) e o português clássico e moderno.

Pelo que li, nenhuma das histórias referidas se baseou em *corpus* ou *corpora* previamente definidos. Quanto às teorias, a de Silva Neto não assume uma teoria; o *Curso* (1991) e as *Histórias* de Ivo Castro, do que se infere da leitura feita, embora falem de variação e

mudança, não aplicam a teoria, mas apresentam ricas informações, com destaque para a de 2006. Só o livro de Fernando Tarallo, sociolinguísta que era, tenta aplicar a teoria laboviana, o que ocorre nos exercícios que sucedem cada capítulo do seu livro.

3. Para finalizar

Colocarei alguns pontos a serem discutidos ou apenas como reflexão:

- Se a mudança das línguas no seu tempo histórico é um fenómeno complexo, uma única teoria não poderá dar conta de todas as mudanças ocorridas;
- No gerativismo diacrônico, que se centra na mudança sintática e assume que as mudanças se passaram de geração para geração e não nas línguas, mas nas gramáticas individuais, essa teoria dará conta de mudanças sintáticas, sem considerar o quadro histórico em que se situam as línguas;
- Na teoria da variação e mudança, teoria extrassistêmica que considera fatores de natureza sócio-histórica – classe, idade, etnia –, profundamente desenvolvida no tempo aparente, ou seja, na diacronia sincrônica, o que se coloca, neste caso, é como utilizá-la no tempo real de longa duração.

Assim sendo e, voltando ao princípio, para a reconstrução da história de uma língua, tudo indica que mais de uma teoria se faça necessária. Além disso, para a reconstituição do passado de qualquer língua se faz necessária a utilização de *corpora* representativos, que documentem o passado em todas as sincronias possíveis.

No nosso grupo de pesquisas – PROHPOR e Projeto para a História do Português Brasileiro (PHPB) –, é o que se vem fazendo. O primeiro vem trabalhando desde o século XIII ao fim do século XVI. Muitas teses, dissertações, artigos e coletâneas têm sido publicadas, mas a síntese desse tempo está por fazer e é o que se pretende no futuro. No PHPB, síntese parecida sobre a sócio-história foi feita por Tânia Lobo e Klebson Oliveira. Está publicada no *site* da Associação de Lingüística e Filologia da América Latina (ALFAL); também nesse

site está uma síntese sobre os estudos lingüísticos até então realizados. Essas sínteses, contudo, precisam ser atualizadas.

Pesquisar o passado de qualquer língua demanda paciência, saber e paixão pelo que se está pesquisando, ou seja “ouvir o inaudível” ou “a arte de fazer o bom uso dos maus dados”.

Termino com o historiador medieval José Mattoso que, no seu livro, *A escrita da história: teoria e métodos* diz:

“Prefiro... tentar descobrir as harmonias resultantes dessa fantástica sinfonia que é a História, feita da incomensurável mistura de elementos de toda espécie, tão dispersos e contraditórios com a própria vida, mas de cuja rede infinitamente complexa e fascinante procuram os eixos, os encontros e desencontros”. (1988: p. 10).

O que disse o historiador para a História aplica-se, a meu ver, à história da(s) língua(s).

Referências

- Câmara Jr., J. M. 1976. *História e estrutura da língua portuguesa*. Rio de Janeiro: Padrão.
- Câmara Jr., J. M. 1975. *História da lingüística*. Petrópolis: Vozes. [1990].
- Castro, I. et al. 1991. *Curso de história da língua portuguesa*. Lisboa: Universidade Aberta.
- Castro, I. 2004. *Introdução à história do português*. 2 ed. Revista e ampliada. Lisboa: Colibri.
- Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N.; Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row.
- Coseriu, E. 1979. *Sincronia, diacronia e história: o problema da mudança lingüística*. Rio de Janeiro: Presença.
- Foucault, M. 1972. *Arqueologia das ciências e história dos sistemas de pensamento*. Tradução de Elisa Monteiro. Rio de Janeiro: Forense Universitária. [2000].
- Guedes, M.; Berlinck, R. 2000. *E os preços eram commodos ... Anúncios em jornais brasileiros. Século XIX*. São Paulo: Humanitas.
- Hermann, P. 1880. *Princípios fundamentais da história da língua*. Lisboa: Gulbenkian. 1970.
- Jakobson, R. 1931. *Principes de phonologie historique*. Paris: Klincksieck. [1964].
- Joseph, B. D.; Janda, R. J. 2002. *Handbook of historical linguistics*. New York: Blackwell.

- Labov, W. 1982. Building on empirical foundations. In: W. Lehmann; Y. Malkiel (Eds.). *Perspectives in Historical Linguistics*. Amsterdam: John Benjamins, 17-82.
- Lass, R. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.
- Lass, R. 1980. *On explaining language change*. Cambridge: Cambridge University Press.
- Lightfoot, D. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.
- Lobo, T. et al. 2001. *Cartas baianas setecentistas*. São Paulo: Humanitas.
- Martinet, A. 1955. *Économie des changements phonétiques: traité de phonologie diachronique*. Bern: Francke.
- Massini-Cagliari, G. 1999. *Do poético ao lingüístico no ritmo dos trovadores: três momentos da história do acento*. Araraquara: Cultura Acadêmica Editora UNESP.
- Meillet, A. 1912. L'évolution des formes grammaticales. In: A. Meillet (1958). *Linguistique historique et linguistique générale*. Paris: Klincksieck.
- Nunes de Leão, D. 1606. *Ortografia e origem da língua portuguesa*. Edição de M. L. C. Buescu. Lisboa: Imprensa Nacional – Casa da Moeda. [1983].
- Roberts, I.; Kato, M. (Orgs., 1993). *Português Brasileiro: uma viagem diacrônica*. Campinas: Editora da UNICAMP.
- Sapir, E. 1921. *A linguagem: uma introdução ao estudo da fala*. 2 ed. Rio de Janeiro: Acadêmico. [1971].
- Schendl, H. 2001. *Historical linguistics*. Oxford: Oxford University Press.
- Schuchardt, H. 1928. *Brevier: ein vademecum der allgemein Sprachwissenschaft*. 2 ed. Harle.
- Silva Neto, S. 1976. *História da língua portuguesa*. Rio de Janeiro: Livros de Portugal.
- Tarallo, F. 1990. *Tempos lingüísticos: itinerário histórico da língua portuguesa*. São Paulo: Ática.
- Teyssier, P. 1980. *História da língua portuguesa*. São Paulo: Martins Fontes. 2004
- Weinreich, U.; Labov, W.; Herzog, M. 1968. Empirical foundations for a theory of language change. In: W. Lehman; Y. Malkiel (Eds.). *Directions for Historical Linguistics*. Austin: University of Texas Press, 95-188.

Locativos preposicionados em posição de sujeito:
uma possível contribuição das línguas Bantu
à sintaxe do português brasileiro

Juanito Avelar
juanitoavelar@iel.unicamp.br
Universidade Estadual de Campinas (Brasil)

Sonia Cyrino
cyrino@iel.unicamp.br
Universidade Estadual de Campinas (Brasil)

RESUMO. Observando o contraste existente entre o português brasileiro e o português europeu em relação a locativos preposicionados em construções encaixadas, argumentamos que esses constituintes podem ocupar a posição gramatical de sujeito em sentenças com verbos transitivos no português brasileiro. Após examinar a chamada *inversão locativa* (em que um constituinte locativo não-argumental é realizado em posição de sujeito) encontrada em várias línguas, focalizamos os casos da construção em línguas Bantu, uma vez que algumas delas eram faladas pela maioria dos escravos chegados ao Brasil entre os séculos XVI e XIX. Sugerimos, preliminarmente, que as inovações do português brasileiro são o resultado do português usado como segunda língua por falantes nativos de línguas Bantu. A possibilidade de um locativo ser pré-verbal e concordar com o predicado nessas línguas, aliada ao enfraquecimento do paradigma flexional presente nos dados lingüísticos primários, gerou ambigüidade para a criança que aprendia o português como primeira língua. O sintagma locativo pré-verbal, em uma posição periférica no português, é reanalisado e passa a poder figurar na posição gramatical de sujeito no português brasileiro.

PALAVRAS-CHAVE. Locativos preposicionados, inversão locativa, português brasileiro, línguas Bantu, sintaxe gerativa, mudança sintática.

ABSTRACT. We observe the existing contrast between Brazilian and European Portuguese in relation to prepositional locatives in embedded sentences, and we argue that those constituents can occupy the grammatical subject position in transitive sentences in Brazilian Portuguese. After examining the so-called *locative inversion* (in which a non-argumental locative constituent may be in subject position) found in various languages, we focus on the cases of the construction in the Bantu languages, since some of them were spoken by the majority of slaves brought to Brazil between the 16th and 19th centuries. We suggest that the Brazilian Portuguese innovations are the result of the

Portuguese used as a second language by Bantu languages native speakers. The possibility for a locative to be pre-verbal and agree with the predicate in those languages combined with the impoverishment of the inflectional paradigm present in the primary linguistic data generated ambiguity for the child acquiring Portuguese as her first language. The pre-verbal locative phrase, which was in a peripheral position in Portuguese, is reanalyzed as possible in the position of grammatical subject in Brazilian Portuguese.

KEY-WORDS. Prepositional locatives, locative inversion, Brazilian Portuguese, Bantu languages, generative syntax, syntactic change.

1. Introdução

O português brasileiro (PB) e o português europeu (PE) apresentam um contraste bastante proeminente em torno de sentenças como aquela exemplificada em (1) a seguir, que traz um sintagma locativo preposicionado na posição pré-verbal da oração encaixada. Construções desse tipo recebem, em PE, uma interpretação na qual o sujeito nulo da encaixada é correferente ao sujeito da principal, tal como indicado em (2a). Em contraste, falantes do PB atribuem a essa construção uma interpretação indeterminadora, similar a casos em que o *se*, no PE, é empregado como índice de indeterminação, como em (2b).¹

- (1) Ele disse que **naquela loja** vende livros.
 (2) a. Elei disse que **naquela loja** *cvi* vende livros.
 b. Ele disse que se vende livros **naquela loja**.

Uma vez que o emprego de *se* indeterminador não é usual no PB, a versão indeterminadora da construção em (1) poderia resultar de um simples desuso desse pronome nos contextos em questão.² Essa visão, contudo, é insuficiente para explicar o fato de, no PB, o locativo ser obrigatório entre esses casos, como podemos observar

¹ Em PB, a única situação em que se pode atribuir uma leitura referencial para um sujeito nulo no interior da encaixada em construções desse tipo parece-nos ser aquela em que o locativo recebe uma leitura de foco contrastivo, quando é pronunciado com uma entonação marcada, como destacado em (i) a seguir.

(i) O Pedro disse que, NAQUELA LOJA (e não em outra), (ele) vende livros.

² Para uma abordagem em torno de construções indeterminadoras sem o emprego do *se* no PB, ver Galves (2001).

em (3). Se toda a história pudesse ser resumida a uma mera rejeição ao *se*, a construção em (3a), sem o locativo, deveria ser licenciada no PB da mesma forma que o PE licencia a construção com *se* em (4), na qual a presença ou ausência do locativo é indiferente para garantir a aceitabilidade.

- (3) a. *Vende muitos livros.
b. **Lá / Naquela loja** vende muitos livros.
- (4) (Naquela loja) vende(m)-se muitos livros

Assumindo a versão minimalista da Teoria de Princípios e Parâmetros (Chomsky 1995; 2000; 2001), partiremos dos contrastes em questão para argumentar que, diferentemente do que se observa no PE (e, até onde saibamos, nas outras línguas românicas), locativos preposicionados (doravante, PPloc) podem ocupar a posição gramatical de sujeito em sentenças com verbos transitivos do PB. Nestes termos, a impossibilidade, no PB, de o sujeito da principal e da encaixada ser correferentes em (1) resulta do fato de o locativo ocupar a posição de sujeito no interior da encaixada e, portanto, impedir que qualquer outro elemento ocorra na mesma posição.³

Iremos sugerir, em seguida, que essa inovação do PB pode ser uma contribuição de línguas da família Bantu, faladas pela maioria dos escravos chegados ao Brasil entre os séculos XVI e XIX. Uma vez que essas línguas admitem, em larga escala, a chamada *inversão locativa* (denominação comumente atribuída a casos em que um constituinte locativo não-argumental é realizado em posição de sujeito), é possível que os fatos do PB sejam o resultado de um processo de aquisição em que um PPloc pré-verbal (no PE, um constituinte adjungido) é reanalisado como alocado na posição gramatical de sujeito.⁴

³ Contudo, como destacado na nota 1, casos em que o locativo pré-verbal é interpretado como foco contrastivo parecem admitir a leitura em que o sujeito nulo da encaixada seja correferente ao sujeito da principal. Em tal situação, o locativo estaria não na posição gramatical de sujeito, mas concatenado em alguma posição periférica, na qual receberia a leitura contrastiva. Nessa condição, a posição de sujeito da encaixada estaria livre para alocar uma categoria vazia com o mesmo referente que o sujeito da principal.

⁴ Um dos revisores deste artigo destaca o fato de o trabalho não discutir hipóteses alternativas, bem como de não apresentar fatos de outras línguas românicas

O trabalho é dividido como se segue: na seção 2, serão apresentados alguns contrastes interlingüísticos relacionados a padrões de inversão locativa, com base em dados do inglês e de algumas línguas Bantu; na seção 3, apresentamos evidências de que, no PB, locativos preposicionados pré-verbais podem, em sentenças com verbos inacusativos, inergativos e transitivos, ocupar um lócus normalmente identificado como sendo a posição de sujeito; na seção 4, sugerimos preliminarmente que as inovações do PB quanto à inversão locativa podem ser o resultado de um processo em que os dados lingüísticos primários, na aquisição da língua, deixaram de ser transparentes quanto à posição do PPloc pré-verbal, em função do aprendizado do português como segunda língua por falantes nativos de línguas Bantu; na seção 5, apresentamos as considerações finais.

parecidos aos atestados no português brasileiro. De acordo com o revisor, “a hipótese de substrato talvez seja inútil” à luz desses fatos.

Sobre a discussão de hipóteses alternativas, este é, até onde saibamos, o primeiro trabalho a abordar o fenômeno da inversão locativa no português brasileiro de uma perspectiva comparativa. Não há, considerando o caráter ainda incipiente das discussões sobre o tema, como apresentar hipóteses alternativas acerca desses mesmos fatos.

Sobre a apresentação de fatos parecidos em outras línguas românicas, não temos, até aqui, notícias de que os padrões relevantes identificados no português brasileiro (inversão locativa com verbos inergativos e transitivos) ocorram entre outras línguas da mesma família (ver a seção 3). A esse respeito, fazemos menção ao estudo de Fernandez-Soriano (1999) para o espanhol (ver seção 3.3), que revela serem bastante restritos os contextos de inversão locativa na gramática do castelhano. Cremos que os fatos do espanhol podem, guardadas as particularidades de cada língua, ser representativos do que ocorre em outras gramáticas ibéricas e, talvez, em outras línguas românicas.

A respeito da noção de substrato, acreditamos que o quadro teórico aqui assumido (a vertente diacrônica desenvolvida a partir da Teoria de Princípios e Parâmetros, particularmente na perspectiva de Roberts 2007) nos permite prescindir desse conceito. Da perspectiva a que recorremos, o que entra em jogo são os seguintes elementos: (a) os fatos de natureza mentalista, biológica, relacionados à aquisição de uma língua; (b) as possíveis correlações, dessa mesma natureza, que podemos observar na aquisição de uma segunda língua por indivíduos adultos (como aquela que caracterizou a aquisição do português por falantes de línguas Bantu – ver a seção 4.1) e (c) as condições que possibilitam a transmissão de possíveis inovações desencadeadas durante o processo de aquisição às futuras gerações de aprendizes da língua. Não vemos, dessa perspectiva, razões para sustentar este trabalho na noção clássica de substrato.

2. Inversão locativa: alguns contrastes interlingüísticos

O fenômeno a que se costuma denominar *inversão locativa* corresponde a casos em que constituintes locativos, normalmente tidos como não-argumentais, ocorrem em uma posição da sentença que pode ser identificada como a posição gramatical do sujeito (para discussões mais detalhadas, ver Bresnan & Kanerva (1989), Mok (1992), Levin & Rappaport Hovav (1995), Fernandez-Soriano (1999), Demuth & Mmusi (1997), Baker (2003), Marten (2006) E Mendikoetxea (2006), dentre vários outros). As línguas naturais não se comportam uniformemente quanto às propriedades que envolvem o fenômeno, podendo diferir, entre outros aspectos, quanto a propriedades de concordância e à especificidade da estrutura argumental que licencia a inversão.⁵

Em relação à concordância, a diferença está na possibilidade de o verbo concordar ou com o constituinte locativo em posição pré-verbal ou com o constituinte pós-verbal que corresponde ao sujeito argumental. As sentenças do inglês e do chichewa (família Bantu) em (5) e (6) abaixo exemplificam cada possibilidade. No inglês, o verbo sempre concorda com o sujeito argumental pós-verbal; em chichewa, é o locativo pré-verbal que desencadeia a concordância.

- (5) a. *From such optical tricks arise all the varieties of romantic hallucination...*
 b. *...out of the house came a tiny old lady...*

(Levin & Rappaport Hovav 1995: 220-221)

- (6) CHICHEWA

a. *ku-um-dzi ku-na-bwér-á a-lendô-wo*
 17-3-village SC17-PAST-come-FV 2-visitors-those
 'To the village came those visitors.'

b. *m-mi-têngo mw-a-khal-a a-nyãni*
 18-4-tree SC18-PERF-sit-FV 2-baboons
 'In the trees are sitting baboons.'

(Bresnan & Kanerva 1989)

Quanto à estrutura argumental que licencia a inversão locativa, há pelo menos dois padrões: (i) a inversão locativa é admitida apenas (ou, pelo menos, preferencialmente) em construções inacusativas; (ii) a inversão locativa ocorre em uma variedade mais ampla de cons-

⁵ Para um quadro contrastivo detalhado, ver Salzmann (2004).

truções. O inglês e o chichewa exemplificam o primeiro caso, como podemos observar em (7) e (8) abaixo.

- (7) a. In the distance APPEARED the towers and spires of a town [...].
 b. * In the cafés of Paris TALK many artists.
 (Levin & Rappaport Hovav 1995: 220-221)

(8) CHICHEWA

- a. *ku-mu-dzi ku-na-bwér-á a-lendô-wo*
 17-3-village SC17-PAST-come-FV 2-visitors-those
 'To the village came those visitors.'
- b.* *m-mi-têngo mu-kú-imb-á a-nyăni*
 18-4-tree SC18-PROGR-sing-FV 2-baboons
 Lit.: 'In the trees are singing baboons.'

(Bresnan & Kanerva 1989)

Em contraste, línguas Bantu como o setswana e o kinyarwanda, com dados apresentados a seguir, admitem a inversão locativa com verbos inergativos e transitivos (Salzamann 2004). A única restrição que impõem é para que os argumentos interpretados como agente e tema não co-ocorram na mesma sentença.

(9) SETSWANA

- Mó-le-fátshé-ng gó-fúla di-kgomo
 18-5-country 17SM-graze 10-cattle
 'In the country are grazing the cattle.'

(10) KINYARWANDA

- a. *kw' íisôko ha-Ø-guz-w-e ibi-íntu bi-taandátu*
in market 16-PST-buy-PAS-PRF 8-thing 8-six
 'At the market were bought six things.'
- b. *um cyumba ha-Ø-ríi-r-iyé umwána*
in room 16-PST-eat-APL-PRF 1:child
 'The child ate in the room.'

Destacamos que não é simples classificar as línguas naturais, em termos tipológicos, quanto aos diferentes padrões de estruturas argumentais que admitem a inversão locativa. Há casos, por exemplo, como o do castelhano, em que a inversão locativa é admitida apenas com uma subclasse restrita de verbos inacusativos (ver Fernandez-Soriano 1999). Há, ainda, situações como a encontrada em otjiherero, que compõe um grupo de línguas mais liberais que o setswana

e o kinyarwanda quanto às estruturas argumentais que admitem a inversão (ver Marten 2006). Mesmo no inglês, em que a inversão locativa costuma ser usada como um diagnóstico de inacusatividade, ocorrem estruturas com verbos tipicamente inergativos que trazem um PPloc em uma aparente posição de sujeito (ver, por exemplo, Levin & Rappaport Hovav (1995) e Mendikoetxea 2006).

Para os objetivos deste trabalho, concentramo-nos, de forma simplificada, no recorte entre (a) línguas que admitem a inversão locativa apenas na situação em que um único argumento, interpretado como tema, estiver presente na estrutura (inglês e chichewa), e (b) línguas que afrouxam esta restrição, admitindo que esse argumento único seja ou o agente, ou o tema (setswana e kinyarwanda).

3. Inversão locativa no português brasileiro

Nesta seção, apresentamos alguns fatos que corroboram a idéia de que, diferentemente das demais línguas românicas, o PB admite inversão locativa em sentenças construídas com verbos inergativos e transitivos, similarmente ao observado em línguas como o setswana e o kinyarwanda. Os conjuntos em (11)-(14) abaixo mostram que verbos inacusativos, transitivos ergativizados (ou seja, sem o agente) e inergativos admitem, em PB, a ocorrência de um locativo pré-verbal, sem que qualquer DP/NP argumental apareça na posição de sujeito.

(11) **verbos inacusativos**

- a. Na casa da Maria *chegou* algumas cartas.
- b. No meio da festa *apareceu* uns convidados estranhos.
- c. Naquele documento *consta* o nome da Maria.

(12) **verbos inergativos**

- a. Naquele quarto *dormiu* várias pessoas.
- b. Naquela fábrica *trabalha* muitos amigos meus.
- c. Na universidade *estuda* a filha de uma amiga minha.

(13) **verbos transitivos ergativizados**

- a. Naquele bairro *aluga* casa de todos os preços.
- b. Na loja do Pedro não *conserta* sapato de couro.
- c. Naquele fazenda *plantava* beterraba.

(14) **verbos transitivos e inergativos sem tema e/ou agente**

- a. Nas cidades do interior não *seqüestra* tanto como nas grandes capitais.
- b. No casa do João *cozinha* todos os dias.
- c. Na casa da Maria *dorme* cedo.

Cabe ressaltar que o fato de as sentenças acima apresentarem um PPloc pré-verbal não é, por si só, revelador da posição em que este se encontra. Os testes a seguir, contudo, evidenciam que os PPloc estão licenciados no lócus comumente identificado como posição de sujeito.

3.1. Obrigatoriedade do constituinte locativo em contextos com sujeitos pós-verbais

Os dados em (a) apresentados em (15)-(16) mostram que o PPloc é opcional nos casos em que o sujeito argumental é pré-verbal; em contraste, nos casos em que o sujeito é pós-verbal, como em (b), o PPloc passa a ser obrigatório. Essa obrigatoriedade pode ser adequadamente explicada se assumirmos que o locativo satisfaz a condição EPP quando o sujeito é pós-verbal (como em (15b)) ou não está presente na sentença (como em (16b)-(17b)).⁶ Para tanto, é necessário assumir

⁶ É importante observar, contudo, que constituintes locativos podem “salvar” as sentenças se também forem inseridos em posição final, tal como abaixo:

- a. Vende muitas calças *naquela loja*.
- b. Grava todo tipo de filme *nesse meu DVD*.
- c. Trabalha vários amigos meus *naquela fábrica*.

Esses casos poderiam ser contra-argumento à idéia de que o constituinte locativo “salva” a sentença com o sujeito posposto por meio de sua entrada em posição de sujeito, que seria, no PB, uma posição pré-verbal. Os fatos em (ii), contudo, mostram que sujeitos podem ser licenciados em posição final, se forem interpretados como tópico e/ou forem retomados por um pronome pessoal em posição pré-verbal.

- (ii) a. (ele_i) não quis almoçar hoje, o Roberto_i
- b. (ela_i) vai comprar um carro novo, a Maria_i
- c. (elas_i) comeram a comida toda, as crianças_i

Da mesma forma, locativos em posição final podem ser retomados por elementos adverbiais pronominais presentes na posição pré-verbal, como nas sentenças em (iii). Esses fatos mostram que, ao contrário do que possa parecer, a possibilidade de locativos em posição final não é um contra-argumento à hipótese de que esses elementos podem garantir a aceitabilidade da sentença ao entrar na posição tipicamente ocupada pelo sujeito argumental, já que sujeitos argumentais também podem aparecer, nas condições apontadas, em posição final.

- (iii) a. (*lá*_i) vende muitas calças, *naquela loja*_i
- b. (*af*_i) grava todo tipo de filme, *nesse meu DVD*_i
- c. (*lá*_i) trabalha vários amigos meus, *naquela loja*_i

que o locativo esteja em [Spec,TP/IP], sem o quê a obrigatoriedade não poderá ser tratada em termos de satisfação do EPP.

- (15) a. *(Naquele quarto)* várias pessoas dormiram.
 b. **(Naquele quarto)* dormiu várias pessoas.
- (16) a. *(Naquela loja)* todos os tipos de livro vendem.
 b. **(Naquela loja)* vende todos os tipos de livro.
- (17) a. *(Na casa do João)* o Pedro cozinha todos os dias.
 b. **(Na casa do João)* cozinha todos os dias.

3.2. Alçamento

As construções com o verbo de alçamento *parecer* mostram que o PPloc deve ocorrer em posição pré-verbal se o sujeito argumental da sentença permanecer posposto. Dados desse tipo revelam que os locativos funcionam da mesma forma que sujeitos nominais quanto à obrigatoriedade de serem movidos para [Spec,TP] em construções de alçamento. Seguindo a mesma linha de raciocínio sugerida para os dados em (15)-(17) acima, podemos afirmar que o PPloc pode satisfazer a condição EPP também nesses tipos sentenciais.

- (18) a. * Parece *na casa da Maria* chegar muitas cartas.
 b. *Na casa da Maria* parece chegar muitas cartas.
- (19) a. * Parece *naquele shopping* trabalhar muita gente.
 b. *Naquele shopping* parece trabalhar muita gente.
- (20) a. * Parece *naquela loja* vender bastante sapato.
 b. *Naquela loja* parece vender bastante sapato.
- (21) a. * Não parece *nas cidades do interior* seqüestrar tanto como nas grandes capitais.
 b. *Nas cidades do interior* não parece seqüestrar tanto como nas grande capitais.

3.3. Estatuto de foco e locativos pré-verbais

Segundo Fernandez-Soriano (1999), somente alguns tipos de verbos monoargumentais admitem a ocorrência de PPloc em posição de sujeito no espanhol. Dentre os fatos explorados pela autora, estão aqueles que dizem respeito ao contraste observado em (22) a seguir: sentenças em que verbos do tipo *faltar* e *sobrar* ocorrem com locativos pré-verbais e sujeitos pospostos podem ser apresentadas como respostas a um pergunta do tipo *¿Qué pasa / pasó?*, mas não sentenças com o inacusativo *llegar* ou com o inergativo *rodar*.

- (22) a. – ¿Qué pasa / pasó?
 b. – En esta casa *falta* café.
 – En esta classe *sobran* estudantes.
 – # Por la colina *rodó* el carrito del niño.
 – # A esta casa *llegaron* estudantes.

Para a autora, a explicação para esse contraste passa pelo fato de o PPloc ocupar uma posição argumental nas construções com *faltar* e *sobrar*; nas outras, o PPloc ocuparia a posição periférica de tópico, que é não-argumental. Isso significa que aquelas sentenças, mas não estas, podem receber integralmente o estatuto de foco e, como tal, serem dadas como resposta à indagação *¿Qué pasa / pasó?*. Ou seja, somente na posição de sujeito, mas não na posição de tópico, pode o PPloc integrar uma resposta a essa pergunta.

Os exemplos em (23) mostram que o PB não faz distinções tão restritas quanto o espanhol. Se a proposta de Fernandez-Soriano for válida, isso significa que o PPloc deve estar na posição de sujeito, e não em uma posição periférica à esquerda, em todos os casos a seguir:

- (23) a. – O que está acontecendo? / – O que aconteceu?
 b. – Na casa da Maria *chegou* umas pessoas estranhas.
 – Na casa da Maria *dormiu* umas pessoas estranhas.
 – Naquela loja *vendeu* mais de cem pares de sapatos ontem.
 – Na casa do João não está mais *cozinhando* aos finais de semana.

3.4. *Inversão locativa e construções transitivas ergativizadas*

De um ponto de vista meramente descritivo, a inversão locativa corresponde a uma troca de posições entre um constituinte locativo e o sujeito. Essa visão traria um aparente problema para classificarmos os casos com verbos transitivos ergativizados,⁷ exemplificados em (13), como instâncias canônicas de inversão locativa, uma vez que, em tais construções, não há qualquer sujeito argumental que possa “trocar” de posição com o locativo.

Os dados em (24)-(27) a seguir, contudo, revelam que, ao contrário do que parece à primeira vista, constituintes locativos podem trocar de posição com o próprio tema na ocupação de [Spec,TP/InflP].

⁷ Para uma discussão mais detalhada sobre construções transitivas ergativizadas, ver Franchi, Negrão & Viotti (1998).

Observe-se que tanto o PPloc quanto o tema podem ser alternados na posição pré-verbal, situação que caracteriza descritivamente os casos classicamente tratados como inversão locativa.

- (24) a. Naquela loja *vende* todas as edições do livro do Harry Potter.
b. Todas as edições do livro do Harry Potter *vendem* naquela loja.
- (25) a. No meu DVD não *grava* filmes antigos.
b. Filmes antigos não *gravam* no meu DVD.
- (26) a. Naquela fotocopiadora não *xeroca* os livros novos da biblioteca.
b. Os livros novos da biblioteca não *xerocam* naquela fotocopiadora.
- (27) a. Nas lojas do centro só *conserta* sapatos importados.
b. Sapatos importados só *consertam* nas lojas do centro.

Os exemplos em (28)-(29) adiante, que apresentam sentenças com os verbos transitivos *adorar* e *odiar*, confirmam o caráter de inversão atribuído às construções acima. Nestes casos, o PPloc não é licenciado em posição pré-verbal. Essa propriedade pode ser explicada pelo fato de, nestas mesmas sentenças, o tema não ser um candidato à posição de sujeito, conforme se pode atestar pela clara inaceitabilidade das construções em (b). Em outras palavras, construções com os transitivos *adorar* e *odiar* não admitem a inversão locativa porque também não permitem que o tema ocorra em [Spec,TP], não havendo, nesses casos, qualquer elemento argumental que possa inverter sua posição com o PPloc.

- (28) a. * Na casa da Maria *adora* os livros do Harry Potter.
b. * Os livros do Harry Potter *adoram* na casa da Maria.
- (29) a. * Na faculdade *odeia* os professores conservadores.
b. * Os professores conservadores *odeiam* na faculdade.

4. Especulando sobre as causas da inovação no português brasileiro

Nesta seção, iremos delinear uma hipótese, em nível ainda bastante preliminar, sobre uma possível causa para a emergência dos padrões de inversão locativa identificados no PB. Essa hipótese passa por assumir uma influência da gramática de línguas Bantu sobre a sintaxe do PB, bem como uma possível correlação entre padrões de inversão locativa e propriedades do paradigma flexional.

4.1. *Ambigüidade em dados lingüísticos primários*

Em propostas como as de Roberts (2007), **dados lingüísticos primários** (PLD, de *Primary Linguistic Data*) podem levar a mudanças gramaticais, quando não são transparentes o suficiente para permitir que a criança fixe a marcação de um determinado parâmetro. Situações de contato entre membros falantes de línguas distintas figuram entre os fatores extralingüísticos que ocasionam essa condição. Em tais situações, é plausível pressupor que os falantes, ao se comunicarem em uma segunda língua, empreguem estruturas que correspondam, o mais próximo possível, à configuração de sua língua materna.

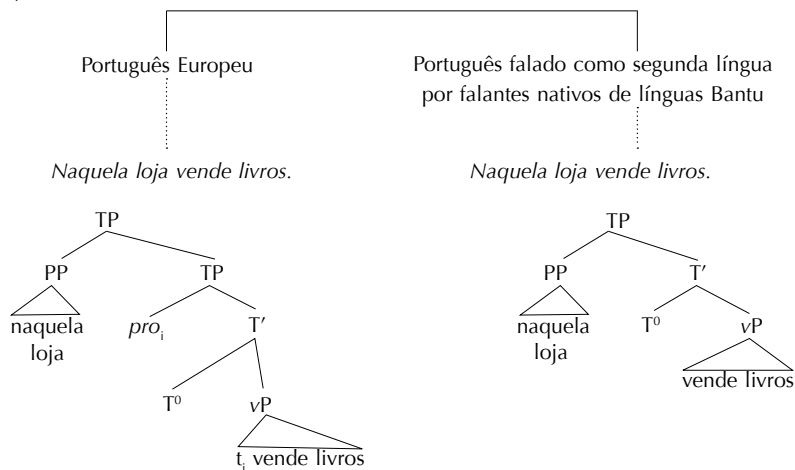
Roberts & Roussou (2003) argumentam em favor da idéia de que, para lidar com a falta de transparência nos dados, a criança “procura” a representação *menos marcada* para determinada categoria. Para definir o que seria uma representação *menos marcada*, considera-se uma hierarquia em que a operação MERGE (nos termos de Chomsky 1995) seria mais simples (e, portanto, menos marcada) do que a operação MOVE.

Tendo em mente essa idéia, um possível quadro para a situação de aquisição da língua em terras brasileiras durante o período colonial seria aquele em que os escravos chegados da África, falantes (em sua grande maioria) de línguas Bantu, aprendiam e moldavam o português de acordo com a gramática de sua língua materna. Considerando que as línguas Bantu são mais permissivas que as línguas românicas no que diz respeito aos contextos que admitem a inversão locativa (ver a seção 2), podemos imaginar uma situação em que os africanos analisariam locativos pré-verbais como categorias instadas na posição gramatical de sujeito, tal como admitido em sua primeira língua, e não em uma posição periférica.

Seriam comuns situações em que as crianças se defrontariam com sentenças do tipo *Naquele loja vende roupas*, à qual poderiam ser associadas duas estruturas possíveis, tal como no esquema apresentado abaixo, a depender do falante que estivesse produzindo a construção.

(30)

PLD



Se a proposta de Roberts & Roussou estiver correta, a criança optaria pela estrutura gerada pelos aprendizes de português como segunda língua, por ser esta *mais simples* em relação à estrutura gerada pelos falantes de português como língua materna: nesta, é preciso assumir uma categoria pronominal nula que se move de [Spec,vP] para [Spec,TP], com o constituinte locativo sendo concatenado em alguma posição periférica; naquela, basta assumir a concatenação do constituinte locativo diretamente em [Spec,TP]. Em outras palavras, a criança optaria pela estrutura fornecida pelos escravos, com inversão locativa, em função de estas serem menos marcadas que as produzidas por falantes de PE.

Se nossa hipótese estiver correta, a possibilidade de concatenação do constituinte locativo diretamente em [Spec,TP] seria acompanhada de um processo em que a presença de um argumento externo em [Spec,vP], a ser interpretado como agente, deixa de ser obrigatória. Trata-se do mesmo padrão observado entre construções passivas analíticas, em que o argumento externo do verbo é dispensado, com a diferença de que, entre as passivas, não há qualquer constituinte locativo sendo requerido para garantir a aceitabilidade da sentença.

Os fatos observados por Gonçalves & Chimbutane (2004) no português moçambicano reforçam a idéia de que as inovações demonstradas pelo português brasileiro para a sintaxe dos sintagmas locativos resultam de influências das línguas Bantu. No português

moçambicano, constituintes introduzidos por *em* também são licenciados em posição de sujeito, como nos exemplos a seguir.

- (31) PORTUGUÊS MOÇAMBICANO
 a. **Em casa dele** é aqui em frente.
 b. **Na nossa zona** era fértil.
 c. **Na igreja** é pequeno.

(Gonçalves & Chimbutane 2004)

De acordo com Gonçalves & Chimbutane, esse fato seria reflexo de propriedades das línguas Bantu faladas em Moçambique pelos aprendizes de português. Para destacar paralelos entre essa inovação e propriedades da língua primeira desses aprendizes, os autores apresentam fatos do Changana, uma das línguas faladas em território moçambicano. Em Changana, há padrões de inversão locativa similares aos identificados no português moçambicano, como se pode facilmente atestar pela comparação das construções em (32) a seguir com aquelas apresentadas em (31). Observe-se, em (32), que o constituinte locativo *kerek-eni* 'na igreja' ocorre em posição de sujeito, desencadeando a manifestação da marca de concordância locativa junto ao verbo, que corresponde ao índice 17. Um constituinte do mesmo tipo, introduzido pela preposição *em*, aparece na posição de sujeito das sentenças do português moçambicano apresentadas em (31).

- (32) CHANGANA
 a. *Kerek-eni* **ku-nghen-ile** *tintombhi*
 9igreja-Loc 17SM-entrar-PS 10-rapariga
 'Na igreja entraram raparigas.'
 b. *Kerek-eni* *i* **ku-tsongo**.
 9igreja-Loc é 17-pequeno
 'Na igreja é pequeno.' (sem equivalente no português europeu)

(Gonçalves & Chimbutane 2004)

O fato de variedades do português africano apresentarem inovações similares às identificadas no português brasileiro corrobora a idéia de que os padrões brasileiros de inversão locativa, sem paralelo entre as línguas românicas, resultam do contato entre o português e as línguas Bantu.⁸

⁸ A proposta de Gonçalves & Chimbutane (2004) para explicar as particularidades de sintagmas introduzidos por *em* no português moçambicano é diferente da que

4.2 Inversão locativa e concordância

Ainda que correta, a hipótese sugerida em 4.1 não pode ser a história completa. É preciso capturar não apenas a opção pela entrada do locativo em uma posição tipicamente destinada a constituintes argumentais, mas também o fato de o aprendiz do PB deixar de empregar o pronome nulo referencial nessas construções. Em relação a este segundo fato, tem sido consensual na literatura que as mudanças na morfologia verbal tiveram um papel determinante nas inovações do PB – uma delas diz respeito exatamente ao desaparecimento de pro nulo referencial, que teria deixado de ser licenciado devido ao desenvolvimento de um paradigma flexional empobrecido (ver, por exemplo, Ferreira 2000).⁹

Seguindo essa linha de raciocínio, Avelar (2006) sugere que, no PB popular, a opcionalidade da concordância sujeito-verbo, como exemplificado em (31) a seguir, está associada ao fato de a categoria

sugerimos para o português brasileiro (ver 4.2). Para aqueles autores, os constituintes locativos introduzidos por *em* foram, na variedade africana, reanalisados como constituintes nominais, com a preposição passando a corresponder a uma categoria interna ao DP/NP. O licenciamento desses constituintes na posição de sujeito estaria, dessa forma, relacionado ao fato de sintagmas como *na igreja* corresponderem, no português moçambicano, a verdadeiros sintagmas nominais, e não a sintagmas preposicionados.

De fato, apesar das similaridades entre as variedades brasileira e moçambicana do português, há contrastes que justificam um tratamento diferenciado para a inversão locativa nas duas línguas. Esses contrastes podem, no que diz respeito à história externa, ser atribuídos a diferenças nos contextos sociolingüísticos que marcam a formação de uma e outra variedade (Avelar, Cyrino & Galves 2008). Como destacado por Gonçalves & Chimbutane, a aquisição do português como segunda língua em Moçambique passa por pelo processo de escolarização (diferentemente da forma como os negros chegados ao Brasil aprendiam o português), somando-se a isso o forte peso das línguas africanas, evidentemente muito maior em Moçambique do que no Brasil colonial. Além disso, sabemos que, no Brasil, diferentemente do que se observa entre a população moçambicana nos dias atuais, os aprendizes da língua tornaram-se monolíngües, o que certamente pode ser apontado como um fator que contribui para as diferenças entre o português brasileiro e a gramática emergente no território moçambicano.

⁹ Para outros estudos sobre as conseqüências da simplificação da morfologia verbal em PB, ver Duarte (1995), Figueiredo Silva (1996), Galves (1996; 2001), Kato (2000), Rodrigues (2002), Martins & Nunes (2008), dentre outros.

T poder ser licenciada sem traços- ϕ . Isso significa que, em PB, constituintes nucleados por categorias sem traços- ϕ (como as preposições) podem ser licenciadas na posição de sujeito, já que a concordância sujeito-verbo não precisa ser desencadeada. Em PE, ao contrário, a categoria T sempre porta traços- ϕ , o que implica a obrigatoriedade da concordância sujeito-verbo e, conseqüentemente, a presença de constituintes nucleados por categorias com traços- ϕ (DPs/NPs) em posição de sujeito.

(31) Os menino comeu/comeram o bolo.

Além de capturar o contraste entre o PE e o PB em relação ao licenciamento de PPloc em posição de sujeito, uma análise desse tipo abre espaço para tratar, sob uma perspectiva unificada, a emergência diacrônica dos padrões de inversão locativa e a simplificação da morfologia verbal. Retornando à especulação a respeito de uma possível influência das línguas Bantu sobre a sintaxe do PB, poderíamos afirmar que ambas as inovações podem ser correlacionadas como parte de um mesmo “pacote de mudanças”.

Há pelo menos duas evidências que reforçam a idéia de que essa correlação existe. Uma delas tem a ver com particularidades do verbo possessivo *ter* no PB, destacadas em Avelar (2008). Conforme o autor, o uso de *ter* como verbo existencial no PB está justamente atrelado ao fato de, nessa língua, a estrutura das sentenças possessivas admitir um PPLOC em [Spec,TP/IP]. Se estiver correta, essa hipótese implica que o verbo possessivo começou a ser usado como existencial a partir do momento em que passou a licenciar o locativo preposicionado em posição de sujeito. A esse respeito, é importante salientar que as primeiras ocorrências de *ter* como existencial em documentos brasileiros são detectadas na segunda metade do século XIX (Callou & Avelar 2003), mesmo período em que, de acordo com Duarte (1995), a freqüência na ocorrência do sujeito nulo de 3ª pessoa começa a declinar. Se, da mesma forma que a queda na freqüência de sujeitos nulos, o licenciamento de PPLOC em posição de sujeito for uma conseqüência da morfologia verbal empobrecida, podemos explicar o porquê de a emergência de *ter* como existencial ser temporalmente paralela aos primeiros sinais da queda na freqüência de sujeitos nulos.

A outra evidência está no paralelismo que se pode observar entre o PB e algumas línguas Bantu, no que concerne a uma possível correlação entre padrões de inversão locativa e morfologia verbal.¹⁰ No conjunto das línguas Bantu, as gramáticas do sesotho, do setswana e do kinyarwanda destacam-se por apresentar um sistema de concordância locativa empobrecido. Em chichewa e chishona, com exemplos em (32)-(33) abaixo, a marca de concordância locativa presente no verbo (o índice 17 em (a) e o índice 18 em (b)) é a mesma que se manifesta no elemento pré-verbal. Em sesotho, setswana e kinyarwanda, ao contrário, com exemplos em (34)-(36), a marca de concordância locativa no verbo é sempre a mesma (17 em sesotho e setswana, e 16 em kinyarwanda), independentemente da especificidade do índice presente no locativo pré-verbal. Esse contraste entre chichewa e chishona, de um lado, e sesotho, setswana e kinyarwanda, de outro, permite considerar que este último grupo tem um paradigma de concordância locativa extremamente simplificado.

(33) CHICHEWA

- a. ku-mu-dzi ku-na-bwér-á a-lendô-wo
 17-3-village SC17-PAST-come-FV 2-visitors-those
 'To the village came those visitors.'
- b. m-mi-têngo mw-a-khal-a a-nyãni
 18-4-tree SC18-PERF-sit-FV 2-baboons
 'In the trees are sitting baboons.'

(34) CHISHONA

- a. ku-ru-kova kw-áka-ón-ék-w-á shúmba né-um-kómaná
 17-11-river 17-PST-see-NEUT-PAS-IND 9:lion by-1-boy
 Lit.: 'At the river was seen a lion by the boy.'
- b. mu-marikete mú-no-téngés-éw-a mi-cheró
 18-market 18-PRS-sell-PAS-IND 4-fruit
 'In the market is sold fruit.'

(35) SESOTHO

- Mo-tsé-ng hó-tl-il-é ba-eti.
 3-village-LOC 17-come-PRF 2-travelers
 'To the village came the travelers.'

(36) SETSWANA

- Mó-le-fátshé-ng gó-fúla di-kgomo
 18-5-country 17SM-graze 10-cattle
 'In the country are grazing the cattle.'

¹⁰ Para uma discussão mais detalhada, ver Demuth & Mmusi (1997).

(37) KINYARWANDA

um	gisagára	ha-ra-rirímbir-a	aba-shyitsi
<i>in</i>	<i>7:village</i>	<i>16-PRS-sing-IPFV</i>	<i>2-guest</i>

Lit.: 'In the village are singing guests.'

(Salzmann 2004)

Os fatos do sesotho, do setswana e do kinyarwanda fazem lembrar, em alguns aspectos, o sistema de concordância atestado no PB, que leva o verbo a ocorrer preferencialmente na terceira pessoa do singular em casos de inversão locativa. Nos casos em (37) a seguir, por exemplo, que trazem locativos em posição pré-verbal no PB, os verbos não precisam ser flexionados para concordar com o sujeito argumental da construção, em posição pós-verbal. Na verdade, o emprego do verbo no singular soa bem mais natural que o seu emprego no plural; é possível que a manifestação de concordância seja, em tais casos, um fato da periferia gramatical, adquirido por meio de imitação da língua escrita culta, e não de uma propriedade da gramática nuclear.

- (38) a. Naquele quarto dorme(m) três crianças.
 b. Naquela loja trabalha(m) vários amigos meus.
 c. Na Unicamp estuda(m) dois filhos da Maria.

É importante frisar que línguas Bantu como o sesotho, o setswana e o kinyarwanda aceitam a inversão locativa com verbos inergativos e transitivos (desde que agente e tema não coocorram na mesma sentença – ver seção 2), da mesma forma que em PB. É possível, portanto, que exista algum tipo de relação entre morfologia empobrecida e a possibilidade de ocorrência de inversão locativa em sentenças inergativas e transitivas, tal como sugerido em Demuth & Mmusi (1997). Estaríamos, dessa forma, diante de indícios bastante fortes para a idéia de que as inovações em PB quanto à inversão locativa vêm no mesmo pacote de mudanças dos fatores que provocaram o enfraquecimento da concordância sujeito-verbo.

5. Consideração finais

Se as idéias aqui apresentadas forem confirmadas em pesquisas futuras, propriedades gramaticais de línguas Bantu poderão estar na

base tanto da simplificação do paradigma flexional quando das inovações que envolvem a inversão locativa no português brasileiro.

À guisa de conclusão, cabe ressaltar que este estudo não se encontra sozinho na tentativa de correlacionar as marcas de línguas ibéricas na América às propriedades mais gerais de línguas africanas. A abordagem de Negrão & Viotti (2008), por exemplo, reúne indícios de que a ergativização de verbos transitivos pode ser um reflexo da entrada maciça, em território brasileiro, de falantes de línguas africanas. No mesmo caminho, o estudo de Lipsky (2007) procura associar uma série de marcas encontradas no chamado *black Spanish* à gramática das línguas africanas. Os trabalhos publicados recentemente em Petter & Fiorin (2008), por sua vez, evidenciam que o debate sobre a influência das línguas africanas, ainda longe de uma visão consensual, pode ser tomado como um dos principais pontos da agenda de pesquisas, no século XXI, sobre a história e estrutura do português brasileiro.

Apesar de seu nível ainda bastante preliminar, este estudo sobre as “origens” da inversão locativa no português brasileiro pode, portanto, ser integrado a um debate mais amplo sobre os efeitos lingüísticos da entrada de africanos no continente americano, ao longo de quase quatro séculos. Olhar para as condições de formação do português brasileiro, tanto no plano intralingüístico quanto extralingüístico, é um elemento imprescindível ao sucesso desse debate.

Referências

- Avelar, J. 2006. *Adjuntos Adnominais Preposicionados no Português Brasileiro*. Tese de Doutorado. Instituto de Estudos da Linguagem – Universidade Estadual de Campinas.
- Avelar, J. 2008 [forthcoming]. On the Emergence of TER as Existential Verb in Brazilian Portuguese. A sair em: P. Crisma; G. Longobardi (Eds.). *Historical Syntax and Linguistic Theory*. Oxford: Oxford University Press.
- Avelar, J; Cyrino, S; Galves, C. 2008. *Locative inversion and agreement patterns: some parallelisms between Brazilian Portuguese and Bantu languages*. Apresentado durante o Special Wocal 6 (World Congress of African Linguistics), realizado em agosto/2008 na Universidade de São Paulo.
- Baker, M. 2003. Agreement, Dislocation, and Partial Configurationality. In: A. Carnie; H. Harley; M. Willie (Eds.). *Formal Approach to Function*. Amsterdam: John Benjamins, 107-134.

- Bresnan, J.; Kanerva, J. 1989. Locative Inversion in Chichewã. *Linguistic Inquiry*. **20**: 1-50.
- Callou, D.; Avelar, J. 2003. Estruturas com “ter” e “haver” em anúncios do século XIX. In: T. Alkmim (Ed.). *Para a história do português brasileiro – Vol. III*. São Paulo: Humanitas-USP, 47-68.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge MA: The MIT Press.
- Chomsky, N. 2000. Minimalist Inquiries: The Framework. In: R. Martin; D. Michaels; J. Uriagereka (Eds.). *Step-by-step: essays in minimalist syntax in honor of Howard Lasnik*. Cambridge MA: The MIT Press, 89-155.
- Chomsky, N. 2001. Derivation by Phase. In: M. Kenstowicz (Ed.). *Ken Hale: A life in language*. Cambridge MA: The MIT Press, 1-52.
- Demuth, K.; Mmusi, S. 1997. Presentational focus and thematic structure in comparative Bantu. *Journal of African Languages and Linguistics*. **18**: 1–19.
- Duarte, E. 1995. *A perda do princípio ‘Evite Pronome’ no português brasileiro*. Tese de Doutorado. Instituto de Estudos da Linguagem – Universidade Estadual de Campinas.
- Fernandez-Soriano, O. 1999. Two types of impersonal sentences in Spanish: locative and dative subjects. *Syntax*. **2(2)**: 101-140.
- Franchi, C.; Negrão, E.; Viotti, E. 1998. Sobre a gramática das orações impessoais com Ter/Haver. *D.E.L.T.A.* **14**: 113-140.
- Ferreira, M. 2000. *Argumentos nulos em português brasileiro*. Dissertação de Mestrado. Instituto de Estudos da Linguagem – Universidade Estadual de Campinas.
- Figueiredo Silva, M. C. 1996. *A posição sujeito no português brasileiro*. Campinas SP: Editora da Unicamp.
- Galves, C. 1996. O enfraquecimento da concordância no português brasileiro. In: I. Roberts; M. Kato (Eds.). *Português brasileiro – Uma viagem diacrônica*. Campinas SP: Editora da Unicamp, 387-408.
- Galves, C. 2001. *Ensaio sobre as gramáticas do português*. Campinas SP: Editora da Unicamp.
- Gonçalves, P.; Chimbutane, F. 2004. O papel das línguas Bantu na gênese do português de Moçambique: o comportamento sintático de constituintes locativos e direccionais. *Papia*. **14**: 7-30.
- Kato, M. 2000. The Partial Prodrop Nature and the Restricted VS Order in Brazilian Portuguese. In: M. Kato; E. Negrão (Eds.). *The Null Subject Parameter in Brazilian Portuguese*. Frankfurt: Vervuert – Latino-Americana, 223-258.
- Levin, B.; Rappaport Hovav, M. 1995. *Unaccusativity: at the Syntax-Lexical Semantics Interface*. Cambridge MA: The MIT Press.
- Lipsky, J. 2007. Afro-Yungueño Speech – The long-lost black Spanish. *Spanish in Context*. **4(1)**: 1-43.
- Marten, L. 2006. Locative inversion in Otjherero: more on morphosyntactic variation in Bantu. *ZAS Papers in Linguistics*. **43**: 97-122.

- Mendikoetxea, A. 2006. Unergatives that 'become' unaccusatives in English locative inversion: a lexical-syntactic approach. In: C. Cop; L. Gournay (Eds.). *Point de Vue sur l'Inversion. Cahiers de Recherche en Grammaire Anglaise de l'Énonciation*. Tome 9. Paris: Editions Orphys, 133-155
- Mok, S. 1992. Locative inversion in Cantonese. In: *Kansas Working Papers in Linguistics*. **17**: 115-157.
- Negrão, E. V.; Viotti, E. 2008. Estratégias de impessoalização no português brasileiro. In: M. Petter; J. L. Fiorin (Eds.). *África no Brasil: a formação da língua portuguesa*. São Paulo SP: Contexto, 179-203.
- Martins, A; Nunes, J. 2008 [forthcoming]. Syntactic Change as Chain Reaction: the Emergence of Hyper-Raising in Brazilian Portuguese. A sair em: P. Crisma; G. Longobardi. (Eds.). *Historical Syntax and Linguistic Theory*. Oxford: Oxford University Press.
- Petter, M.; Fiorin, J. L. (Eds., 2008). *África no Brasil: a formação da língua portuguesa*. São Paulo SP: Contexto, 179-203.
- Roberts, I. 2007. *Diachronic Syntax*. Oxford: Oxford University Press.
- Roberts, I; Roussou, A. 2003. *Syntactic Change: a Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.
- Rodrigues, C. 2002. Morphology and Null Subjects in Brazilian Portuguese. In: D. Lightfoot (Ed.). *Syntactic Effects of Morphological Changes*. Oxford: Oxford University Press, 160-178.
- Salzmann, M. 2004. *Theoretical Approaches to Locative Inversion*. Dissertação de Mestrado. Philosophical Faculty of the University of Zurich.

L'utilisation des anglicismes dans la langue du sport en français et leur variation

Mélanie Bernard-Beziade
melanie7979@yahoo.com

Université Michel de Montaigne, Bordeaux 3 (France)

RÉSUMÉ. Les langues romanes ont beaucoup évolué depuis quelques années. La mondialisation des échanges a entraîné un contact entre elles et les transformations sociales et technologiques induisent un changement linguistique: la mondialisation se manifeste dans tous les secteurs d'activité et par conséquent, elle ne peut laisser le vocabulaire qui s'y rapporte intact, des nouveaux mots sont nécessaires pour exprimer les changements sociaux. Dans cette étude, nous nous intéressons à ce qui se passe sur le plan linguistique dans le domaine du sport: les relations étroites entre la France et les pays anglo-saxons ont conduit à un fort développement des termes anglais dans la langue française et les anglicismes dans le domaine du sport occupent une place de plus en plus importante dans notre société actuelle. Depuis quelques décennies, le monde du sport a changé: de plus en plus, les sportifs sont des professionnels, et les médias jouent un rôle important dans la diffusion des sports. Par voie de conséquence, d'une part, la langue du sport est en train de changer, et, d'autre part, l'utilisation de termes propres au sport déborde largement du cadre des activités sportives. Cependant, tous les locuteurs n'ont pas les mêmes usages. En effet, la connaissance et / ou l'emploi de telle ou telle lexie contribuent à marquer l'appartenance à une catégorie sociale; tous les anglicismes ne sont pas utilisés de la même façon par les membres d'une même communauté. Partant de ce constat simple, nous avons voulu savoir de quelle manière se manifeste le changement linguistique dans le domaine du sport. Nous avons tout d'abord mené une recherche empirique en recueillant des lexies dans le quotidien sportif *L'Équipe*. Le choix du journal peut être discuté mais deux facteurs décisifs ont guidé notre sélection: tout d'abord l'histoire du quotidien démontre plus d'un siècle de rédaction et de reportages tournés vers le sport; de plus, les chiffres nous indiquent que c'est le journal le plus lu en France. La constitution du corpus s'est faite majoritairement durant les Jeux Olympiques d'été qui se sont déroulés à Athènes du 14 au 30 août 2004; nous avons recueilli 220 lexies définies comme des anglicismes et nous les avons datées. Cette datation permet une description diachronique de ces termes. Puis nous avons mené une enquête sociolinguistique auprès de 250 personnes réparties sur tout le territoire français, qui a permis de décrire les particularités qui régissent les pratiques langagières des locuteurs

en prenant en considération les variables sociales qui les caractérisent, de manière à déterminer dans quelle mesure la présence des anglicismes – dans le sport et dans les autres domaines d'activité – est elle-même en relation avec ces variables.

MOTS-CLÉS. Anglicismes, sport, utilisation, variation.

ABSTRACT. For the past few years, the Roman languages have evolved a lot. The globalization of exchanges triggered a contact between them and social and technological transformations inferred a linguistic change: globalization shows itself in all sectors and as a result it cannot leave the vocabulary unchanged. New words are necessary to express social changes. In this study, I am interested in what takes place from a linguistics' perspective in the field of sports: the narrow relationships between French and English speakers have led to a strong development of English terms in the French language. Anglicisms in the field of sports take a more and more important part in our society. For a few decades, the world of sports has changed: more and more sportsmen are professional, and mass-media play an important role in sports broadcasting. On the one hand, the language of sports is changing, on the other hand the use of words belonging to the sports area goes beyond the scope of sports activities. However, not all speakers have the same way of speaking. As a consequence, the knowledge and / or the use of such or such lexe contributes to stress the feeling of belonging to a social category: Anglicisms are not used by the members of the same community the same way. Through to this report, I wanted to analyse in which way the linguistic change expresses itself in the field of sports. First, I have made an empirical investigation, gathering lexes in sports newspaper *L'Équipe*. The choice of a newspaper can be discussed but two elements led to my choice: the history of the newspaper shows more than one century of report directed towards sports, and sales figures show that it is the most read newspaper in France. The gathering of the corpus took place during the Summer Olympics Games in Athens 14-30 August 2004; I studied 220 lexes referred to as "anglicisms" and dated them. This dating allows a diachronic description of these lexes. Then, I did a sociolinguistic survey among 250 people spread over all the French territory. Taking into consideration social variables, this allowed me to describe the particularities that rule the language experience of speakers, in order to assess to what extent the presence of anglicisms in sports and in other field of activities is related to these variables.

KEY-WORDS. Anglicisms, sport, use, linguistic variation.

Les langues romanes ont beaucoup évolué depuis quelques années et le changement linguistique est induit à la fois par la mondialisation

des échanges, qui a entraîné des contacts croissants entre elles, et les transformations sociales et technologiques. La mondialisation se manifeste dans tous les secteurs d'activité et de nouveaux mots sont nécessaires pour refléter les changements sociaux; par conséquent, le vocabulaire subit nécessairement des transformations.

Dans cette étude, nous nous intéressons aux changements en cours sur le plan lexical dans le domaine du sport: les relations étroites entre la France et les pays anglo-saxons ont conduit à un fort développement des termes anglais dans la langue française et les anglicismes dans le domaine du sport occupent une place de plus en plus importante dans notre société actuelle. Déjà, en 1939, MacKenzie écrivait:

[Bien] que les sports modernes se soient acclimatés très rapidement en France et y soient devenus un complément important de la vie, surtout dans les grandes villes, ils n'en sont pas originaires, et le plus léger coup d'oeil suffit à révéler l'origine étrangère d'une forte partie de leur terminologie. Tout écrit français qui traite de sport s'agrémente en effet de nombreuses tournures anglaises, indice permanent de l'influence exercée sur ce terrain par l'Angleterre. L'histoire de l'introduction dans n'importe laquelle des couches sociales de tel ou tel jeu dans un pays présente un réel intérêt. (p. 131)

Depuis quelques décennies, le monde du sport a changé: de plus en plus, les sportifs sont des professionnels, et les médias jouent un rôle important dans la diffusion des sports. Par voie de conséquence, d'une part, la langue du sport est en train de changer, et, d'autre part, l'utilisation de termes propres au sport déborde largement du cadre des activités sportives. Cependant, tous les locuteurs n'ont pas les mêmes usages. En effet, l'emploi de telle ou telle lexie contribue à marquer l'appartenance à une catégorie sociale; tous les anglicismes ne sont pas utilisés de la même façon par les membres d'une même communauté.

Partant de ce constat simple, nous avons voulu savoir de quelle manière se manifeste le changement linguistique dans le domaine du sport. Dans cette étude, nous analysons d'abord le contexte dans lequel les lexies ont été recueillies puis nous décrivons la façon dont l'enquête a été distribuée. Nous décrivons ensuite les particularités linguistiques en fonction des variables sociales choisies. Car selon Martinet (1974: 34-35), procéder à une description dynamique des

usages de la langue consiste à caractériser les «usages divergents comme des stades différents d'un même processus évolutif ou comme le résultat de divergences à partir d'un même stade initial, en s'efforçant, chaque fois, de dégager ce qui, dans un usage antérieur, peut contribuer à expliquer ce qui conduit à l'usage postérieur».

1. Le contexte

Afin d'étudier l'utilisation et la variation des anglicismes dans la langue du sport, nous avons tout d'abord mené une recherche empirique en recueillant des lexies dans le quotidien sportif *L'Équipe*. Le choix du journal peut être discuté mais deux facteurs décisifs ont guidé notre sélection: tout d'abord l'histoire du quotidien démontre plus d'un siècle de rédaction et de reportages tournés vers le sport; de plus, les chiffres nous indiquent que c'est le journal le plus lu en France: Charon (2003) dans *Les Médias en France* confirme que *L'Équipe* est le quotidien national le plus lu avec 2 495 000 lecteurs devant *Le Monde* (2 164 000 lecteurs).

Nous pouvons également nous interroger sur les raisons de l'utilisation des anglicismes (à la place de lexies française) par la presse française. De nombreux auteurs ont effectivement manifesté leur désapprobation vis-à-vis de l'utilisation des anglicismes; ce phénomène est fréquemment désigné sous l'appellation de «lutte contre le français», cependant il est généralement admis chez les linguistes que l'emprunt linguistique a toujours constitué une importante source d'enrichissement lexical et que les lexies empruntées reflètent les nouveaux concepts des locuteurs de la langue emprunteuse. Selon Dauzat (1952), une cause de l'entrée des anglicismes du sport dans la langue française peut être la concision et la rapidité qui les caractérisent. En effet, la langue anglaise possède de nombreuses ellipses, des troncations, des abréviations de mots trop longs, des amputations des composés avec la conservation d'un seul terme, et son vocabulaire est clair et concis. N'oublions pas que les Anglais ont été les premiers à codifier un certain nombre de sports et que ces derniers ont été empruntés avec le vocabulaire qui leur est propre dans diverses langues. En France, la Commission ministérielle de terminologie du sport créée en 1984 est chargée de trouver des «équivalents français»

aux anglicismes du sport mais la tâche est ardue et les dissensions fréquentes. Le professeur Guillerrou, membre de cette commission, résume par exemple une rencontre (cité dans Depecker, 1994): «Le mot *corner* avait été l'objet d'âpres discussions. L'un de nous – je ne sais plus qui – ayant proposé *coup de coin*, un journaliste fit observer: «Vous voyez d'ici un stade tout entier se mettre à crier *coup de coin*? ce serait une basse-cour à canards !» Mais la locution fut adoptée.» Mais «adopter» par la Commission de terminologie ne signifie pas «utiliser» par les locuteurs de français, c'est ce que nous essaierons de montrer dans cet article.

La constitution du corpus s'est faite majoritairement durant les Jeux Olympiques d'été qui se sont déroulés à Athènes du 14 au 30 août 2004. Nous avons donc établi un corpus de 220 lexies définies comme des anglicismes (c'est-à-dire des lexies dont le signifié et / ou le signifiant procède de la langue anglaise, quelle que soit leur variété géographique (anglais britannique, anglais américain...), utilisées par les journalistes de la presse sportive française et nous les avons datées. Cette datation permet une description diachronique de ces termes. Nous avons consulté de manière exhaustive toutes les pages du journal *L'Équipe* se rapportant aux Jeux Olympiques. Pour les deux sports non olympiques, nous avons consulté de manière sélective les pages se rapportant au Tournoi des VI Nations de rugby (du 17 au 29 mars 2004) et aux tournois de golf du *British Open* et de *L'Évian Masters* (du 16 au 24 juillet 2004). Le choix des Jeux pour huit sports (athlétisme, basket-ball, boxe, équitation, football, natation, tennis, voile) a été simple: c'est un événement important, suivi par des millions de spectateurs et téléspectateurs et nous pouvons par conséquent supposer que les lexies utilisées par les journalistes dans ce contexte le seront également par la population. Puis nous avons mené une enquête sociolinguistique auprès de 250 personnes réparties sur tout le territoire français, qui a permis de décrire les particularités qui régissent les pratiques langagières des locuteurs en prenant en considération les variables sociales qui les caractérisent, de manière à déterminer dans quelle mesure la présence des anglicismes est elle-même en relation avec ces variables.

2. L'enquête sociolinguistique

Les lexies ayant été recueillies dans les journaux, nous les avons soumises à la population dans le but d'analyser quelles étaient leur utilisation et leur variation dans la langue française. Dans ce travail, nous nous attardons sur 30 lexies (les plus récentes, c'est-à-dire celles qui ont été introduites dans la langue française dans la seconde moitié du XXe siècle) afin d'analyser s'il existe un changement linguistique en cours pour les lexies désignées. Nous savons que le changement a déjà eu lieu dans le domaine du sport pour d'autres anglicismes; par exemple, la lexie *goal*, troncation de l'anglais *goalkeeper*, très utilisée il y a quelques années, ne s'emploie guère plus, remplacée par gardien (de but).

Ce travail relève donc d'une étude sociolinguistique. Le sociolinguiste fait porter son attention sur le locuteur en tant que membre d'une communauté linguistique, en tant que sujet dont la langue peut caractériser l'origine ethnique, la profession, le niveau de vie, l'appartenance à une classe; nous savons que le chercheur doit tenter de mettre en évidence la variation d'un locuteur à l'autre (dans notre étude il s'agit de la variation dans le domaine lexical) et il doit essayer de montrer dans quelle mesure celle-ci est déterminée par des variables sociales. Comme le rappelle Gadet (2003: 7), «les façons de parler se diversifient selon le temps, l'espace, les caractéristiques sociales des locuteurs et les activités qu'ils pratiquent.» Le but de toute étude sociolinguistique est donc de décrire les particularités linguistiques qui régissent la langue d'un groupe déterminé de personnes en prenant en considération les facteurs sociaux; dans le cadre de notre étude, la présence des anglicismes dans le sport est elle-même due à des phénomènes d'ordre social. Nous nous intéressons à cinq variables sociales:

- le sexe: les femmes utiliseraient moins de lexies relatives au domaine du sport que les hommes (car elles pratiquent moins le sport que les hommes);
- l'âge: il jouerait un rôle dans l'utilisation des lexies du sport. En effet, les locuteurs plus âgés peuvent avoir été en contact avec plus de sports que des locuteurs plus jeunes, ils connaîtraient donc plus de lexies s'y rapportant et ils auraient également un

- souci plus important de la conservation de la langue française; ils utiliseraient donc aussi d'autres lexies;
- l'intérêt pour le sport: si un enquêté a un certain intérêt pour le sport, il utilisera un vocabulaire en adéquation avec chaque sport;
 - la pratique d'un sport: un article en ligne de Blin et Nouveau (1996) sur les Français et le sport¹ rappelle qu'à la fin des années soixante, 28% des Français de plus de 15 ans déclaraient pratiquer une activité sportive, de façon régulière ou irrégulière, et que ce pourcentage est passé à 38% au débuts des années quatre-vingts pour atteindre près d'un Français sur deux aujourd'hui. Ces pourcentages indiquent la place importante que tient le sport dans la société française.
 - lecture de publications se rapportant au sport: la presse sportive en France constitue un marché florissant. À côté des journaux spécialisés, tous les quotidiens nationaux et régionaux comportent une rubrique sportive. Nous supposons qu'un individu qui lit des magazines de presse sportive ou des journaux se rapportant au sport aura plus de chances d'utiliser les lexies spécifiques se rapportant à ce domaine qu'un individu qui ne lit pas la presse.

Afin de mener cette enquête sociolinguistique, nous avons administré un questionnaire par écrit. Ce mode de recueil des données possède aujourd'hui un caractère pratique non négligeable: la possibilité d'utiliser les nouvelles technologies comme Internet, et ce à moindre coût, l'administration du questionnaire s'est faite majoritairement par voie informatique (75%); l'autre moyen utilisé pour sonder l'échantillon a été le courrier traditionnel (25%).

Notre questionnaire est organisé en deux parties:

- 1) La première partie a permis de recueillir les données d'identification des répondants. Pour des raisons pratiques, certaines de ces données sont formulées sous forme de questions ouvertes, notamment en raison de la difficulté à faire figurer tous les

¹ Cet article est disponible à l'adresse suivante: <http://hcsp.ensp.fr/hcspi/docspdf/adsp/adsp-14/ad141819.pdf#search='les%20fran%C3%A7ais%20et%20le%20sport'>

cas possibles (dans les cas du type de profession, du lieu de résidence, du capital scolaire). Certaines données (comme l'âge, le sexe, le revenu) correspondent à des questions fermées pour deux raisons: soit il n'existe que deux cas possibles (le sexe) soit nous avons proposé une classification pour éviter de dévoiler des données que les personnes interrogées peuvent juger trop personnelles (comme le niveau de revenu ou l'âge).

- 2) La deuxième partie du questionnaire porte sur le corpus recueilli et interroge les locuteurs de français sur l'utilisation et la variation des lexies proposées. Les questions étaient les suivantes: *Quels sont [les mots] pour lesquels vous utilisez un autre mot (ou expression) pour exprimer la même chose? Quel est ce mot?* Les questions ont été élaborées de façon à permettre une vérification de la cohérence interne des réponses. Pour l'analyse de l'utilisation et de la variation, nous sommes donc en présence de trois modalités dans les réponses: certains enquêtés ont répondu *non*; certains enquêtés ont répondu *oui, j'utilise un autre mot* mais aucun mot n'était donné (nous l'avons nommée 'réponse positive'); d'autres ont répondu à l'affirmation *j'utilise un autre mot* et un autre mot était noté (nous l'avons nommée 'réponse effective'). Dans ce dernier cas, il existe également deux catégories de répondants: ceux qui proposent des variantes appropriées et ceux qui avancent des variantes qui ne concernent pas le champ sémantique du mot (ici principalement le domaine du sport) et qui témoignent ainsi soit de la méconnaissance du terme soit de son utilisation dans un autre champ sémantique. Les résultats concernant la question sur l'utilisation et la variation des lexies font appel à la compétence des individus, et plus spécifiquement à la disponibilité de leur vocabulaire dit «passif»². Quelquefois, la

² Un vocabulaire passif correspond à un vocabulaire comprenant des unités correctement interprétées par les locuteurs mais dont ils ne font pas forcément usage. Rappelons la définition de «lexique et vocabulaire» donnée par Arrivé (1986): «C'est sur [l'] hétérogénéité des usages que repose la notion de lexique individuel qui permet, notamment, de tenter d'estimer la compétence lexicale d'un locuteur. Mais la prise en compte de ses productions langagières (performances enregistrées dans un corpus) permet d'obtenir un *vocabulaire*, à partir duquel la notion de lexique est

distinction entre le vocabulaire actif et le vocabulaire passif est peu claire et dans notre enquête certaines personnes ont déclaré connaître des lexies et en utiliser d'autres, mais leurs réponses sont parfois 'hors sujet', c'est-à-dire qu'elles ne correspondent pas à la réponse attendue. Dans nos résultats, les réponses à cette question sont parfois non appropriées en raison d'une erreur d'interprétation, mais elles indiquent au moins au chercheur que les lexies concernées appartiennent aussi à d'autres domaines que celui dans lequel elles ont été relevées. Pour cette question, les personnes répondent donc sur ce qu'elles estiment faire avec le lexique disponible, ou sur ce qu'elles croient qu'il faudrait faire dans les situations linguistiques proposées: quelquefois on croit connaître les lexies mais en réalité on ne les connaît pas vraiment, ce qui nous renvoie à la notion de compétence³ que nous devons différencier de la notion de performance⁴. N'oublions pas que la variation permet à l'individu d'exprimer son identité en fonction de valeurs sociales mais également en fonction de sa conscience métalinguistique.

3. Les résultats

Les 250 enquêtés ont répondu 282 fois positivement à l'affirmation «j'utilise un autre mot»; à la question «quel est ce mot?» ils ont produit 261 réponses (*tokens*); et le corpus de variantes⁵ produit est de 106 réponses différentes (*types*). Les variantes citées par

difficilement inférable. On considère, en effet, qu'un locuteur n'utilise qu'une partie des mots qu'il est capable de comprendre, d'où la distinction entre un *vocabulaire actif* et un *vocabulaire passif* aux limites extrêmement floues.» (p. 378)

³ La notion de compétence désigne l'ensemble des aptitudes linguistiques que tout locuteur d'une langue est censé posséder et qui le rendent capable de produire ou de comprendre un nombre infini de phrases.

⁴ La notion de performance se rapporte à la réalisation effective des énoncés dans des situations concrètes.

⁵ Deux unités linguistiques qui peuvent être substituées l'une à l'autre sans qu'il y ait de différence dans le sens dénotatif du mot sont appelées variantes (Dubois, 2001).

les enquêtés ne sont quelquefois pas acceptables du point de vue sémantique. Certaines variantes sont ‘fantaisistes’ par exemple, pour *wishbone*, la seule variante proposée est «haricot magique»; cette variante doit avoir un sens pour la personne enquêtée mais nous ne savons pas à quoi elle correspond et ne pouvons pas l’expliquer. Il est intéressant de noter qu’aucune variante n’a été donnée pour la lexie *bogey*. Ce terme est spécifique au domaine du golf et si la lexie n’est pas connue par les enquêtés, ils sont dans l’impossibilité de donner une variante.

Nous constatons tout d’abord que l’écart entre la réponse à «j’utilise un autre mot» et «quel est ce mot» est d’environ 9%, ce qui signifie que certains enquêtés déclarent utiliser une variante mais n’indiquent pas quelle est celle-ci. Ce résultat renvoie clairement à la différence entre les deux notions auxquelles nous avons déjà fait référence: la notion de compétence et celle de performance.

Le Tableau 1 présente les résultats globaux de l’enquête.

Tableau 1. résultats globaux

Lexie	Réponses positives	Réponses effectives	Occurrences	Variantes	
Antidopage	1	1	1	Antidoping (1)	
Un birdie	5	4	3	Oiselet (2) Oiseau (1)	Expression golf (1)
Un bogey	0	0	0		
Un break	41	38	9	Pause (24) Temps mort (4) Arrêt (3) Coupure (2) Avantage (1)	Coupe (1) Point gagné sur service adverse (1) Pose (1) Temps arrêt (1)
Un chip	10	10	6	Jeton (4) Puce (2) Pièce (1)	Coup d’approche roulé (1) Processeur (1) Chop (1)
Un cut	5	3	3	Découpage (1) Faire la coupe (1)	Arrêt (1)

Lexie	Réponses positives	Réponses effectives	Occurrences	Variantes	
Un débriefing	15	14	8	Compte-rendu (5) Bilan (2) Réunion (2) Mise au point (1)	Décorticage après séance (1) Réunion de conclusion (1) Réunion de synthèse (1) Réunion finale (1)
Dopant	1	1	1	Produit illicite (1)	
Un dunk	8	8	3	Smash (5) Trempe (2)	Panier en force (1)
Un fan	17	15	8	Supporter (6) Groupie (2) Ventilateur (2) Admirateur (1)	Amateur (1) Amateur excité (1) Fanatique (1) Supporter inconditionnel (1)
Un flanker	9	9	4	Troisième ligne (4) Troisième ligne aile (3)	Ailier (1) Numéro 7 (1)
Indoor	43	41	6	Intérieur (25) En salle (12) Dedans (1)	Fenêtre (1) Interrogatoire (1) Sous abri (1)
Un maul	12	11	3	Regroupement (5) Mêlée (5)	Cocotte (1)
Un negative split	4	3	3	Course en rattrapage (1)	Pente négative (1) Répartition de l'effort (1)
Un performer	4	3	3	Habitué aux bons résultats (1)	Champion (1) Performeur (1)
Un performeur	3	3	3	Performant (1) Athlète (1)	Gagnant (1)
Une performeuse	1	1	1	Performante (1)	
Un play-off	6	6	5	Mort subite (2) Tournoi final (1) Phase finale (1)	Accessit (1) Dernier carré (1)
Un practice	9	9	5	Terrain d'entraînement (3) Entraînement (2)	Exercice (2) Stage (1) Terrain d'exercice (1)
Un putter	2	1	1	Fer droit (1)	
Un ruck	3	3	2	Mêlée spontanée (2)	Se faire piétiner (1)

Lexie	Réponses positives	Réponses effectives	Occurrences	Variantes	
Un scramble	3	3	3	Alerte (1) Vegas (1)	Cake (1)
Un sponsor	11	10	4	Mécène (6) Partenaire (2)	Bailleur de fonds (1) Investisseur (1)
Un sponsoring	7	6	3	Mécénat (4) Partenaire (1)	Investisseur (1)
Une sprinteuse	5	3	1	Coureuse (3)	
Une standing-ovation	25	23	8	Acclamation (5) Ovation (4) Ovation debout (4) Applaudissements (4)	Ola (3) Tout le stade se lève (1) Clameur (1) Triomphe (1)
Un tie-break	13	13	4	Jeu décisif (10) Manche décisive (1)	Pause (1) Temps mort (1)
Un wedge	3	3	2	Fer (2)	Angle (1)
Une wild-card	15	15	2	Invitation (14)	Élément imprévu (1)
Un wish-bone	1	1	1	Haricot magique (1)	
30 Lexies	282	261	106		

Nous avons récolté 250 questionnaires, les variables sociales qui nous intéressent sont réparties de la façon suivante:

3.1. Sexe

En ce qui concerne la distribution par sexe de nos enquêtés, nous pouvons observer une proportion quasi-égale d'hommes et de femmes, comme le montre le Tableau 2.

Tableau 2. Distribution des enquêtés par sexe

Sexe	Nombre d'enquêtés	%
Hommes	130	52
Femmes	120	48
Total	250	100

Les hommes ont répondu 180 fois positivement à l'affirmation «j'emploie un autre mot»; ils ont produit 167 réponses et 76 réponses différentes (soit un rapport *type / token* de 45,5% et une moyenne de 1,3 variante par personne). Les femmes ont répondu 102 fois positivement à cette affirmation; elles ont produit 94 réponses et 30 réponses différentes (soit un rapport *type / token* de 31,9% et une moyenne de 0,78 variante par personne).

Les femmes utilisent donc moins de variantes que les hommes dans le domaine du sport. Ce résultat paraît logique puisque nous avons déjà indiqué que les femmes s'intéressent moins au sport que les hommes, nous avons établi qu'elles connaissent moins le vocabulaire proposé dans le corpus que les hommes. Ces deux éléments font que les femmes sont moins performantes quand il s'agit de produire des variantes car il est évident que pour donner une variante de la lexie, il faut savoir ce qu'elle signifie.

3.2. Âge

Concernant la distribution par rapport à la variable âge, la tranche d'âge «plus de 55 ans» est la moins représentée; cela est probablement dû au moyen utilisé dans l'administration du questionnaire (à l'aide de l'outil informatique). Le Tableau 3 montre cette distribution.

Tableau 3. Distribution des enquêtés en fonction de l'âge

Tranche d'âge	Nombre d'enquêtés	%
18-34 ans	93	37,2
35-54 ans	96	38,4
+ de 55 ans	61	24,4
Total	250	100

D'une génération à l'autre, on peut, en règle générale, constater une certaine variation dans utilisation du vocabulaire. L'analyse du comportement linguistique par tranche d'âge permet de vérifier si cette variation se manifeste dans le domaine du sport et de dire si nous sommes en présence d'un changement linguistique en cours. Pour cela, il faut examiner si une même variante se manifeste à l'intérieur d'un certain groupe d'âge alors qu'elle ne le fait pas chez un autre.

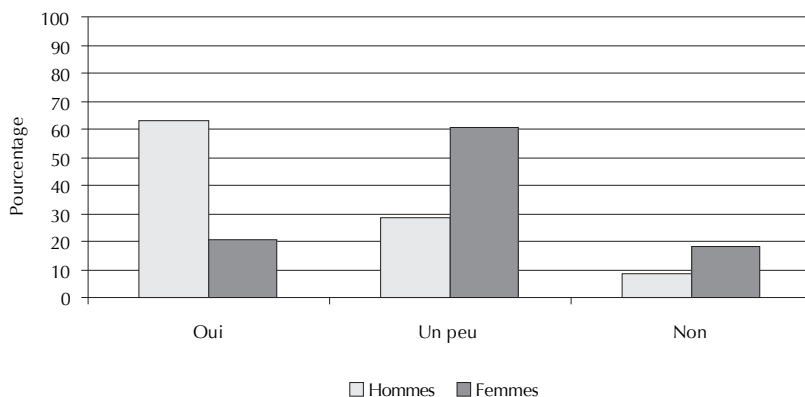
Les 93 enquêtés de la tranche des 18-34 ans ont répondu 114 fois positivement et ont produit 113 réponses et 51 réponses différentes (soit un rapport *type / token* de 45,1% et une moyenne de 1,2 variante par personne); les 96 enquêtés de la tranche des 35-54 ans ont répondu 119 fois positivement et ont produit 98 réponses et 23 réponses différentes (soit un rapport *type / token* de 23% et une moyenne de 1,02 variante par personne) et les 61 enquêtés de la tranche des plus de 55 ans ont répondu 61 fois positivement et ont produit 50 réponses et 32 réponses différentes (soit un rapport *type / token* de 64% et une moyenne de 0,80 variante par personne).

Il semble que la tranche des 18-34 ans utilise plus de variantes que les autres tranches d'âge. Ces résultats peuvent être vus comme une tendance en ce qui concerne un possible changement linguistique en cours, mais il faudrait analyser de façon qualitative les réponses des enquêtés afin de déterminer si les variantes utilisées par un certain groupe d'âge ne le sont pas par un autre.

Le fait que les locuteurs des tranches d'âge comprises entre 35 et plus de 55 ans utilisent moins de variantes que la tranche des 18-34 ans est contraire à notre hypothèse. Suite aux modifications qui sont intervenues dans le vocabulaire du sport au fil des ans, nous pensions que les locuteurs plus âgés avaient adopté le nouveau vocabulaire suite aux différents changements linguistiques qui ont pu avoir lieu (notamment dans le domaine du sport où les règles changent assez souvent pour pouvoir s'adapter aux nouveaux besoins des sportifs) tout en conservant l'ancienne terminologie.

3.4. Intérêt pour le sport

Selon notre enquête, seules 13,2% des personnes interrogées déclarent ne pas s'intéresser au sport. Ce pourcentage confirme la place importante que le sport tient dans la société française actuelle: 44% déclarent s'intéresser un peu au sport et 42,8% s'y intéressent vraiment. Les hommes s'intéressent nettement plus au sport que les femmes (63,1% des hommes de notre échantillon contre 20,8% des femmes).

Figure 1. Intérêt pour le sport

Pour les 33 enquêtés qui ne s'intéressent pas au sport, il y a 31 réponses positives; ils ont produit 26 réponses effectives et 17 variantes différentes (soit un rapport *type / token* de 65,3% et une moyenne de 0,78 variante par personne).

Les 110 enquêtés qui s'intéressent un peu au sport ont répondu 98 fois positivement et ont produit 87 réponses effectives et 39 variantes différentes (soit un rapport *type / token* de 44,8% et une moyenne de 0,79 variante par personne).

Les 107 enquêtés qui s'intéressent au sport ont répondu 153 fois à cette question; ils ont produit 148 réponses effectives et 50 variantes différentes (soit un rapport *type / token* de 33,8% et une moyenne de 1,38 variante par personne).

Il est normal de constater que les enquêtés qui prétendent s'intéresser au sport ont produit plus de variantes que les enquêtés qui prétendent n'avoir aucun ou peu d'intérêt pour le sport.

Le phénomène de la variation linguistique est donc bien présent dans le groupe social des personnes qui s'intéressent au sport.

3.5. Pratique du sport

Les réponses des enquêtés indiquent que 56% d'entre eux pratiquent un sport. La pratique effective d'un sport n'est pas aussi prononcée que l'intérêt général porté au sport. Nous voyons donc que les Français se déclarent assez sportifs mais que peu pratiquent

dans un club avec une licence et les contraintes que cela implique. Les hommes pratiquent plus d'activités sportives que les femmes (pour notre échantillon, 60,8% des hommes contre 50,8% des femmes).

Les 110 enquêtés qui ne pratiquent pas de sport ont répondu 105 fois à cette question; ils ont produit 94 réponses effectives et 48 variantes différentes (soit un rapport *type / token* de 51,1% et une moyenne de 0,85 variante par personne).

Les 140 enquêtés qui pratiquent un sport ont répondu 117 fois à cette question; ils ont produit 167 réponses effectives et 58 variantes différentes (soit un rapport *type / token* de 34,7% et une moyenne de 1,19 variante par personne).

Dans les deux ensembles de résultats ci-dessus, la moyenne de variantes produites est bien supérieure lorsque les enquêtés pratiquent un sport, et nous pouvons remarquer que la variation linguistique est donc fonction de la pratique d'un sport.

3.6. Lecture de publications se rapportant au sport

Les enquêtés lisent peu la presse sportive spécialisée: seuls 32,4% affirment lire des publications consacrées au sport et parmi ceux-ci 38,8% lisent au moins *L'Équipe*. Rien de bien étonnant quand on sait que ce journal est le plus diffusé en France (d'où notre choix de travailler sur ce quotidien). Les hommes lisent plus la presse spécialisée que les femmes (47,7% des hommes de notre échantillon contre 15,8% des femmes).

Les 81 enquêtés qui lisent des publications en rapport avec le sport ont répondu 136 fois à cette question; ils ont produit 135 réponses effectives et 58 variantes différentes (soit un rapport *type / token* de 42,9% et une moyenne de 1,6 variante par personne).

Les 169 enquêtés qui ne lisent pas de publications ont répondu 146 fois à cette question; ils ont produit 126 réponses effectives et 48 variantes différentes (soit un rapport *type / token* de 38,1% et une moyenne de 0,74 variante par personne).

Les pratiques de lecture semblent donc être une variable sociale liée à la variation linguistique: un enquêté qui lit des publications en relation avec le sport serait donc capable de produire plus de variantes, son vocabulaire disponible dans le domaine du sport étant plus important.

4. Conclusion

Selon les résultats présentés ci-dessus, les cinq variables sociales analysées rendent compte de la variation dans l'utilisation et la variation des lexies.

Les hommes, les jeunes (tranche d'âge comprise entre 18 et 35 ans), les personnes présentant un intérêt pour le sport, le pratiquant et lisant des publications en rapport avec ce domaine utilisent et proposent un nombre de variantes plus important que les autres groupes sociaux analysés.

Nous savons également que l'attitude du locuteur dans l'utilisation des variantes est liée à l'acceptation ou au rejet de l'anglicisme, c'est-à-dire au rapport que le locuteur entretient avec sa langue; et également à sa conscience métalinguistique (utilisation de variantes préconisées par la Commission de terminologie du sport). S'il est communément admis chez les linguistes que la véritable synonymie n'existe pas, que deux lexies ne sont jamais complètement identifiables du point de vue de leur signification car chaque lexie a sa valeur propre et ses usages, le locuteur ordinaire reconnaît la relation de synonymie entre deux formes lexicales différentes (qui se distinguent par leurs signifiants) mais de même sens (qui ont le même signifié) d'où son utilisation de variantes synonymiques. Cette étude sur l'utilisation et la variation des anglicismes extraits d'un journal sportif (le quotidien *l'Équipe*) et durant une période déterminée et contemporaine (les Jeux Olympiques d'Athènes de 2004) n'est menée que sur les anglicismes les plus récents; elle a cependant permis de contribuer à décrire de façon quantitative l'ampleur du changement linguistique en cours, ainsi que de déterminer dans quelle mesure des variables individuelles telles que le sexe, l'âge, l'intérêt pour le sport, la pratique du sport et les pratiques de lecture influent sur le choix des locuteurs dans leur choix d'utiliser des anglicismes et la variation qui y est attachée. Quant à l'étude qualitative de la variation utilisée, nous savons que le changement linguistique se réalise dans le temps, et la variation de la lexie empruntée étant une manifestation du changement linguistique, son appropriation prend aussi du temps.

Références

- Arrivé, M.; Gadet, F.; Galmiche, M. 1986. *La grammaire d'aujourd'hui: guide alphabétique de linguistique française*. Paris: Flammarion.
- Blin, P.; Nouveau, A. 1996. *Les Français et le sport*. <http://hcsp.ensp.fr/hcspi/docspdf/adsp/adsp-14/ad141819.pdf#search='les%20fran%C3%A7ais%20et%20le%20sport'> (Mars 1996)
- Charon, J.-M. 2003. *Les médias en France*. Paris: Éditions de la Découverte.
- Dauzat, A. 1952. *Voyage à travers les mots*. Paris: Bourrelief.
- Depecker, L. 2001. *L'invention de la langue: le choix des mots nouveaux*. Paris: Armand Colin.
- Deroy, L. 1956. *L'emprunt linguistique*. Paris: Les Belles Lettres.
- Dubois, J. et al. 2001. *Dictionnaire de linguistique*. Paris: Larousse.
- Gadet, F. 2003. *La variation sociale en français*. Paris: Ophrys.
- Laks, B. 1992. La linguistique variationniste comme méthode. *Langages*. **108(34)**: 34-50.
- Mackenzie, F. 1939. *Les relations de l'Angleterre et de la France d'après le vocabulaire I. Les infiltrations de la langue et de l'esprit français: les anglicismes français*. Paris: Droz.
- Martinet, A. 1974. *Le français sans fard*. Paris: P.U.F.

Relativização de Predicados Nominais na História do Português¹

Adriana Cardoso

adrianaprincipe@gmail.com

Centro de Linguística da Universidade de Lisboa (Portugal)

RESUMO. Este artigo apresenta uma análise de orações relativas do tipo (...) *ho s(enh)or dom ãRique governador Comprou a dona guyomar de meneses molher q(ue) foi de symão fogaça (...)*, que ocorrem em textos medievais portugueses (literários e não literários), mas que não são possíveis no português actual. O principal objectivo deste trabalho é o de caracterizar as propriedades sintáctico-semânticas destas construções, cuja existência não é mencionada nas gramáticas históricas do português nem nos estudos de referência de história da língua portuguesa.

Neste trabalho, propõe-se que estas orações relativas envolvem uma lacuna na posição de predicado da oração copulativa e que é esta intrigante particularidade que contribui para a construção das propriedades semânticas que estas construções exibem. Do ponto de vista da análise sintáctica, a não adjacência entre o PP e o núcleo nominal é explicada por uma proposta que combina a ‘análise por elevação’ das relativas restritivas com a teoria do movimento por cópia. Para a explicação da mudança relativa à possibilidade de adjacência entre o núcleo e o PP, propõe-se que o mecanismo de apagamento de partes de diferentes elos das cadeias (‘scattered deletion’) deixou de estar disponível em português, pelo menos ao nível do DP, contribuindo desta forma para a existência de relações menos transparentes entre estrutura sintáctica e estrutura informacional.

PALAVRAS-CHAVE. Orações relativas; relativização de predicados nominais; teoria do movimento por cópia; mudança sintáctica

ABSTRACT. This paper investigates relative clauses as in (...) *ho s(enh)or dom ãRique governador Comprou a dona guyomar de meneses molher q(ue) foi de symão fogaça (...)* (lit. ‘the Mr. Dom ãRique, governor, bought from Dona guyomar de meneses, woman that was of symão fogaça’), which occurred in literary and non-literary Medieval Portuguese texts, but are not possible in Modern Portuguese. The central goal of this paper is to describe and analyse the syntactic and semantic properties of these constructions, whose

¹ Este trabalho foi desenvolvido no âmbito da Bolsa de Doutoramento atribuída pela Fundação para a Ciência e a Tecnologia (SFRH/BD/22475/2005).

existence has not been mentioned in the Historical Portuguese Grammars nor in the reference works on the History of Portuguese Language.

The claim is that these relative clauses have a gap in a postcopular position (predicate of the copular sentence) and that this striking particularity contributes to the specific semantic properties that they exhibit. As far as the syntactic analysis is concerned, the non-adjacency between the head noun and the PP is explained by the combination of the 'raising analysis' of relative clauses and the copy theory of movement. As for the syntactic change w.r.t. the non-adjacency between the head noun and the PP, it is claimed that the mechanism that allowed the realization of different pieces of different chain links ('scattered deletion') is not allowed anymore, at least at the DP level. This change can explain why in Modern Portuguese we find a less transparent relation between syntax and information structure.

KEY-WORDS. Relative clauses; relativization of nominal predicates; copy theory of movement; syntactic change

0. Introdução

O principal objectivo deste trabalho é o de analisar as propriedades sintáctico-semânticas de uma construção de relativização que foi pela primeira vez descrita para o português por Costa (2004), mas que não é referida nas gramáticas históricas nem nos estudos de história da língua portuguesa. Pretende-se ainda, tendo como referência o quadro teórico da gramática generativa, apresentar uma análise sintáctica para esta construção.

O *corpus* considerado neste estudo é constituído por textos não literários:

- séc. XIII a XVI (1^a metade) – documentos notariais editados por Martins (1994, 2001).²
- séc. XVI (2^a metade) – documentos relativos à Inquisição publicados em Pereira (1987);
- séc. XVII (1^a metade) – documentos relativos à Inquisição publicados em Coelho (1987);

² Estes documentos estão inseridos no *Corpus Informatizado do Português Medieval* (CIPM), disponível em: <http://cipm.fcsh.unl.pt>. A transcrição e a referência dos textos corresponde a esta versão digitalizada (correspondendo a sigla CHP aos textos editados em Martins (1994)). Neste estudo serão ainda analisados os textos que foram introduzidos na edição de Martins (2001).

- séc. XVII (2ª metade) – documentos relativos às sessões inquisitoriais do Padre António Vieira, publicadas em Muhana (1995: 47-174).

1. Problema

No decurso da história do português, ocorre com alguma frequência a construção exemplificada em (1) e (2):

- (1) ho s(enh)or dom ãRique governador Comprou a dona guyomar de meneses [mulher q(ue) foi de symão fogaça] toda a ffazenda (e) foros (e) paues q(ue) ella tem na dita vylla daueiras (CHP210 1544)
- (2) A q(ua)l uïha tem [A mulher q(ue) foy d(e). lourêçe Anes] ãplazada do d(i)to Mom(steiro). (CHP074 1310-1311)

Nestas sequências, encontra-se destacada uma expressão nominal que contém uma oração relativa introduzida por *que*. Numa posição imediatamente adjacente à esquerda de *que* ocorre um nome que é o antecedente da oração relativa. No interior da oração relativa, a posição relativizada é a de predicado da oração copulativa.

Por esta razão, esta construção será doravante referida como Relativa de Predicado Nominal (RPN), um termo que pretende reflectir a particularidade mais intrigante desta construção, que será apresentada com mais detalhe na secção 4.

O problema que se coloca é então o de que, em português actual, sequências estruturalmente equivalentes a estas não são gramaticais:³

- (3) *Encontrei a Rita, [mulher que foi do Rui].
- (4) *Encontrei [a mulher que foi do Rui].

³ Rosa Virgínia Matos e Silva e Clara Barros questionaram a não ocorrência destas construções no português actual, nomeadamente em textos jurídicos. Ainda que não tenha procedido até ao momento à análise de textos jurídicos actuais, é importante salientar que esta construção não ocorria apenas em textos notariais, mas também em textos literários e em textos não literários pertencentes a outras tipologias (cf. secção 3). Como tal, a eventual ocorrência destas construções em textos jurídicos modernos poderá apenas corresponder a uma ‘fossilização’ de uma construção que outrora foi produtiva na língua, mas que actualmente não ocorre noutras tipologias textuais nem é aceite pela generalidade dos falantes.

Tendo em conta estes constrastes, neste trabalho procurar-se-á responder às seguintes questões: (i) Quais são as propriedades sintáctico-semânticas da RPN?; (ii) Qual a sua estrutura sintáctica?; (iii) Como explicar a agramaticalidade destas construções no português actual?

2. Dados comparativos

Para o castelhano e para o galego encontram-se atestadas construções equivalentes às RPN em análise neste trabalho. Estas construções ocorrem em estádios anteriores do castelhano (cf. (5)), mas também em castelhano actual, sobretudo em registos escritos formais (cf. (6)):⁴

- (5) la Reina de Hungria ... [muger que habia sido del buen rey Matias de Hungria]. (Gen 425, 12, *apud* Keninston 1937: 173, séc. XVI)
- (6) Gallo, [gobernador civil que fue de Barcelona], llevaba muchos años retirado de la política. (Brucart 1999: 489)

No *Corpus de Referencia do Galego Actual*⁵, também foi possível encontrar algumas atestações de RPN (cf. (7)). É de notar, contudo, que os falantes consideram tratar-se de uma construção muito formal, que nunca é utilizada em situação de fala espontânea.

- (7) Todo llo teño que agradecer a don Tomás Dapena Espinosa, [alcalde que foi de Betanzos], que informou que eu xa de pequeno tiña unhas tendencias que non eran normais. (Ruibal, E. 1994. *De corpo enteiro*. Vigo: Edicións Xerais de Galicia.)

⁴ No castelhano actual, estas estruturas têm as seguintes propriedades: (i) a expressão nominal que contém a relativa nunca ocorre em posições argumentais mas sim em ‘incisos parentéticos’ (Brucart 1999: 489); (ii) o PP não ocorre em adjacência relativamente ao nome (**Flórez, cónsul de Colombia que fue ...*) (Gutiérrez Ordóñez 1989: 260); (iii) o seu valor aspectual é perfectivo, dado que a construção não permite tempos verbais que expressem uma qualidade vigente no momento da enunciação (**gobernador civil que es de Barcelona*) (Brucart 1999: 489); (iv) a construção não expressa conteúdo de tipo circunstancial (causal, temporal, concessivo) e, por isso, não se encontra habitualmente em posição inicial de frase (Gutiérrez Ordóñez 1989: 260).

⁵ Disponível *on-line* em: <http://nunes.cirp.es/corgaxml/>. Agradeço estes dados a Damián Suárez.

Em provençal antigo, esta construção também se encontra atestada:⁶

- (8) La comtessa Garsenda [moiller que fo del comte de Proenza] (Schultz-Gora 1924: § 211)

3. Dados relativos à história do português

O número de ocorrências de RPN no *corpus* em análise é apresentado no QUADRO 1:

Quadro 1. Ocorrências de RPN no *corpus*

Martins 1994, 2001				Pereira 1987	Coelho 1987	Muhana 1995
XIII	XIV	XV	XVI (1ª metade)	XVI (2ª metade)	XVII (1ª metade)	XVII (2ª metade)
11	27	23	7	5	1	1

Apesar dos limites cronológicos do *corpus* considerado neste estudo, é de notar que a consulta não sistemática de outra documentação permitiu apurar que as RPN se encontram documentadas, pelo menos, a partir do final do século XII e que ainda eram utilizadas no século XVIII:

- (9) domũ que est de [muliere que fuit de gueicha] (Doc. de 1175. Mitra de Braga, maço 1)⁷
- (10) Guiomar Prra. mer. meretris natural | da cide. de Lxa. fa. de João Pinto Pra. já defunto [criado q. foi do fi- | dalgo do Bomjardim, e de sua mer. Mna. Carvo.] (CARDS, nº 431, 1717, fl. 360r)⁸

A consulta de outra documentação permitiu ainda apurar que as RPN não ocorrem apenas em documentos legais, mas também em textos que se integram em outras tipologias:

⁶ Agradeço esta referência a Jan-Wouter Zwart.

⁷ Agradeço este dado a Ana Maria Martins.

⁸ *Cartas Privadas Arquivadas nos Cadernos do Promotor da Inquisição de Coimbra* – Projecto CARDS – Cartas Desconhecidas (www.clul.ul.pt/sectores/filologia/project_cards.php). Agradeço a Rita Marquilha a disponibilização desta documentação.

- (11) «Senhor cavalleiro, eu vos direi meu nome pois me tanto rogades. Eu ei nome Meraugis de Porlegues e som de Cornualha, [cavaleiro que nom som ainda de muita nomeada], ca nom ha muito que foi cavaleiro». (*Demanda do Santo Graal*, cap. CCLXXV, fol. 102v, séc. XV)⁹
- (12) E os nossos fezerão hũa hyda cõ elles, na qual Fernã Martiz de Vascomçellos, [neto que hera //(p.615) do mestre de Samtiago dom Mẽ Rodriguez], matou hũ mouro de cavallo (*Crónica do Conde D. Pedro de Meneses*, Livro II, cap. XXXVI, 629, final do séc. XV)¹⁰

Este é um dado muito relevante, dado que exclui a hipótese de as RPN ocorrerem apenas em expressões fixas e fossilizadas, frequentes nas ‘partes formulaicas’ dos documentos legais (Sabatini 1968).

4. Propriedades das RPN

4.1. Lacuna em posição de predicado nominal

As RPN distinguem-se de outras orações relativas na medida em que a lacuna se encontra na posição de predicado nominal, como se ilustra em (13) e (14):

- (13) ho s(enh)or dom ãRique gouernador Comprou a dona guyomar de meneses [molher q(ue) foi __ de symão fogaça] toda a ffazenda (e) foros (e) paues (CHP210 1544)
- (14) A q(ua)l uïha tem [A molher q(ue) foy __ d(e). lourçe Anes] êplazada do d(i)to Mom(steiro). (CHP074 1310-1311)

O verbo que ocorre é tipicamente o verbo copulativo *ser*¹¹, que selecciona uma oração pequena. O sujeito dessa oração não tem realização fonética e o predicado é uma expressão nominal, que corresponde à posição relativizada.

⁹ Piel, J. M.; Nunes, I. F. (Eds.) 1988. *Demanda do Santo Graal*. Lisboa: Imprensa Nacional – Casa da Moeda.

¹⁰ Brocardo, M. T. 1997. *Crónica do Conde D. Pedro de Meneses – Gomes Eanes de Zurara / edição e estudo*. Lisboa: FCG/JNICT.

¹¹ Existem porém algumas atestações de RPN com verbos com funcionamento transitivo predicativo em vez de copulativo:

(i) a q(ua)ll posse o d(i)to p(r)ioill ouue po(r) tomada p(er) pedra (e) t(e)rra (e) colmo (e) folhas daruores abryndo a po(r)ta do d(i)to logar (e) ffechando (e) emt(re)gando [o d(i)to p(ro)curado(r) q(ue) sse djzya __] a chause do d(i)to logar (CHP155 1437)

A relativização do predicado nominal é a característica mais intrigante destas construções, dado que as posições sintáticas que são tipicamente relativizadas são as de argumento (cf. (15)) ou de adjunto (16)), i.e., correspondem a posições não predicativas:

- (15) (e) meteu em posse o d(i)cto Giral do(mingu)iz d[as meyas das. d(i)ctas Casas q(ue) lhj el(e)s venderõ __] (CHP092 1353)
- (16) E mândo A uos Chañçeller do d(i)cto Senh(o)r Arçeb(is)po q(ue) seelledes [os stro(mento)s (e) as c(ar)tas das veedorías dos emp(ra)zamêtos (e) scambhos en q(ue) derdes a d(i)cta autoridade __] (CHP093 1356)

4.2. (Não) adjacência entre o nome e o PP

O nome que funciona como antecedente da RPN ocorre tipicamente associado a um PP. No *corpus* em análise, há apenas um exemplo em que tal não acontece:¹²

- (17) Johan coelhejro (e) váásqe est(evêe)z m(ar)notej'ro (e) Johan galego (e) ffrâncisco p(er)ez [p(or)t(e)jro q(ue) ffoj __] (CHP123 1341)

Contudo, a situação mais frequente é aquela em que os nomes se encontram associados a PPs.

No *corpus* analisado, até à 1ª metade do século XVI, o PP nunca se encontra em adjacência relativamente ao nome mas sim numa posição à direita do verbo copulativo:¹³

- (18) ho s(enh)or dom ãRique governador Comprou a dona guyomar de meneses [mulher q(ue) foi __ de symão fogaça] toda a ffazenda (e) foros (e) paues q(ue) ella tem na dita vylla daueiras (CHP210 1544)

¹² Na documentação consultada relativa ao século XVIII também se registou uma ocorrência deste tipo:

(i) Mto. Illustres Snres. Denuncia a V. sra. Maria Martins Va. de Mel. Alvres | lavrador nal. e moradora do lugar e frga. de Aguiar de Souza de Anta. solteira | mulher e a velha sua vezinha [secadeira q. foi] (CARDS, nº 563, 1720, fl. 240r)

¹³ A consulta de outros textos permitiu contudo encontrar uma atestação de adjacência entre o nome e o PP no século XIV:

(i) Vêo [o priol de nãdim que entõ era __]. E o de olueira.

(Brocardo, M. T. 2006. *Livro de Linhagens do Conde D. Pedro*. Edição do fragmento manuscrito da Biblioteca da Ajuda (século XIV). Lisboa: IN-CM, p. 136)

Porém, na segunda metade do século XVI e no século XVII registam-se ocorrências do PP em adjacência relativamente ao nome:

- (19) mandamos a[lo nosso recebedor do dito bispado que ora é __ e pelo tempo for __], e assim ao rendeiro da massa de Bragança, que é do dito bispado, (1581, Pereira 1987: 122)
- (20) para o que poderá ajudar muito [o Rei de Portugal, que então for __,] como fizeram seus antecessores em todas as quatro partes do mundo (1664, Muhana 1995: 80)

4.3. Função na matriz

O NP que contém a RPN ocorre tipicamente como aposto não restritivo relativamente a um DP. Este DP, por sua vez, é tipicamente um antropónimo (cf. (18)) ou um pronome pessoal que tem como aposto um antropónimo:

- (21) Conheçuda Cousa sseia a quantos esta. C(arta). viren e léér ouiren q(ue) Eu dona Brigueira ayras [molh(er) q(ue) foy __ de don Roy g(ar)cia d(e) pau||h||a.] (CHP041 1290)

Dado que, nas posições, uma expressão nominal estabelece com o antecedente uma relação predicativa, pode assumir-se que, nestes exemplos, o NP que contém a RPN é um NP predicativo (Gratacós 1999: 525). Ou seja, há coincidência entre a função sintáctica correspondente à subordinada e à matriz: a posição de relativização tem a função sintáctica de predicado e o NP que contém a RPN também é predicativo.¹⁴

¹⁴ No *corpus* analisado, estas construções podem também ser encabeçadas por *como*:

(i) Sabham todos q(ue) Eu M(ar)cos mjg(u)eez cõme [p(ro)curado(r) q(ue) ssoom __ da P(ri)oressa E donas do Moesteyro dachelas da p(ar) da Çidade de libbõa] emp(ra)zo A uos Joham vicent(e) (CHP132 1363)

(ii) E q(ue) pero elle d(i)cto auto(r) como [he(r)deiro q(ue) Era __ do d(i)cto p(er) o vallent(e) Ja finado p(er) cuja herança a elle esto todo p(er)teeçia] Requerera p(er) muytas vezes ao d(i)cto Reeo (e) ao d(i)cto seu avoo q(ue) manteuesem (e) comp(ri)sem o d(i)cto cont(ra)uto (CHP172 1422)

Nestes exemplos ‘como’ tem o significado equivalente a ‘na qualidade de’, ‘enquanto’ e introduz um constituinte não oracional. Por essa razão, alguns autores defendem que se trata de um ‘como’ preposicional, que selecciona elementos que são interpretados predicativamente (Magro 2007; Brucart 1999: 514).

Por fim, é de notar que, quando o NP que contém a RPN desempenha a função de aposto, surge por vezes coordenado a outro NP. Como se pode observar em (22)-(23), nesse caso, o NP que contém a RPN pode ocorrer no primeiro ou no segundo membro de coordenação:

- (22) dereito de hũa de manda que ho dito mosteiro (e) outras pessoas como herdeiros transuersaes ab intestado de dona ysabel de souza [mulher que foy __ de christouam de britto] (e) hirmã da dita dona ynes de melo (CHP211 1548)
- (23) nas cassas da morada de njcolosso maraboto/?/ morador na d(i)cta cidade a santo esp(ri)to estãdo hí ljanor ychoa sogra do d(i)cto njcolosso E [mulher que foy __ de diegaluarez de ualladares] (CHP187 1482)

Contudo, o NP que contém a RPN pode também ser argumental. No *corpus* analisado, as funções sintácticas que este constituinte pode desempenhar são as de sujeito (cf. (24)), de objecto directo (cf. (25)), de objecto indirecto (cf. (26)) e de complemento de preposição (cf. (27)).

- (24) A q(ua)l uíha tem [A mulher q(ue) foy __ d(e). lourçe Anes] ãplazada do d(i)to Mom(steiro). (CHP074 1310-1311)
- (25) Aos que esta nossa certidão virem digo eu André de Sousa, escrivão dante os corregedores do crime de Lisboa, que eu fiz notificar [os almotacés da almotaçaria que ora são], por mandado do corregedor Diogo da Fonseca (1572, Pereira (1987: 100))
- (26) mandamos a[o nosso recebedor do dito bispado que ora é e pelo tempo for], e assim ao rendeiro da massa de Bragança, que é do dito bispado, (1581, Pereira 1987: 122)
- (27) o qual Johã de Mérida logo pareceu (e) díssi q(ue) era tut(or) de [filha q(ue) ffoj __ de Johã lopez] (e) díssi q(ue) q(ue)ría s(e)r outor das d(i)ctas Casas (CHP022 1289)

Como se pode verificar, nestes casos não há coincidência entre a função sintáctica na matriz e a função sintáctica na subordinada: a posição de relativização corresponde a uma posição predicativa, enquanto a posição na matriz é argumental.

4.4. Determinação do antecedente

O antecedente das RPN só ocorre associado a um artigo nos casos em que o NP que contém a relativa desempenha a função de argumento na matriz.

- (28) as quaes agora tras sem tytolo (e) por comsemty(m)ento das ditas donas
[a molher q(ue) ffoy __ dallu(aro) f(e)r(nande)z] (CHP209 1540)

Porém, há contextos do mesmo tipo nos quais o antecedente não surge associado ao artigo:¹⁵

- (29) o qual Johã de Mérida logo pareceu (e) díssi q(ue) era tut(or) de [[filha
q(ue) ffoj __ de Johã lopez] (e) díssi q(ue) q(ue)ría s(e)r outor das d(i)ctas
Casas (CHP022 1289)
- (30) Nem vender alhear nẽ escanbar nẽ tresmudar O d(i)cto enp(ra)zamêto nẽ
part(e) delle ssem líçença E auctoridade da d(i)cta sua lg(re)ía E [abade
q(ue) aaquelle tempo for __ della] (CHP161 1472)

A ausência de artigo nos casos em que o NP que contém a relativa desempenha a função de argumento na matriz deixa de ocorrer em documentos posteriores: a partir da 2^a metade do século XVI, é sistemática a ocorrência de artigo nestes contextos.

No que diz respeito à construção da referência nominal, é de notar que estes NPs podem ter quer uma leitura referencial (cf. (28)-(29)), quer uma leitura atributiva (cf. (30)).

Porém, o que é surpreendente nestas RPN é que a introdução de um referente no discurso é feita em simultâneo com uma predicação. Ou seja, no português actual, o que seria de esperar numa construção equivalente era que houvesse primeiro a introdução de um referente e que, só posteriormente, se predicasse sobre ele. Seria então possível (31) mas não (32):

- (31) Encontrei a rapariga que foi mulher do Lourenço.
(32) *Encontrei a mulher que foi do Lourenço.

¹⁵ Parece contudo haver justificação independente para a ausência do artigo em (29)-(30), podendo manter-se que tipicamente as RPN argumentais incluem o artigo definido. Em (29), a ausência de artigo pode indicar que se trata de 'uma das filhas de *Johã lopez*' e não de 'uma filha única'. Em (30), *abade* pode ser o segundo membro de uma estrutura de coordenação introduzida por artigo definido (o qual poderia concordar apenas com o primeiro membro da estrutura de coordenação).

Conclui-se, assim, que nestes contextos, a lacuna da RPN corresponde a uma posição predicativa, enquanto o antecedente ocorre numa posição argumental. Desta combinação resulta a especificidade mais surpreendente destas relativas: a codificação simultânea de uma entidade e da predicação que recai sobre ela.

Por contraste, nos casos em que o NP é um aposto, este nunca ocorre com determinante (cf. (22)). Tal impossibilidade parece resultar do facto de o NP ter, nesses contextos, um valor predicativo (cf. Brucart 1999: 423).

4.5. Predicado nominal com valor de ‘etiqueta de classe’

Nas RPN, os nomes que correspondem ao predicado nominal relativizado designam tipicamente papéis, funções ou relações sociais e, por isso, são designados por alguns autores como ‘etiqueta de classe’ (Leborans 1999: 2371).

Como se pode observar no QUADRO 2, os predicados nominais que ocorrem nas RPN correspondem sobretudo a nomes de profissão e relações de parentesco.

Quadro 2. Nomes que ocorrem como predicados nominais nas RPN

nomes de profissões / cargos		nomes de parentesco	outros
prioll pryoresa clerigo abade dona padre freira	almotáce rebedor rei scripuã tabelliõ cryado (a) porteiro	procurador ouuydor veedor tesoureiro inquisidor	mulher filho (a,s) morador (e vizinho) herdeiro

Estes nomes caracterizam o sujeito pela sua pertença a uma determinada classe ou categoria na teia de relações que se estabelecem em sociedade. O sujeito é, como tal, distinguido por pertencer a um protótipo de classe ou categoria socialmente reconhecidos (Leborans 1999: 2371).

4.6. Copulativas de caracterização vs copulativas de identificação

Segundo Leborans (1999: 2368), as orações com *ser* copulativo podem ser classificadas em dois subtipos: copulativas de caracterização / copulativas de identificação¹⁶.

Nas construções em análise neste trabalho, ocorrem estes dois tipos de copulativas. Assim, as RPN que ocorrem em apostos contêm claramente copulativas de caracterização. Neste sentido, o predicado nominal predica sobre o sujeito uma característica: neste caso a pertença a uma determinada classe ou categoria na teia de relações que se estabelecem em sociedade (x foi abade de...).

Quando as RPN ocorrem em constituintes em posições argumentais, o predicado nominal tem sobretudo a função de ‘identificar’ o sujeito e não a de o caracterizar. Ou seja, nestas construções ocorrem copulativas de identificação, dado que a sua interpretação preferencial parece ser a de que ‘x foi a mulher de...’:

4.7. Tempo e aspecto

O tempo verbal do verbo copulativo nas RPN é tipicamente o pretérito perfeito. A selecção deste tempo verbal indica que determinada propriedade já não se aplica ao indivíduo, construindo-se desta forma um valor aspectual perfectivo. Por esta razão, Keninston (1937) afirma que esta construção transmite a noção de que a ‘classificação’ já não é verdadeira, significando “o último”, “o anterior” ou “ex-”.

Assim, uma sequência como (33)a poderia, no português actual, corresponder a (33)b:

- (33) a. ho s(enh)or dom ãRique gouernador Comprou a dona guyomar de meneses [molher q(ue) foi __ de symão fogaça] toda a ffazenda (e) foros (e) paues q(ue) ella tem na dita vylla daueiras (CHP210 1544)
 b. ... Comprou a *dona guyomar de meneses*, ex-mulher de *symão fogaça*, ... (dado construído)

¹⁶ Estas distinguem-se pela função semântica que o predicado estabelece com o sujeito e com a intenção e pressuposições do falante. Assim, nas copulativas de caracterização, uma expressão predica sobre o sujeito algum tipo de característica, permitindo a sua individuação no universo do discurso. Nas copulativas de identificação, estabelece-se uma relação de identidade ou identificação entre as duas expressões vinculadas por *ser* (Leborans 1999).

Contudo, ao contrário do que acontece por exemplo em castelhano (cf. Brucart 1999: 489), o verbo copulativo pode encontrar-se no presente do indicativo, expressando simultaneidade relativamente ao momento de enunciação (cf. (34)-(35)).¹⁷ Nestes casos, há a construção de um valor aspectual imperfectivo, na medida em que a propriedade ainda se aplica no momento da enunciação.¹⁸

- (34) q(ue) pore[m] lhe ap[er]ra)zya q(ue) b[er]asco m[er]art)jz [p[er]ri]oll q(ue) ora h[er]e ___ do d[er]i)to most[er]eiro] tome ou mande tomar e posse e tenças E demynyo E ssenhoryo dos d[er]i)tos casaaes (CHP155 1437)
- (35) Sabhã todos q[ua]ntos este st[er]o)mêto virẽ q[ue] eu Johã f[er]nãdez morador no burgo de Poõbeiro. (e) [filho q[ue] s[ã]o ___ de f[er]nã de st[er]ev)ez (e) de Mag[er]ari)da m[er]art)j(n)z] (CHP097 1365)

No *corpus* em análise, o verbo copulativo só ocorre no imperfeito do indicativo quando a construção é introduzida por ‘como’ (cf. (36)). Contudo, em Costa (2004) é apresentado um exemplo de RPN de aposto com o verbo no imperfeito (cf. (37)).

- (36) o d[er]i)cto fe[er]nam Ca[er]neiro q[ue] p[re]sente estaua Como [p[er]ro]Curador q[ue] era ___ do honrrado dom Joham uaa[s]q[ue]z p[er]ri)ol do m[er]osteiro) de villarynho da hordem de santo agostynho p[er] h[ui]a p[er]ro]curaÇõ q[ue] Eu sobr[e]d[er]i)cto t[er]abaliã) tenho Registada em meu liu[er]o) de notas sofiçient[er]e) p[er]ra) o q[ue] se adjante segue] djse q[ue] (CHP160 1467)
- (37) conteudos nas primeiras partições que se fezerom cõ Johã afomso tio e tetor que entõ era dos horffõs (1425, Costa 2004)

Nestes exemplos, o imperfeito apresenta características de tempo relativo a um ponto de perspectiva temporal do passado (Oliveira 2003: 157). Mais concretamente, verifica-se uma relação de inclusão com um tempo do passado, uma vez que, em (36), o acto de ‘dizer’

¹⁷ É de notar que esta possibilidade de ocorrência do verbo copulativo em outros tempos verbais sugere que as RPN que ocorrem em castelhano se encontram ‘fossilizadas’ (e por isso só admitem a ocorrência do pretérito perfeito do indicativo). Pelo contrário, no período da história do português em análise, esta construção pode ocorrer com outros tempos verbais, facto que aponta para um uso mais produtivo e menos marcado desta construção.

¹⁸ Como refere Oliveira (2004: 519, n. 10), a utilização do presente com estados transmite a informação de tempo presente relativamente ao ‘tempo da enunciação’, podendo alargar-se para além deste pelo facto de ser um estado.

(*djse*) está incluído no intervalo de ‘ser procurador’ e, em (37), o acto de ‘fazer’ (*fezerom*) também está incluído no intervalo de ‘ser tio e tutor’.

Regista-se ainda a ocorrência do verbo copulativo no futuro do conjuntivo (cf. (38)). Quando tal acontece, a predicação não é validada no plano enunciativo, mas é validável no plano potencial/fictício, num momento posterior ao da enunciação. Ou seja, no futuro, haverá uma entidade relativamente à qual se atribuirá a predicação ‘ser abade de...’.

- (38) Nem vender alhear nẽ escanbar nẽ tresmudar O d(i)cto enp(ra)zamẽto nẽ part(e) delle ssem lĩçença E auctoridade da d(i)cta sua lg(re)ia E [abade q(ue) aaquelle tempo for ___ della] (CHP161 1472)

Estes diferentes valores temporais e aspectuais são por vezes construídos mediante a co-ocorrência de adjuntos com valor temporal, compatíveis com os valores construídos no enunciado. Assim, quando o verbo copulativo se encontra no pretérito perfeito, no presente, no imperfeito e no futuro do conjuntivo podem co-ocorrer na RPN, respectivamente, localizadores temporais como *en outro tẽpo*¹⁹, *ora* (cf. (34)), *entõ* (cf. (37)) e *aaquelle tempo* (cf. (38)).

Esta possibilidade de delimitação da ‘duração da predicação’ (Oliveira 2004: 505) está intrinsecamente relacionada com o tipo de predicados envolvidos. De acordo com a terminologia de Carlson (1977), os predicados envolvidos nas RPN são de ‘indivíduo’ (*individual-level predicates*). Os predicados de indivíduo (por oposição aos predicados de ‘fase’), “(...) dizem respeito a propriedades estáveis e nessa medida caracterizam o indivíduo por um período suficientemente alargado de tempo, se não mesmo por toda a sua existência.” (Oliveira 2004: 511)

Nas RPN, os predicados de ‘indivíduo’ que ocorrem são tipicamente atribuídos de forma estável mas não permanente a um indivíduo (Oliveira 2004). Ou seja, dado o seu valor semântico de ‘categorias sociais virtualmente contingentes’ (Leborans 1999: 2372) (cf. secção

¹⁹ Como em (i):

(i) Sabahã/sic/ todos q(ue) eu Sancha do(mingu)jz [molh(er) q(ue) ffuy **en outro tẽpo** ___ de J(o)h(an)e amjgo mor[a]dor ã Lixbõa] Dou (e) out(or)go (CHP119 1329)

4.5.), estes predicados admitem uma delimitação temporal, ou seja, uma delimitação da ‘duração da predicação’.²⁰

4.8. Modalidade

No *corpus* verifica-se o recurso a uma estratégia através da qual o enunciador não assume o valor de verdade do conteúdo proposicional RPN:

(39) pareceu no d(i)to logo Johã vjçent(e) [p(ro)curador q(ue) dj'z q(ue) h(e) __; da P(ri)oressa (e) cõuento do M(osteiro) de chelas] da h[ua] part(e) (CHP146 1397)

Em (39), existe uma oração subordinada completiva encaixada na relativa *e*, como tal, trata-se de um caso de movimento relativo longo. Com a introdução desta oração, o enunciador atribui a asserção a *Johã vjçent(e)*, não se comprometendo com o seu valor de verdade. Trata-se, desta forma, de uma estratégia de distanciação e não comprometimento do enunciador relativamente ao valor de verdade da predicação.

5. Análise das RPN

5.1. Análise de Costa (2004)

Costa (2004) analisa esta estrutura como uma relativa apositiva, com o verbo *ser*, que envolve a extraposição de um PP. A hipótese avançada é a de que a estrutura oracional relativa com o verbo *ser* envolve um argumento que é uma categoria vazia. Esta relativa estaria posicionada em adjunção a DP, como modificador apositivo. De acordo com esta análise, seria ainda necessário pressupor

²⁰ Note-se que mesmo com um predicado de ‘indivíduo’ como ‘ser filho de...’, que à partida diria respeito a toda a existência do indivíduo, também se verifica a possibilidade de ocorrência do verbo *ser* no pretérito perfeito, tendo como efeito a interpretação de que os ‘pais’ já não são vivos:

(i) Sablha~/sic/ q(ua)ntos esta c(ar)ta virẽ q(ue) eu Johã ro(drigu)iz (e) eu Gil ro(drigu)iz (e) eu A(fonso) ro(drigu)iz (e) eu váásco ro(drigu)iz (e) eu Gonçalo ro(drigu)iz da Castanheyra [filhos q(ue) fomos __ d(e) Rodrigo. affonso (e) d(e) Giom(ar) ro(drigu)iz] (CHP025 1295)

a existência de uma adjunção recursiva para dar conta da posição extraposta do PP:

- (40) [_{DP} [_{XP} ffreira profressa [PP k]] [CP que [I é [VP [-]]]]] [PP do Moesteiro dachellas k]] (Costa 2004: 417)

Esta análise levanta porém os seguintes problemas: (i) a proposta de elipse de VP parece um pouco estipulativa, uma vez que este não é um contexto típico de legitimação desta estrutura²¹; (ii) de acordo com esta hipótese, *ffreira profressa* é o antecedente de uma relativa apositiva. Porém, a possibilidade de ocorrência do verbo copulativo no conjuntivo demonstra que se trata de uma relativa restritiva, visto que uma apositiva nunca admite o conjuntivo (cf. Brito & Duarte 2003: 673); (iii) se a posição de relativização fosse a de sujeito e o antecedente fosse sempre o constituinte nominal que precede *que*, então esperar-se-ia que o verbo da oração relativa se encontrasse sempre na 3ª pessoa (singular ou plural). Porém, tal predição não se verifica, dado que existem algumas ocorrências em que o verbo copulativo se encontra na primeira/segunda pessoa do singular ou plural. Tal facto demonstra que o antecedente não corresponde à posição de sujeito no interior da relativa:

- (41) Sablhã/sic/ q(ua)ntos esta c(ar)ta virẽ q(ue) eu Johã ro(drigu)iz (e) eu Gil ro(drigu)iz (e) eu A(fonso) ro(drigu)iz (e) eu váásco ro(drigu)iz (e) eu Gonçalo ro(drigu)iz da Castanheyra [filhos q(ue) fomos __ d(e) Rodrigo. affonso (e) d(e) Giom(ar) ro(drigu)iz] damos (e) outorgamos (CHP025 1295)
- (42) me obrígo p(er) my´ E p(er) todos me(us) bêes assy´ mouíjs como de Rajz auõdos E por aue(r) aa dar E paga(r) a uos v(asco) nogey´ra escudeyro morador na d(i)cta çidade e [p(ro)curador q(ue) sodes __ do moestey´ro da çhellas] mj´ll (e) duzentos Reaes bran(cos) (CHP176 1436)

²¹ Como é referido em Matos (2003: 891), a elipse de VP pode ocorrer com verbos copulativos. Contudo, nestes casos, há sempre um elemento verbal antecedente, com o qual o verbo que legitima o constituinte elíptico estabelece um paralelismo lexical, como se observa nos seguintes exemplos:

(i) Esse cão é um perdigueiro, mas aquele não é.

(ii) Estamos desejosos de saber os resultados dos exames e pensamos que eles também estão.

(Matos 2003: 891-892).

Ora é precisamente a ausência de um predicado que funcione como antecedente que enfraquece a hipótese de as RPN envolverem elipse de VP.

5.2. Proposta de análise²²

A análise que proponho para estas construções assenta nos seguintes pressupostos: (i) o nominal imediatamente adjacente à esquerda de *que* é o antecedente da relativa; (ii) a posição relativizada é a de predicado nominal da oração pequena; (iii) as RPN são relativas restritivas e não apositivas.

5.2.1. A representação das relativas restritivas

Muito do debate acerca da sintaxe das orações relativas tem-se centrado na oposição entre duas análises: a análise ‘standard’ (envolvendo adjunção) e a análise por ‘elevação’ (envolvendo o movimento do núcleo a partir de uma posição interna à oração relativa).

De acordo com a análise ‘standard’ (e seguindo, por exemplo, a proposta de Demirdache 1991), a representação de uma relativa como (43) seria a apresentada em (44). Em (44), o antecedente da relativa é gerado numa posição externa e, como tal, o PP é extraposto (i.e., movido para a direita), para uma posição de adjunção à direita do DP. Nesse caso, as RPN em análise envolveriam extraposição no interior do DP.

(43) filhos q(ue) fomos d(e) Rodrigo

(44) [_{DP} [_{NP} [_{NP} filhos [_{PP} t_i]] [_{CP} Op_k [_C que [_{IP} pro_j [_{V'} fomos: [_{VP} [_{V'} t_i [_{SC} t_j t_k]]]]]]]]]] [_{PP} de Rodrigo]]

Pelo contrário, a adopção de uma análise por ‘elevação’ do núcleo, tal como proposta por Kayne (1994), Bianchi (1999), levaria à representação apresentada em (45). Neste caso, o núcleo nominal

²² Uma hipótese alternativa para a análise destas construções seria a de considerar que se trata de relativas apositivas de sujeito, que envolvem a anteposição de um constituinte motivada por factores discursivos (cf. Brucart 1999; Schutz-Gora 1924). Por razões de limitação de espaço, não é possível refutar detalhadamente esta hipótese. Contudo, em traços gerais, pode afirmar-se que esta análise não é adoptada por não ter o poder explicativo suficiente para dar conta das seguintes questões:

(i) Por que razão um movimento motivado por factores discursivos envolveria apenas predicados de orações copulativas e não outros constituintes?

(ii) Como explicar que um constituinte movido para uma posição dedicada à estrutura informacional possa ocorrer numa posição argumental (cf. secção 4.1.)?

(iii) Como explicar a possibilidade de coordenação das RPN com outros sintagmas nominais (cf. (22)-(23))?

é gerado no interior da oração relativa e é movido para Spec, CP, permanecendo o PP na sua posição básica. De acordo com esta representação, as RPN em análise não envolvem extraposição (i.e. movimento para a direita), mas sim movimento ‘parcial’ do constituinte relativizado e ‘stranding’ do PP.

(45) $[_{DP} [_{CP} \text{ filhosi} [_c \text{ que} [_{IP} \text{ pro} [_r \text{ fomosj} [_{VP} [_v \text{ t}_i [_{sc} \text{ t}_i [_{DP} \text{ t}_i \text{ de Rodrigo}]]]]]]]]]]]]]$

Contudo, até ao momento, não foram apresentadas neste trabalho evidências empíricas que justifiquem a adopção de uma análise em detrimento de outra. Note-se que, em todos os exemplos analisados, o PP ocorre na posição mais à direita, podendo como tal encontrar-se numa posição no interior da oração relativa ou em adjunção ao DP.

Porém, a análise por elevação permite fazer uma predição importante: se o PP pode permanecer *in situ* e não ser alvo de movimento para Spec, CP, então espera-se que possa ocorrer em outras posições para além da de predicado nominal. Ou seja, a possibilidade de ocorrência do PP numa posição interna à oração relativa deve também ser possível quando existe relativização de outras posições sintácticas.

E, de facto, esta predição é correcta: em (46) e (47), o PP ocorre claramente no interior da oração relativa:

(46) E dar autoridades aos arrêdamêtos [q(ue) __ das Egreias fforem ff(e)ctos]
(CHP156 1441)

(47) e qualquer [q(ue) __ de nos primeiro morer]
(CHP033 1266)

Este constitui, como tal, um importante argumento a favor da análise por elevação do núcleo: se o PP (adjunto ou complemento) é gerado no interior da relativa, então o núcleo nominal também deve ser gerado nessa posição. A análise ‘standard’ dificilmente poderia fornecer uma explicação satisfatória para a posição ocupada pelo PP, dado que, neste caso, estaria envolvido um movimento para a direita, para uma posição que não c-comanda o vestígio.²³

²³ Uma hipótese alternativa seria a de considerar que nestes contextos o PP não seria composto com o núcleo da oração relativa, mas gerado numa posição

5.2.2. *Combinação da análise por elevação com a teoria de movimento por cópia*

A representação em (45) coloca contudo um problema: aparentemente, o elemento movido para Spec, CP não é um constituinte. Tal facto pode ser facilmente observável em (48)-(49):

- (48) sancha ãnes [f(e)ira p(ro)fesa que foy ___ do d(i)to moest(eir)o]
(CHP171 1417)
- (49) Eu. vaasco m(art)jz [p(ri)or q(ue) ora sóo ___ (con)f(ir)mado do d(i)cto moest(eiro) de villarínho]
(CHP154 1434)

Em (48), um dos modificadores do núcleo é movido e outro (o PP) permanece *in situ*, enquanto em (49) os dois modificadores permanecem *in situ*. Como tal, estes dados demonstram claramente que: (i) na análise por elevação não é apenas o núcleo que é gerado no interior da relativa, dado que há modificadores que permanecem *in situ* (*contra* Cecchetto & Donati 2008); (ii) o elemento movido para Spec, CP não parece ser um constituinte, dado que forma um constituinte com o material que permanece *in situ*.

Uma possível solução para este ‘puzzle’ passa então por adoptar uma análise que combina a análise por elevação para as relativas com a teoria do movimento por cópia (Bošković & Nunes 2007).

De acordo com os princípios descritos por Bošković & Nunes (2007), e devido a questões relacionadas com a economia das derivações, são normalmente as cópias mais baixas na estrutura sintáctica que são apagadas. Assim, para derivar os casos de adjacência entre o núcleo nominal e o PP (cf. (19)-(20)), basta postular que este princípio geral se aplica e que se verifica o apagamento da cópia mais baixa:

- (50) [[filhos de rodrigo]ⁱ [que fomos [filhos de Rodrigo] ^j]]

Contudo, cópias estruturalmente baixas na estrutura sintáctica ou diferentes partes de diferentes elos da cadeia (‘scattered deletion’)

autónoma, por exemplo como adjunto. Contudo, neste caso, perder-se-ia a relação sintáctico-semântica que existe entre os dois elementos. Note-se, por exemplo, que em (46) o PP é argumento do nome *arrêdamêtos*.

podem ser realizados, desde que algum princípio da componente fonológica assim o determine. Desta forma, e seguindo Stjepanović (2007), proponho que a cópia mais baixa do PP pode ser pronunciada de forma a satisfazer requisitos de atribuição do acento frásico. Mais precisamente, quando o PP é interpretado como foco informacional, tem de ocorrer na posição mais à direita da frase, dado que em português (antigo e actual) a posição de final de frase é a posição *default* para a atribuição do acento frásico (e, como tal, é a posição ocupada pelos constituintes que são interpretados como foco informacional) (Costa 1998, Martins 2002).

Assim, como se pode observar em (51), os contextos de não adjacência entre o núcleo nominal e o PP podem ser derivados por ‘scattered deletion’, considerando que há uma regra da componente fonológica (atribuição do acento frásico) que exige o apagamento da cópia mais alta do PP:

(51) [[filhos ~~de rodrigo~~]ⁱ [que fomos [filhos de Rodrigo]]^j]

Para o efeito é necessário assumir, na sequência da proposta de Stjepanović (2007), que em algumas línguas, como é o caso do português, a atribuição do acento ocorre antes do apagamento de cópias. Por contraste, noutras línguas, como o Inglês, o apagamento de cópias não pode interagir com o mecanismo de atribuição do acento.²⁴

5.2.3. *Uma possível explicação para a mudança*

Tendo por base a análise apresentada, a mudança relativa às restrições observadas no decurso da história do português quanto à possibilidade de não adjacência entre o núcleo nominal e o PP pode

²⁴ Note-se que os exemplos em (46) e (47) não são derivados considerando apenas a análise apresentada até ao momento. Dado que, nestes exemplos, os PPs são preferencialmente interpretados como ‘foco identificacional’ (Kiss 1998), proponho que se verifica em primeiro lugar o movimento do PP para uma posição dedicada à estrutura informacional. Posteriormente, o NP (incluindo o vestígio) é movido para Spec, CP (cf. i). Trata-se desta forma de uma operação do tipo ‘remnant movement’ analisada em termos de teoria do movimento por cópia.

(i) [_{DP} os [_{CP} [NP arrēdamētos [~~das Egreias~~]^k] [_C que [_{FocP} [das Egreias]ⁱ] [_{IP} fforem ff(e)ctos [_{NP} arrēdamētos [~~das Egreias~~]^k]]]

explicar-se pelo facto de a operação de apagamento de diferentes partes nos elos da cadeia ('scattered deletion') ter deixado de estar activa em português, pelo menos no domínio do DP.

As RPN em análise na verdade envolvem duas propriedades contraditórias: por um lado, um constituinte deve ser movido para Spec, CP para verificar um traço *wh*; por outro lado, parte desse constituinte deve permanecer *in situ* para ser interpretado como foco informacional. Em português antigo, ambos os requisitos podem ser satisfeitos, dado que 'scattered deletion' podia aplicar-se ao nível do DP, assegurando relações mais transparentes entre estrutura sintáctica e estrutura informacional (cf. Martins 2002). Em português actual, uma vez que a 'scattered deletion' deixou de ser possível ao nível do DP, a cópia mais baixa tem de ser apagada e o PP tem de ser pronunciado na cópia mais alta.

É por esta razão que, em português actual, em construções que exibem afinidades estruturais com as RPN analisadas neste trabalho (cf. (52)), a adjacência entre o nome e o PP é obrigatória:

E foi a minha segunda surpresa. [Pobre homem do Norte que sou __], senti qualquer coisa próximo das lágrimas aflorar-me aos olhos! (Goethe, J. 2001. *Viagem a Itália*. Lisboa: Relógio D' Água, p. 272)

Referências

- Bianchi, V. 1999. *Consequences of Antisymmetry: Headed Relative Clauses*. Berlin: Mouton de Gruyter.
- Bošković, Ž.; Nunes, J. 2007. The Copy Theory of Movement. In: N. Corver; J. Nunes (Eds.) *The Copy Theory of Movement*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 13-74.
- Brito, A. M.; Duarte, I. 2003. Orações Relativas e Construções Aparentadas. In: M. H. Mateus; A. M. Brito; I. Duarte; I. H. Faria; S. Frota; G. Matos; F. Oliveira; M. Vigário; A. Villalva. *Gramática da Língua Portuguesa*. Lisboa: Caminho, 653-694.
- Bruce, J. M. 1999. La estructura del Sintagma Nominal: Las oraciones de Relativo. In: I. Bosque; V. Demonte (Orgs.). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa Calpe, 395-522.
- Carlson, G. 1977. *Reference to Kinds in English*. Doctoral Dissertation, University Massachusetts, Amherst. (Publicado em 1980 por Garland Press, Nova Iorque).

- Cecchetto, C.; Donati, C. 2008. Towards a HEAD raising analysis of relative clauses. Comunicação apresentada no XVIII Colóquio de Gramática Generativa. Lisboa, Abril 2008. Retirado em Março de 2008 da página da internet: <http://www.fl.ul.pt/18CGG/ing/programme.htm>.
- Costa, A. 2004. Aspectos das Construções de Relativização no Português do séc. XV. In: T. Freitas; A. Mendes (Orgs.). *Actas do XIX Encontro da Associação Portuguesa de Linguística*. Lisboa: APL, 409-420.
- Costa, J. 1998. *Word Order Variation. A Constraint-based Approach*. The Hague: Holland Academic Graphics.
- Demirdache, H. 1991. *Resumptive Chains in Restrictive Relatives, Appositives, and Dislocation Structures*. PhD dissertation, MIT.
- Gratacós, A. S. 1999. La aposición y otras relaciones de predicación en el sintagma nominal. In: I. Bosque & V. Demonte (Orgs.). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa Calpe, 523-564.
- Kayne, R. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.
- Kenington, H. 1937. *The Syntax of Castilian Prose: the sixteenth century*. Chicago & Illinois: The University of Chicago Press.
- Kiss, K. 1998. Identificational focus versus information focus. *Language*. **2**: 245-273.
- Leborans, M. J. 1999. La predicación: las oraciones copulativas. In: I. Bosque; V. Demonte (Orgs.). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa Calpe, 2357-2460.
- Magro, C. 2007. (Org.) *CORDIAL-SIN. Manual de Anotação Morfosintáctica*. Realizado em colaboração de E. Carrilho; A. Eliseu; M. Lobo, A. M. Martins & S. Pereira. Centro de Linguística da Universidade de Lisboa. Retirado em Março de 2008 da página da internet: www.clul.u.pt/sectores/variacao/cordialsin/manual_annotacao_morfologica.pdf
- Martins, A. M. 2002. The Loss of IP-scrambling in Portuguese: Clause Structure, Word Order Variation and Change. In: D. Lightfoot (Ed.). *Syntactic Effects of Morphological Change*. Oxford & New York: Oxford University Press, 232-248.
- Matos, G. 2003. Construções elípticas. In: M. H. Mateus; A. M. Brito; I. Duarte; I. H. Faria; S. Frota; G. Matos; F. Oliveira; M. Vigário; A. Villalva. *Gramática da Língua Portuguesa*. Lisboa: Caminho, 869-913.
- Oliveira, F. 2003. Tempo e Aspecto. In: M. H. Mateus; A. M. Brito; I. Duarte; I. H. Faria; S. Frota; G. Matos; F. Oliveira; M. Vigário; A. Villalva. *Gramática da Língua Portuguesa*. Lisboa: Caminho, 127-178.
- Oliveira, F. 2004. O Imperfeito e o Tempo dos Indivíduos. In F. Oliveira; I. M. Duarte (Eds.). *Da Língua e do Discurso*. Porto: Campo das Letras, 505-528.
- Ordóñez, S. G. 1986. *Variaciones sobre la atribucion*. León: Universidad de León [reedição, 1989].
- Sabatini, F. 1968. Dalla «scripta latina rustica» alle «scriptae romanze». *Studi Medievali*. **9**: 320-358.

- Schultz-Gora, O. 1924. *Altprovenzalisches Elementarbuch*. Heidelberg: Carl Winter.
- Stjepanović, S. 2007. Free Word Order and Copy Theory of Movement. In: N. Corver; J. Nunes (Eds.). *The Copy Theory of Movement*. Amsterdam: John Benjamins, 219-248.

Fontes

- Coelho, A. B. 1987. *Inquisição de Évora: dos primórdios a 1668*. Lisboa: Caminho. Vol. 2.
- Martins, A. M. 2001. *Documentos Portugueses do Noroeste e da Região de Lisboa*. Lisboa: Imprensa Nacional Casa da Moeda.
- Martins, A. M. 1994. *Clíticos na História do Português – Apêndice Documental*. Dissertação de Doutoramento, Faculdade de Letras da Universidade de Lisboa.
- Muhana, A. 1995. *Os autos do processo de Vieira na Inquisição* [edição, transcrição, glossário e notas de Adma Muhana]. São Paulo: UNESP.
- Pereira, I. R. 1987. *Documentos para a história da Inquisição em Portugal: séc. XVI*. Lisboa: s.n.. Vol. 1.

(Inter)subjectification des connecteurs: le cas de *car* et *parce que*

Liesbeth Degand
liesbeth.degand@uclouvain.be
UCLouvain – FRS-FNRS (Belgique)

Benjamin Fagard
benjamin.fagard@ens.fr
Lattice, CNRS – ENS (France)

RÉSUMÉ. Dans le paradigme des conjonctions de cause du français, *car* et *parce que* présentent des différences étonnantes. Si à l'écrit (presse écrite et littérature moderne), les fréquences des deux connecteurs sont très similaires, cette proportion change drastiquement à l'oral, laissant la part belle à *parce que* alors que *car* disparaît quasiment. De plus, des études antérieures ont montré que *car* est très stable sémantiquement, tandis que *parce que* varie nettement de l'oral à l'écrit. Est-il possible d'expliquer ces différences?

Afin de chercher des réponses à ce déséquilibre frappant entre écrit et oral, nous étudions les emplois de *car* et *parce que* en ancien français. Cela nous permet d'esquisser l'évolution des deux conjonctions, et de formuler l'hypothèse suivante: *parce que* aurait subi, depuis la langue médiévale, un phénomène de subjectification marqué, l'amenant à remplacer progressivement *car*. Ce processus de remplacement, qui est presque complet à l'oral, ne semble cependant pas encore avoir eu lieu pour la langue écrite.

MOTS-CLÉ. Grammaticalisation, subjectification, causalité, diachronie, étude sur corpus.

ABSTRACT. In the paradigm of French causal conjunctions, *car* and *parce que* make quite an odd pair. While both of them can be translated by «because», their use in Spoken vs Written French is very different: they roughly have the same frequency in written corpora, but *parce que* is by far more frequent in the spoken language than in written texts, whereas *car* is almost absent from Spoken French. Besides, previous studies have shown that, while *car* is quite stable semantically, *parce que* is very different in Written and Spoken French. Is it possible to explain these differences?

The goal of this paper is to find answers to this question in the diachrony of French. A detailed corpus study of *car* and *parce que* in Medieval French enables us to make assumptions about their subsequent evolution. Our

hypothesis is that the grammaticalization of *parce que* was accompanied by a process of subjectification and that it ended up replacing *car* in the spoken language, but has not yet done so in Written French.

KEYWORDS. Grammaticalization, subjectification, causality, diachrony, corpus study.

1. Introduction

Notre étude part du constat familier à tout locuteur (natif) du français que l'emploi des connecteurs de cause *car* et *parce que* est très divergent à l'oral et à l'écrit, du moins en termes de fréquence. Les analyses sur corpus confirment que les fréquences des deux connecteurs sont très similaires à l'écrit, alors qu'à l'oral, *parce que* devient onze fois plus fréquent et que *car* disparaît quasiment. Pour expliquer ce déséquilibre frappant entre écrit et oral, nous formulons l'hypothèse que les deux connecteurs poursuivent des voies de grammaticalisation différentes. *Car* aurait momentanément acquis un équilibre sémantique, alors que *parce que* poursuivrait son évolution sémantique. Cette évolution sémantique se réaliserait en premier lieu sous forme d'une subjectification, et même d'une intersubjectification qui selon Traugott & Dasher (2002: 88 sqq.) constituent des phénomènes sémantiques accompagnant fréquemment la grammaticalisation¹ des marqueurs du discours.

L'article se structure de la manière suivante: La section 2 reprend les études de Degand & Pander Maat (2003) et Simon & Degand (2007) dressant un portrait sémantique de *car* et *parce que* en français contemporain, à l'écrit et à l'oral. Les sections 3 et 4 retracent respectivement l'évolution diachronique de *car* et *parce que* sur base d'un corpus de français médiéval. Une comparaison des deux connecteurs en ancien français est donnée dans la section 5, après quoi nous présentons nos conclusions principales.

¹ Il s'agit là d'une question qui est loin d'être tranchée. Voir, entre autres, Davidse *et al.* (sous presse), Athanasiadou *et al.* (2006), Traugott (1995, 2003) sur la relation entre subjectification et grammaticalisation.

2. *Car* et *parce que* en français contemporain

Les connecteurs *car* et *parce que* ont fait l'objet de nombreuses études linguistiques (entre autres Simon & Degand, 2007; Debaisieux, 2002; Iordanskaja, 1993; Ferrari, 1992; Moeschler, 1987; Bentolila, 1986; Groupe λ -I, 1975). Si tous les auteurs s'accordent pour dire que les deux connecteurs sont différents, ils s'accordent également sur l'observation que les deux connecteurs peuvent apparaître dans des contextes très similaires, sinon identiques. Une distinction récurrente est celle faite par le Groupe λ -I (1975) caractérisant *parce que* comme une 'conjonction de contenu' et *car* comme une 'conjonction marquant un acte de parole', différence sémantique qui est reflétée dans le comportement syntaxique différent des deux connecteurs. Simon et Degand (2007) montrent sur la base de données écrites et orales, que la distinction entre les deux connecteurs en français contemporain peut être ré-exprimée en termes de subjectivité scalaire ou de degré d'implication du locuteur (Pander Maat et Degand, 2001), et que le sémantisme des deux connecteurs varie avec le mode écrit vs. oral.

Nous reprenons ici les résultats principaux de cette étude en vue de formuler des hypothèses quant à l'évolution diachronique des deux connecteurs sur le plan de la distribution sémantique, hypothèse que nous testons sur corpus dans les sections 3 et 4.

En ce qui concerne la fréquence de *car* et *parce que* à l'oral et à l'écrit, les chiffres confirment l'intuition de tout locuteur (natif) du français que *car* est quasi inexistant à l'oral. Ainsi dans la base de données Valibel², nous avons relevé 13.614 occurrences de *parce que*, soit 3,70‰, pour seulement 80 occurrences de *car*, soit 0.02‰. Par contre, dans un corpus de presse écrite (*Le Soir* 1997, +/- 27 millions

² La banque de données textuelles orales Valibel contient actuellement plus de 400 heures d'enregistrement transcrites en orthographe standard, correspondant à +/- 3,7 millions de mots. Environ 42% sont des entrevues portant sur des questions sociolinguistiques; 30% sont des entrevues informelles (entre deux ou plusieurs locuteurs qui se connaissent) et 6% sont des conversations non sollicitées par un enquêteur. 6% relèvent des médias (interviews radiophoniques, débats, émissions d'information), 2,5% relèvent du secteur professionnel (réunions de travail) et un autre 2,5% du secteur scolaire (cours). Enfin, 11% sont constitués de tâches de lecture. Cette répartition évolue chaque année au gré des projets de recherche. Voir aussi Francard, Geron et Wilmet, 2002.

de mots), *car* a une fréquence 0.40‰ contre 0.32‰ pour *parce que*. Il est par ailleurs intéressant d'observer que cette proportion stable dans l'emploi de *car* et *parce que* à l'écrit se maintient dans le genre narratif (romans dans Frantext 1980-2000, 8,6 millions de mots) comptant 0.39‰ de *car* pour 0.44‰ de *parce que*. Si, à l'écrit, la fréquence des deux connecteurs est donc sensiblement similaire, à l'oral *parce que* se rencontre 185 fois plus souvent ! Nos données confirment donc que c'est en français parlé que le fossé est le plus grand entre les deux connecteurs, en termes de fréquence³. *Parce que* semble être un connecteur de l'oral par excellence, avec une fréquence 11,5 fois plus élevée à l'oral qu'à l'écrit (3,70‰ contre 0.32‰, dans les corpus Valibel et *Le soir*, cf. ci-dessus).

2.1. Profil sémantique de *car* et *parce que* à l'écrit

Nous défendons l'idée que les relations de cohérence et leurs marqueurs linguistiques peuvent être décrits en termes d'*Implication du locuteur* ('Speaker Involvement', désormais IdL) (Degand & Pander Maat, 2003; Pander Maat & Degand, 2001). Celle-ci fait référence au degré avec lequel le locuteur joue implicitement un rôle actif dans la construction de la relation, en l'occurrence de la relation causale. Le degré d'implication augmente à mesure que le locuteur investit la relation causale d'un certain nombre de suppositions lui conférant ainsi une fonction légitimante (Rossari & Jayez, 1996). Selon notre hypothèse, les différents connecteurs sont ordonnés sur une échelle allant d'une implication minimale du locuteur (relation objective) à une implication maximale (relation subjective)⁴. Ainsi, dans le domaine de la causalité, l'échelle comporte en ordre croissant d'IdL les relations suivantes:

- causale non volitive, qui établit une relation de cause à effet entièrement objective. Le locuteur met en relation deux états des choses de manière purement factuelle sans s'impliquer dans la relation:

(1) *Le match est annulé parce que le terrain est détremé.*

³ Ces données confirment sensiblement celles de Labbé (2003).

⁴ Pour une conception des relations causales en termes de subjectivité, voir Pit (2003), Pander Maat & Sanders (2001).

– causale volitive, qui implique un protagoniste conscient, capable d'actions et de décisions posant un acte volontaire *q* sur la base d'un jugement d'un état des choses *p*. Il y a donc une certaine évaluation de la situation qui constitue pour le protagoniste une raison valide pour agir, raison qui peut être validée ou non par l'interlocuteur:

(2) *Annie est rentrée tôt parce que son copain n'était pas là.*

– causale mentale, qui donne une raison jugée valide par le protagoniste pour un état d'esprit plutôt que pour une action (concrète):

(3) *Je suis contente parce qu'il y a congé demain.*

– causale épistémique, qui fait intervenir une causalité argumentative. Le protagoniste (par défaut le locuteur) tire une conclusion à partir d'un état des choses, ce dernier fonctionnant comme une prémisse pour une argumentation. Elle doit dès lors être connue et acceptée. En d'autres termes, la conclusion est présentée comme la conséquence d'un raisonnement argumentatif, il s'agit donc d'un état mental du protagoniste. Contrairement à la causalité mentale, la causalité épistémique est 'immédiate', elle se situe dans le *hic et nunc* de la situation causale. Pour être valide, la ou les prémisse(s) doi(ven)t être partagée(s) par l'interlocuteur:

(49) *Il doit avoir près de 90 ans, parce qu'il faisait déjà partie de l'unité en 1932.*

– causale interactionnelle, où le locuteur n'est pas impliqué en tant que personne pensante, mais uniquement dans son rôle de locuteur, la relation causale étant établie avec l'acte de langage proprement dit (justification d'une assertion, d'une question), plutôt que son contenu propositionnel:

(5) *Tu peux me donner ton âge, parce qu'il me le faut pour l'enquête.*

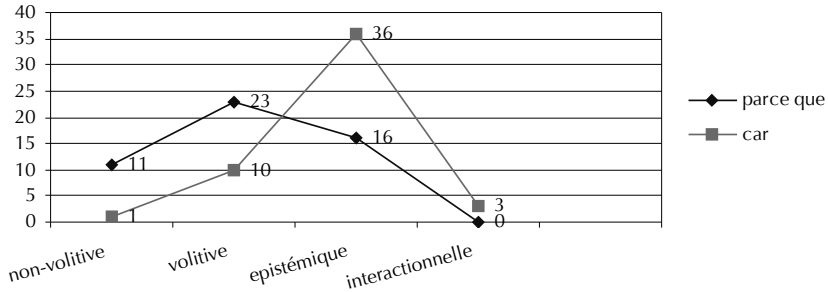
Le niveau d'IdL d'une relation (causale) va de pair avec un certain nombre de caractéristiques discursives. En résumé, une relation causale aura un degré d'IdL bas si elle exprime une relation causale conforme à la causalité dans le monde, si aucun protagoniste n'intervient dans la relation ou, si c'est le cas, si ce protagoniste est nommé explicitement,

et si l'événement causal est relaté indépendamment de la situation énonciative présente. La relation causale aura, au contraire, un niveau d'IdL maximal si elle n'est pas iconique avec la causalité réelle, si elle fait intervenir un protagoniste conscient qui reste implicite dans la relation, et si la situation causative se confond avec la situation énonciative (cf. Pander Maat & Degand, 2001 pour une justification théorique; Degand & Bestgen, 2004 pour une opérationnalisation). Ces caractéristiques concernent les relations causales. En ce qui concerne les connecteurs – en tant que marqueurs prototypiques des relations de cohérence –, nous avons établi qu'un connecteur *encode un certain niveau d'IdL*, qui constitue sa contribution à l'interprétation de son environnement discursif. Comme pour les relations de cohérence, la place qu'occupe un connecteur sur l'échelle se reflète dans son comportement discursif. La représentation scalaire rend compte du fait que les connecteurs ne sont pas strictement liés à un 'niveau' spécifique de signification, mais qu'ils imposent néanmoins des contraintes sur les contextes dans lesquels ils peuvent apparaître, certains contextes étant plus 'naturels' que d'autres.

En ce qui concerne la distribution sémantique de *car* et *parce que* à l'écrit, l'étude de Degand et Pander Maat (2003) confirme que les deux connecteurs peuvent apparaître dans des contextes causaux similaires. Tous deux peuvent exprimer des relations causales non volitives, volitives et épistémiques. Le seul contexte dans lequel on ne retrouve pas le connecteur *parce que* à l'écrit est la relation interactionnelle. Néanmoins, cette similitude n'empêche pas une certaine spécialisation sémantique des connecteurs. On observe, en effet, une tendance significative pour *parce que* à exprimer des relations causales à IdL moins élevée (surtout non volitives et volitives), alors que *car* exprime des relations plus subjectives, y compris les relations interactionnelles ($\chi^2 = 24,15$, $df = 3$ et $p < .0001$)⁵. Il en résulte un profil sémantique différent (même s'il y a une intersection) pour les deux connecteurs (Figure 1).

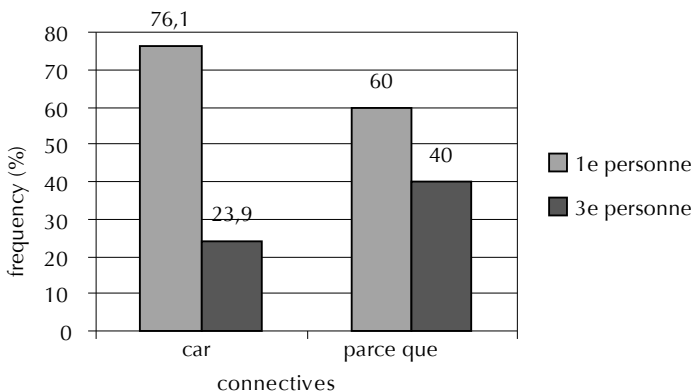
⁵ Pour des raisons de puissance statistique, les relations causales mentales et épistémiques ont été regroupées en une seule catégorie épistémique.

Fig. 1. Distribution relationnelle de *car* et *parce que* dans un corpus journalistique (corpus *Le Soir*, 1997) sur la base des résultats de Degand & Pander Maat (2003) (cf. Figure 1 dans Simon et Degand, 2007).



Un indice supplémentaire de la subjectivité plus élevée de *car* est la tendance pour ce connecteur à apparaître plus fréquemment en co-occurrence avec un verbe à la première personne. L'hypothèse sous-jacente est que les connecteurs à IdL plus élevée se retrouvent plus souvent en cooccurrence avec des protagonistes à la première personne, puisque les actions et conclusions d'un locuteur seront, en général, formulées avec une subjectivité plus grande que les actions et conclusions d'une tierce personne. Un locuteur sera en effet enclin à accepter plus facilement les assomptions liées à ses propres décisions que celles liées aux décisions de tiers.

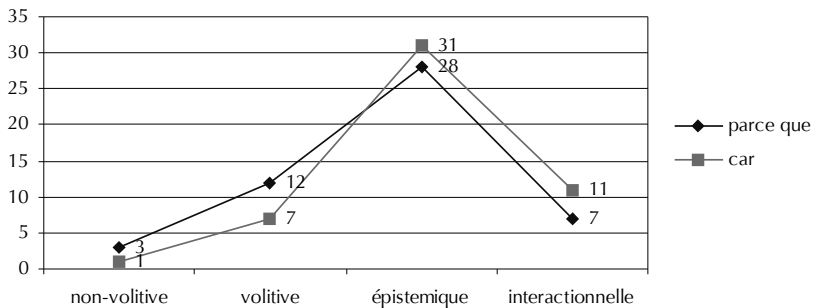
Fig. 2. Identité du protagoniste causal accompagnant *car* et *parce que* sur base des résultats de Degand et Pander Maat (2003).



2.2 Profil sémantique de *car* et *parce que* à l'oral

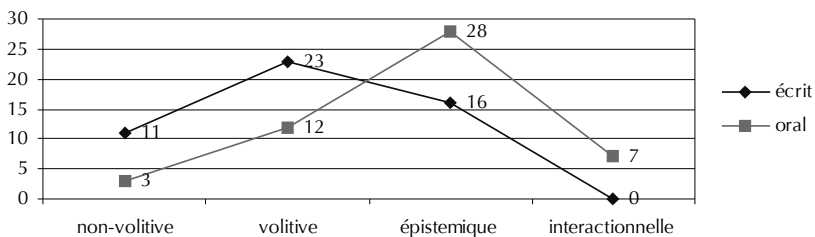
L'étude de Simon et Degand (2007) concernant la distribution sémantique de *car* et *parce que* en français parlé donne les résultats principaux suivants: A l'oral, le profil sémantique des deux connecteurs n'est plus divergent ($\chi^2 = 3.357$, $df = 3$ et $p = .34$ (N.S.)). Il semble que l'on doive attribuer en premier lieu cette similarité de profil à la distribution sémantique plus étendue de *parce que* (Figure 3). Il semble en effet que *parce que* rejoigne *car* dans ses usages les plus subjectifs.

Fig. 3. Distribution relationnelle de *car* et *parce que* dans les corpus oraux (cf. Figure 2 dans Simon et Degand (2007))



Cette observation se confirme si l'on fait une comparaison intermodale de *parce que* (Figure 4). Le connecteur a, en effet, une nette tendance à exprimer des relations avec un degré d'|*Idl*| plus élevé (des relations plus subjectives) à l'oral qu'à l'écrit, et ceci de manière statistiquement significative ($\chi^2 = 18.301$, $df = 3$ et $p < .0001$), ce qui nous porte à parler de la 'subjectivisation' de *parce que* à l'oral.

Fig. 4. Distribution relationnelle de *parce que* à l'oral et à l'écrit (cf. Figure 3 dans Simon et Degand (2007))



A l'inverse, la distribution sémantique de *car* ne diverge pas à l'oral et à l'écrit ($\chi^2 = 5.74$, $df = 3$ et $p = 0.14$ (N.S.)).

Sur base de ces observations, nous pouvons conclure que *car* est un connecteur stable en terme de mode: son profil ne varie pas selon le mode oral ou écrit, même si sa fréquence est, elle, très divergente. Une explication peut venir de l'analyse du profil sémantique de *parce que* qui à l'oral semble 'prendre la place' de *car* en couvrant toutes les situations énonciatives, des plus objectives au plus subjectives. La co-occurrence de *parce que* avec des marques de la première personne pointe dans la même direction: À l'oral, on observe une véritable montée de ce type de marques (de 60% à l'écrit à 86.1% à l'oral) alors que pour *car* cette évolution est moindre (de 76.1% à l'écrit à 82.3% à l'oral). Ceci nous amène à émettre l'hypothèse que les deux connecteurs poursuivent des voies de grammaticalisation différentes. *Car* aurait momentanément acquis un équilibre sémantique, alors que *parce que* poursuivrait son évolution sémantique. Une étude diachronique semble nécessaire pour vérifier ou infirmer cette hypothèse: c'est le but des sections suivantes.

3. Evolution sémantique de *car*

3.1. *Car en ancien français*

En ancien français, *car* conserve des emplois qu'il a hérités du latin (*qua re > quare* avait principalement des emplois comme relatif et interrogatif), mais une étude à partir du corpus de la BFM (Base du Français Médiéval, UMR 5191 ICAR / ENS-LSH, <http://bfm.ens-lsh.fr>)⁶ permet de montrer que les emplois les plus fréquents de *car* en ancien français sont – dès les plus anciens textes – ceux comme conjonction et particule de renforcement de l'impératif (ex. 6), comme on le voit dans la figure 5 ci-dessous.

(6) Renforcement de l'impératif:

*Filz, **quar** t'en vas colcer avoc ta spuse*

Fils, couche-toi près de ta femme (*Saint Alexis*, c. 1050)

⁶ Nous avons également utilisé, en complément, la base Champion Electronique.

Dans ces nouveaux emplois de *car* comme conjonction, on peut noter divers degrés de subjectivité que nous mettons en relation avec les emplois contemporains décrits plus haut: emploi objectif, non volitif (ex. 7), emploi volitif (ex. 8), épistémique (ex. 9), et même intersubjectif avec l'emploi interactionnel (ex. 10).

- (7) Non volitif:

*Ydoine vint à son ostel, Oū il n'avoit ne pain ne sel, **Quar** povreté la destregnoit...*

Ydoine arriva chez elle, où il n'y avait ni pain ni sel, car elle était accablée par la pauvreté. (Du Segretain Moine, *Recueil général et complet des fabliaux des XIII^e et XIV^e s.*, t. V – p. 220)

- (8) Volitif:

*Reposent sei **quar** lassét sunt.*

Ils se reposent, parce qu'ils sont fatigués. (*Saint Brendan*, 12^{ème} siècle)

- (9) Epistémique:

*Forment m'en poise, **quar** mout l'avoie amé; Mes par mon chief ja sera comparé.*

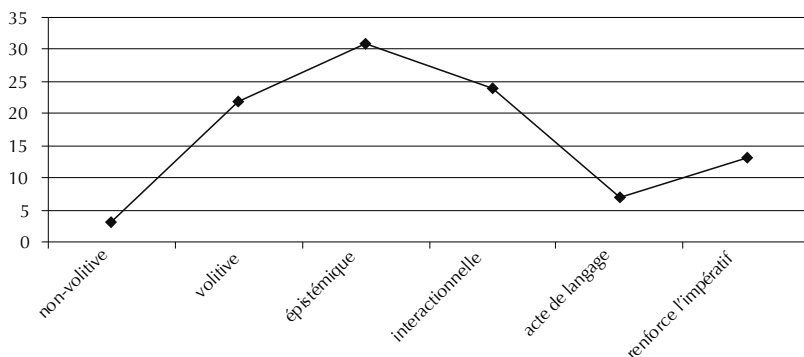
J'en suis fort triste, car je l'aimais beaucoup; mais, je vous l'assure, il sera vengé. (*Aliscans*, 12^{ème} siècle)

- (10) Interactionnel (acte de langage):

*Ferez i, Francs, **kar** très ben les veintrum !*

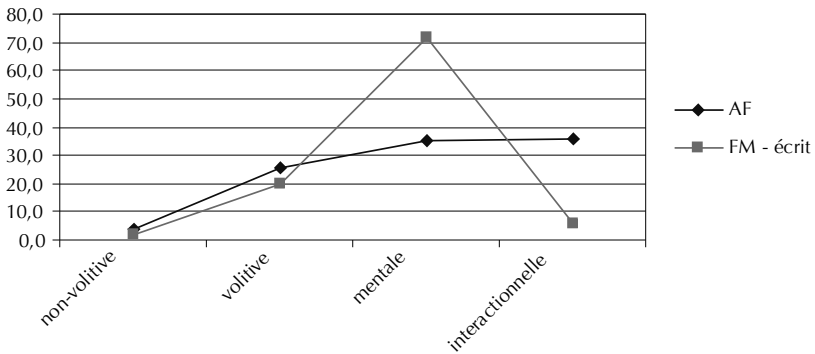
Frappez, Francs, **car** nous les vaincrons facilement ! (*Chanson de Roland*, vers 1100)

La figure 5 illustre la répartition de ces différents emplois dans notre corpus. On voit que les emplois les plus fréquents sont les emplois subjectifs et intersubjectifs, comme en français moderne. Les emplois purement objectifs sont très rares – ils constituent moins de 5% des occurrences; les emplois plus subjectifs (cause volitive et épistémique) et intersubjectifs (cause interactionnelle) font jeu égal, tandis que les cas où *car* renforce l'impératif restent peu fréquents, et disparaissent en moyen français.

Fig. 5. Sens de *car* en ancien français (corpus BFM; en %)

3.2. *Car en diachronie*

De l'ancien français au français moderne, il semble que *car* ait peu évolué, du moins d'après nos corpus. La figure 6 ci-dessous met en regard la répartition des emplois causaux de *car* dans les corpus d'ancien français et de français moderne. On peut observer deux phénomènes majeurs: d'une part, les emplois subjectifs semblent avoir augmenté; d'autre part, les emplois intersubjectifs, relativement fréquents en ancien français ont à peu près disparu en français moderne. Tandis que la première tendance paraît naturelle si l'on pose pour la subjectification le même principe d'unidirectionnalité que pour la grammaticalisation (voir Traugott & Dasher 2002), la seconde pose problème, pour la même raison. Il semble cependant possible de l'expliquer, si l'on prend en compte la quasi absence de *car* à l'oral en français moderne: étant donné que l'intersubjectivité est le propre de l'interaction verbale, elle est nécessairement liée à l'oral; puisque *car* n'est presque plus employé à l'oral, il est normal qu'il perde ses emplois intersubjectifs à l'écrit aussi.

Fig. 6. Evolution sémantique de *car* (emplois causaux; corpus BFM & Le soir)

4. Evolution sémantique de *parce que*

4.1. *Parce que* en ancien français

En ancien français, il est difficile de distinguer entre *par ce que*, *por ce que*, *parce que* et *porce que*: les scribes ne font pas toujours la différence (*par* et *por* pouvant tous deux être abrégés de la même manière, par un *p* barré), laissant toute latitude à l'éditeur du texte pour l'interprétation. Nous rendons compte ici de l'ensemble des graphies trouvées dans notre corpus médiéval, et nous indiquerons autant que possible le lien entre sens et graphie; il faut noter cependant que lorsque nous évoquons *parce que*, nous faisons référence à l'ensemble des graphies trouvées dans le corpus.

Par ailleurs, la locution *parce que* n'est causale que dans une partie de ses emplois; on trouve également des emplois comme conjonction introduisant des subordonnées de but (ex. 10) ou de moyen (ex. 11):

(10) But:

Sunent mil grailles por ço que plus bel seit.

Ils font sonner mille cloches, pour que la fête soit à son comble (*Chanson de Roland*, vers 1100)

(11) Moyen:

En grant peine nus mist Par ço quë il mangat Ço que Eve li dunat Sur le defens de Dé, Ulte sa volenté.

Il nous a mis dans une situation terrible en mangeant ce que lui avait donné Eve, malgré l'interdiction divine (*Comput*, 12^{ème} siècle)

Il y a cependant un certain nombre d'occurrences purement causales; *parce que* introduit alors la cause objective (ex. 12) ou volitive (ex. 13):

(12) Non volitif:

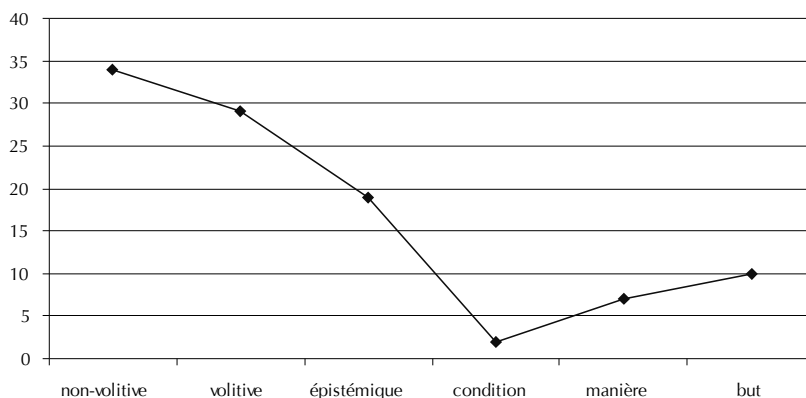
Li anfes ploroit de grant fin por ce que n'avoit que mengier
L'enfant pleurait de faim, parce qu'il n'avait rien à manger (*Roman de Renart*, 13^{ème} siècle)

(13) Volitif:

mort me fis en mi la voie por ce que trop grant faim avoie
Je fis le mort au milieu de la route, parce que j'avais une faim terrible (ibid.)

La figure 7 illustre la répartition de ces différents emplois dans notre corpus. Comme on le voit, le contraste avec *car* est assez net, puisque les emplois purement objectifs sont les plus fréquents, suivis des emplois volitifs et épistémiques; les emplois non causaux sont relativement peu fréquents: moins de 20 % des occurrences du corpus.

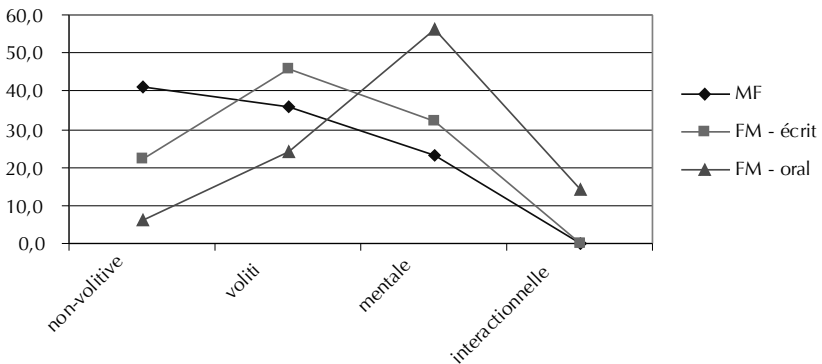
Fig. 7. Emplois de *parce que* en ancien français (corpus BFM)



4.2. Parce que en diachronie

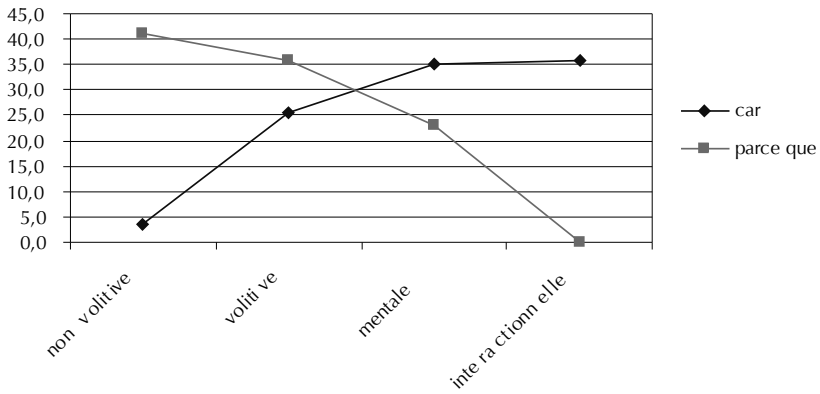
L'évolution des emplois causaux de *parce que* de l'ancien français au français moderne est plus importante que celle de *car*, comme le montre la figure 8 ci-dessous. La comparaison avec les emplois de *parce que* à l'oral, en français moderne, semble de plus confirmer la tendance très nette vers une subjectification de la conjonction. Ainsi, *parce que* perd progressivement ses emplois objectifs, et gagne au contraire des emplois subjectifs; à l'oral, il acquiert même des emplois intersubjectifs, apparemment absents à l'écrit.

Fig. 8. Evolution sémantique de *parce que* (emplois causaux; corpus BFM, *Le soir* & Valibel)

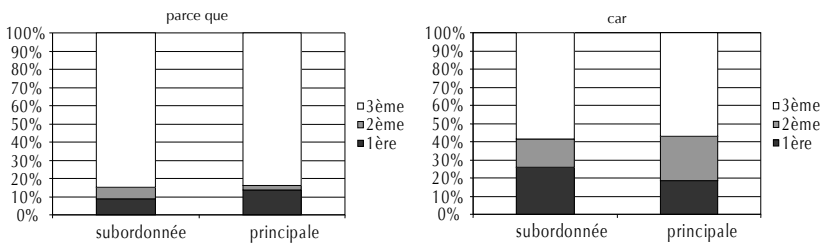


5. Car et *parce que* en ancien français: comparaison sur critères morpho-syntaxiques

La répartition des emplois observée pour les conjonctions *car* et *parce que* en ancien français laisse penser que la première présente plus fréquemment que la seconde des emplois subjectifs, comme l'indique la figure 9 ci-dessous.

Fig. 9. Emplois causaux de *car* et *parce que* en ancien français (corpus BFM)

Afin d'appuyer ce constat, nous avons analysé systématiquement le contexte d'apparition des deux conjonctions, en particulier le temps, l'aspect et le mode (TAM) du verbe principal de chaque proposition (principale et subordonnée). La place manque pour présenter en détail l'ensemble de ces résultats, mais la figure 10 ci-dessous reprend celui qui nous semble le plus révélateur, à savoir le lien entre la personne (sujet du verbe) et la conjonction, pour chaque proposition.

Fig. 10. Personne du verbe de la principale et de la subordonnée introduite par *car* et *parce que* respectivement (corpus BFM)

Les résultats renforcent la conclusion que *car* est une conjonction plus subjective que *parce que*: pour les subordonnées introduites par *car*, comme pour le verbe de la proposition principale, le verbe est à la 1^{ère} ou 2^{ème} personne dans plus de 40 % des cas. Pour les subordonnées introduites par *parce que*, comme pour le verbe de

la proposition principale, ce n'est plus le cas que dans moins de 20 % des occurrences. Nous avons fait une observation similaire pour le français moderne (cf. Figure 2 supra): la comparaison des deux figures permet d'observer une subjectification accrue des deux conjonctions, dans des proportions différentes cependant. Pour *parce que*, la subjectification est très nette, puisque la part des emplois avec un protagoniste à la 1^{ère} personne passe de moins de 20 % en ancien français à 60 % en français moderne. Pour *car*, la proportion est déjà de plus de 40 % de première ou deuxième personne en ancien français, et passe à un peu moins de 75 % (pour la première personne seulement) en français moderne. Ces éléments confirment donc la subjectification des deux conjonctions dans la diachronie du français⁷.

6. Conclusion

Pour conclure, on notera qu'il y a un double contraste entre les deux conjonctions étudiées dans cet article. D'une part, *car* a des emplois surtout causaux et subjectifs dès l'ancien français, tandis que *parce que* est plus tardif (à la fin du 11^{ème} siècle *por ço que*, au 12^{ème} siècle *par ço que*) et connaît à l'origine des emplois causaux plutôt objectifs, ainsi que des emplois non causaux. D'autre part, *car* ne présente quasiment aucune évolution de l'ancien français au français moderne, si ce n'est une baisse très nette de fréquence et la perte de l'emploi devant l'impératif (aux dépens de *donc*), tandis que *parce que* évolue très nettement d'emplois objectifs vers des emplois subjectifs et, à l'oral, intersubjectifs.

Ces deux conjonctions semblent donc présenter un cas d'école du renouvellement grammatical: tandis qu'un morphème déjà grammaticalisé tend à se figer et à devenir moins fréquent, un autre, nouvellement grammaticalisé, acquiert progressivement les emplois

⁷ Il serait intéressant de compléter ces données par une analyse contrastive de l'emploi de *car* dans le récit et le discours direct, en ancien français. Si l'on admet que le discours direct est représentatif de l'oral, comme nous l'avons observé dans le cas de l'adverbe *alors* (Degand & Fagard 2008), cela nous permettrait de faire des hypothèses plus précises sur le degré de subjectification de *car* en ancien français, dans la langue écrite et parlée.

du premier, qu'il finit par remplacer complètement. Le plus étonnant ici est qu'il a fallu un millénaire pour que cette évolution parvienne à son terme, et que la langue «normée», écrite, ne reflète toujours pas cette évolution.

Remerciements

Cette recherche est soutenue financièrement par un Pôle d'attraction interuniversitaire financé par le gouvernement fédéral belge sous le contrat PAI P6/44 «Grammaticalization and (inter)Subjectification». La première auteure est chercheuse qualifiée du Fonds de Recherche Scientifique (FRS-FNRS) financé par la Communauté française de Belgique.

Nous tenons également à remercier les relecteurs anonymes pour leurs commentaires.

References

- Athanasiadou, A.; Canakis, C.; Cornillie, B. (Eds.). 2006. *Subjectification: Various paths to Subjectivity*. Berlin/New York: Mouton de Gruyter.
- Davidse, K.; H. Cuyckens; Vandelanotte, L. (Eds., sous presse). *Subjectification, intersubjectification and grammaticalization*. Berlin: Mouton.
- Debaisieux, J.-M. 2002. Le fonctionnement de *parce que* en français contemporain: étude quantitative. In: C. Pusch; W. Raible (Dir.). *Romanistische Korpuslinguistik – Romance Corpus linguistics*. Tübingen: Gunter Narr Verlag, 349–362.
- Degand, L.; Bestgen, Y. 2004. Connecteurs et analyses de corpus: de l'analyse manuelle à l'analyse automatisée. In: S. Porhiel; D. Klingler (Eds.). *L'Unité texte*. Pleyben: Perspectives, 49-73.
- Degand, L.; Fagard, B. 2008. Alors between discourse and grammar: the role of syntactic position. Présentation orale: *DG2008: International Conference on Discourse and Grammar*, 23-24 mai 2008, Gent, Belgique.
- Degand, L.; Pander Maat, H. 2003. A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In: A. Verhagen; J. van de Weijer (Eds.). *Usage based approaches to Dutch*. Utrecht: LOT, 175-199.
- Ferrari, A. 1992. Encore à propos de *parce que*, à la lumière des structures linguistiques de la séquence causale. *Cahiers de Linguistique française*. **13**: 183-214.
- Francard, M.; Geron, G.; Wilmet, R. 2002. La banque de données VALIBEL: des ressources textuelles orales pour l'étude du français en Wallonie et à Bruxelles.

- In: C. D. Pusch; W. Raible (Eds.). *Romanistische Korpuslinguistik – Korpora und gesprochene Sprache/Romance Corpus Linguistics – Corpora and Spoken Language* (= *ScriptOralia*, 126). Tübingen: Gunter Narr, 71-80.
- Groupe λ-l. 1975. Car, parce que, puisque. *Revue Romane*. **10**: 248-280.
- Iordanskaia, L. 1993. Pour une description lexicographique des conjonctions du français contemporain. *Le Français moderne*. **2**: 159-190.
- Labbé, D. 2003. Coordination et subordination en français oral. *Présentation aux IV^e Journées de l'ERLA, Coordination/Subordination en Français de Spécialité*, Brest, 14-15 nov. 2003. <http://www.upmf-grenoble.fr/cerat/Recherche/PagesPerso/LabbeBrest03.pdf>.
- Moeschler, J. 1987. Trois emplois de *parce que* en conversation. *Cahiers de Linguistique française*. **8**: 97-110.
- Pander Maat, H.; Degand, L. 2001. Scaling causal relations and connectives in terms of Speaker Involvement. *Cognitive Linguistics*. **12(3)**: 211-245.
- Pander Maat, H.; Sanders, T. 2001. Subjectivity in causal connectives: An empirical study of language in use. *Cognitive Linguistics*. **12(3)**: 247-273.
- Pit, M. 2003. *How to Express Yourself with a Causal Connective. Subjectivity and Causal Connectives in Dutch, German and French*. Amsterdam: Rodopi.
- Rossari, C.; Jayez, J. 1996. *Donc* et les consécutifs. Des systèmes de contraintes différentiels. *Linguisticae Investigationes*. **XX(1)**: 117-143.
- Simon, A. C.; Degand, L. 2007. Connecteurs de causalité, implication du locuteur et profils prosodiques. Le cas de *car* et de *parce que*. *Journal of French Language Studies*. **17**: 323-341.
- Traugott, E. 1995. Subjectification in grammaticalization. In: D. Stein; S. Wright (Eds.). *Subjectivity and Subjectivisation*. Cambridge: Cambridge University Press, 37-54.
- Traugott, E. 2003. From subjectification to intersubjectification. In: R. Hickey (Ed.). *Motives for language change*. Cambridge: Cambridge University Press, 124-139.
- Traugott, E.; Dasher, R. 2002. *Regularity in Semantic Change*. Cambridge: C.U.P.

Bases de données

- Base de Français Médiéval (Laboratoire ICAR, ENS-LSH / CNRS).
- Base Champion Electronique (Corpus de la littérature médiévale, éditions Champion).
- Base textuelle du moyen français (Laboratoire ATILF, CNRS).
- Base Frantext (Laboratoire ATILF, CNRS).
- Base Valibel (Groupe de Recherches Valibel, UCLouvain)

The effect of language contact on Romance verbal paradigms: an empirical survey¹

Ana R. Luís

aluis@fl.uc.pt

University of Coimbra (Portugal)

Abstract. Rather than conveying morphosyntactic meaning, conjugation classes determine how such properties are expressed. Conjugation classes are therefore ‘ornamental’ properties of language and clearly not essential to communication. Such purely formal distinctions tend to be absent from contact varieties largely because adults have a natural tendency to bypass linguistic features that are perceived as semantically unnecessary (e.g., Seuren and Wekker 1986, McWhorter 2001). The goal of this paper, however, will be to show that language contact does not necessarily lead to a loss of meaningless forms. Building on recent work by Luís (2008, forthcoming), we provide empirical evidence which shows that Romance conjugation classes respond in different ways to language contact. In particular, we show that conjugation classes may undergo different types of morphological change, such as lexicalization, levelling, retention and extension. Our evidence will be drawn from contact varieties which have derived from the contact between one Romance language (either Portuguese or Spanish) and one (or more than one) non-Romance language.

Key-words. Morphological change, language contact, verbal inflection, theme vowels, conjugation classes, Indo-Portuguese, Spanish Romani.

1. Introduction

In this paper we examine the loss and survival of conjugation classes in language contact varieties. Scarcely any attention has been paid to this question and the reasons for such lack of interest may be two-fold. On the one hand, there appears to be the widespread belief in contact-linguistics that inflectional morphology is an ornamental property of linguistic structure which can, by and large, be discarded without affecting communication (McWhorter 2001). On the other

¹ For valuable suggestions and detailed comments, I would like to thank Hugo Cardoso, Clancy Clements, Andrew Koontz-Garboden, Ingo Plag, and Jeff Siegel. All errors are of course my own.

hand, there is a tendency to focus predominantly on meaningful inflections, i.e. inflections which express some kind of morphosyntactic meaning (e.g., person, number, tense, among other). Conjugation classes have therefore gone entirely unnoticed in contact linguistics largely because they make no contribution to the overall meaning of an utterance. In this paper, we investigate the effect of language contact on conjugation classes by examining the verbal paradigms of language varieties which have emerged from the contact between one Romance language (either Portuguese or Spanish) and one (or more than one) non-Romance language. We will provide evidence which shows that Romance conjugation classes do effectively survive in some contact varieties.

Our findings weaken the idea that the grammar of contact varieties is determined entirely by the pragmatic relevance or the semantic transparency of linguistic forms (Seuren and Wekker 1986; McWhorter 1998). Our data instead shows that other factors must play an important role in shaping the various interlanguages spoken by a given language community. Winford (2003), for example, takes the view that the factors that are responsible for the linguistic outcome can be broadly divided into two classes: internal (linguistic) and external (social and psychological). The linguistic factors include the degree of typological similarity between the contact languages and linguistic constraints that are specific to particular sub-areas of grammar, including factors a more general (perhaps universal) nature. Among the external factors, there is the length and intensity of the contact and also the attitude of speakers towards the languages and the motivations for using them.

More recently, research has also highlighted the importance of Second Language Acquisition in the development of contact varieties. The difference between contact induced language change and other kinds of language change resides in the role played by adult speakers: "adults rather than children or adolescents are ultimately responsible for the changes as a result of second language learning (Siegel 2004:154). Thus, the question as to why certain grammatical features survive while other are discarded has been explored from the perspective of untutored second language learning in combination with both socio-cultural and attitudinal factors (Siegel 2004; Veenstra

2004; Clements 2006; Plag 2008). Overall, then, research indicates that language contact should be treated as a multidimensional process which results from the complex interaction between internal and external forces.

For reasons of space, we leave the factors underlying the partial or total retention of Romance conjugation classes unaddressed. Instead we provide an empirical overview of the kinds of morphological change suffered by conjugation classes². Our paper is structured as follows. Section 2 highlights some of the inflectional properties of Romance verbal paradigms and illustrates the effect of conjugation classes on the shape of verb forms. Section 3 examines verbal paradigms in various ‘Romance-based’ contact languages and argues that conjugation classes may suffer the following diachronic changes: lexicalization (section 3.1), levelling (section 3.2), retention (section 3.3.) and extension (section 3.4). Section 4 offers a brief summary of our main findings.

2. Inflectional paradigms: conjugation classes and theme vowels

In what follows, we will briefly illustrate how conjugation classes determine the shape of verb forms. As noted by Aronoff (1994: 46), conjugations tell us “which inflectional affixes will realize the various morphosyntactic properties that the verb bears in a particular instance”. Knowing the conjugation class of a given verb therefore enables us to predict its inflected forms. In (1), for example, there are two suffixes realising the imperfect tense in Portuguese, namely the suffix *-va* and the suffix *-ia* (cf.(1)). Each one of them is selected by a different conjugation class: while *-va* attaches to verbs belonging to the first conjugation, *-ia* must attach to verbs belonging to the second and third conjugations.

² However, see Clements (1996) and Luís (2008) on the causes underlying the shape of verbal paradigms in Indo-Portuguese. See also Siegel (2003) and Plag (2008), for a more general discussion about the fate of morphology in creoles, and see Matras & Bakker (2003) on the genesis of mixed languages.

- (1) a. *cant* -a -va -s
 sing -cl1 -impf -2sg
 '(you.sg) sang'
 b. *beb* -ia -s
 drink.cl2 -impf -2sg
 '(you.sg) drank'
 c. *sent* -ia -s
 feel.cl3 -impf -2sg
 '(you.sg) felt'

Conjugation classes can also influence the shape of other verbal endings. As shown in (2), there are two past allomorphs for third person singular verb forms. The third person singular suffix is expressed through the suffix *-o* in the first class (cf. (2a)), but through the suffix *-u* [-w] in the second and third classes (cf. (2b) and (2c)).

- (2) a. *cant* -ou [-o]/[-ow]
 sing.cl1 -perf.3sg
 '(s/he) sang'
 b. *beb* -e -u [-w]
 drink -cl2 -perf.3sg
 '(s/he) drank'
 c. *sent* -i -u [-w]
 feel -cl3 -perf.3sg
 '(s/he) felt'

Suffix allomorphy occurs when a given language provides more than one affix for a given set of morphosyntactic features. In the Portuguese examples given in (2), the past tense allomorphy is entirely triggered by the conjugation class of the verbal lexemes.

Theme vowels, as alluded to before, are conjugation class markers. In Portuguese, there are three theme vowels, namely *-a* (cf. (1a)), *-e* (cf. (2b)) and *-i* (cf. (2c)), which follow the root. These vowels are inflectional units which make neither a semantic nor a morphosyntactic contribution to the overall meaning of the inflected verb. Unlike theme vowels, verb roots convey lexical meaning (e.g., *cant-* means 'sing') while verbal endings express morphosyntactic information (e.g., TMA or person/number agreement, as in *-va* 'imperfect.CL1' and *-s* '2nd person singular', respectively).

Theme vowels, as stem-forming units, combine with the root to form the base to which inflectional endings attach. Each conjugation class has its default stem (or primary stem) which contains the default theme vowel. However, theme vowels may also undergo allomorphy and, therefore, conjugation classes may in effect take more than one stem. Secondary stems may differ from the primary stem in various ways. For example, in (2a), the first conjugation theme vowel is absent and therefore both the stem and the root are homophonous; and in (3), the primary stem (containing the default -e theme vowel (cf. (2b))) is systematically replaced by a secondary stem (containing the theme vowel -i). In (4), there are two present subjunctive stems, namely the -a stem in the second and third conjugation classes (cf. (4b-c)) and the -e stem in the first conjugation class (cf. (4a))³.

- (3) *beb* *-i* *-do*
 'drink -cl3 -compl'.
 'drunk'
- (4) a. *cant* *-e* *-s*
 sing - cl2.sbj -2sg
 '(you.sg) sing'
- b. *beb* *-a* *-s*
 drink - cl1.sbj -2sg
 '(you.sg) drink'
- c. *sint* *-a* *-s*
 feel - cl1.sbj -2sg
 '(you.sg) feel'

This section has offered a brief outline of some of the formal properties of Romance verbal paradigms, focussing on the inflectional status of both conjugation classes and theme vowels. Crucial to our discussion is the idea that conjugation classes, despite their lack of meaning, play a crucial role in determining the shape of inflected verb forms in Portuguese.

³ There may be alternative ways of segmenting the subjunctive forms, however the crucial point here is that any inflectional account of these forms must make reference to conjugation class distinctions.

3. The fate of Romance conjugation classes: an empirical survey

When Romance verbal paradigms come into contact with genetically distinct languages, four different types of morphological change can take place: lexicalization, levelling, retention and extension. These four different types of diachronic change differ from each other as follows: a) *lexicalization* triggers the loss of conjugation classes (section 3.1); b) *levelling* leads to a reduction of classes (section 3.2); c) *retention* involves the complete survival of the Romance conjugations (section 3.3), and d) *extension* leads to an expansion of the conjugation classes (section 3.4).

As alluded to above, our evidence is drawn from contact varieties which emerged from the contact between one Romance language (either Portuguese or Spanish) and one (or more than one) non-Romance language. In particular, we examine the verbal paradigm of several Portuguese-based creoles – i.e., Sotavento Kabuverdianu (Baptista 2003; Baptista *et al.* 2007), Kriyol (Kihm 1994), the Indo-Portuguese varieties of Korlai, Daman and Diu (Clements 1996; Clements & Koontz-Garboden 2002; Cardoso 2006) – and the verbal paradigm of one mixed contact language, namely Spanish-Romani, also known as Caló (Boretzky & Iglá 1994; Bakker 2003).⁴

3.1. Theme vowel lexicalization

The borrowing of morphologically complex word forms often involves the loss of morpheme boundaries and, hence, the loss of internal morphological structure. This has happened in a number of Portuguese-based creoles in which the boundary between the root and the ‘Portuguese’ theme vowel has disappeared. As result, theme vowels have been reanalysed as a phoneme and integrated into the phonological form of the root. As we will show, this type of reanalysis seems to have taken place in the Sotavento varieties of Kabuverdianu (spoken in Cape Verde) and in Kriyol (spoken in Guiné-Bissau).

In Kriyol, each verb lexeme exhibits only one (unmarked) verb form. Verb forms therefore are invariable:

⁴ Thanks to Clancy Clements for bringing Spanish Romani to our attention.

- (5) a. *seta* b. *kume* c. *bibi* (Kihm 1994)
 ‘accept’ ‘eat’ ‘drink’

While each one of the verb forms in (5) appears to contain a verb root followed by a theme vowel (i.e., *set-a* (in 6a), *kum-e* (in 5b) and *bib-i* (in 5c)), closer inspection reveals that these verb forms are effectively monomorphemic. Given the lack of inflectional affixes, it cannot be claimed that the verbs in (5) have an internal morphological structure. Instead, we must conclude that the theme vowels have lost their inflectional status and have been integrated into the sound structure of the verb root. The same view is formulated by Kihm (1994) who points out that “in no case could these [Kriyol] vowels be analysed as affixes or thematic vowels as in Portuguese”.

In Sotavento Kabuverdianu, lexemes have two verb forms – one bare form and one inflected past form, as shown in ((6)-(8)). However, the invariable nature of both the suffix *-ba* and the verbal base also suggests that *falta-* (in 6b), *kore-* (in 7b) and *parti-* (in 8b)) are invariable and monomorphemic.

- (6) a. *falta* b. *falta-ba* (Baptista 2003; Baptista *et al.* 2007)
 ‘miss’ ‘miss-ant’
 (7) a. *kore* b. *kore-ba*
 ‘run’ ‘run-ant’
 (8) a. *parti* b. *parti-ba*
 ‘break’ ‘break-ant’

Assuming then that verb stems such as *falta-*, *kore-* and *parti-* cannot be segmented into smaller inflectional units, it seems plausible to treat the root and the stem as homophonous strings. Note also that the only attested suffix is invariable and attaches to an invariable base. Such lack of inflectional alternations clearly shows that the verbal paradigm of Sotavento Kabuverdianu is not organised into conjugation classes. Overall, then, while the *-a*, *-e* and *-i* vowels have derived from genuine superstrate theme vowels, empirical evidence indicates that they have lost their inflectional status in both Kriyol and Sotavento Kabuverdianu.

A lot has been said in the literature about the causes that may have been responsible for the loss of superstrate inflections in creoles. The striking surface similarities between adult interlanguages and creoles

suggest adult language learners play a crucial role in shaping the grammar of creoles (Plag 2008). Lending empirical support to an SLA-perspective of creolization is the striking surface similarity between adult interlanguages and creoles. For example, absence of overt inflection is not only a typical property of creoles but also of adult L2 varieties (DeGraff 2005; Prévost & White 2000). In adult L2 varieties, utterances consist largely of invariable lexical items which carry no inflectional affixes. Verbs exhibit no functional morphemes and therefore information about temporality is expressed lexically. Under one recent approach to creolization (Siegel 2004; Plag 2008), inflectionally poor creoles seem to have derived from such early stages of adult interlanguages, also known as ‘Basic Varieties’ (Klein & Perdue 1997).

3.2. Morphological levelling

Having provided evidence which shows that Romance theme vowels may lexicalize under contact, this section will now argue that language contact can also lead to the levelling of conjugation classes rather than to their complete loss.

Spanish Romani (also known as Caló) is a mixed contact variety which emerged from the contact between (Andalusian) Spanish and Romani (Boretzky & Iglá 1994; Bakker 2003). Sociolinguistically, Caló was created by ex-nomadic groups of Gypsies living in Andalusian Spain who needed a secret language to communicate without being understood by strangers (Winford 2003). As is typical of mixed languages, the grammar of Caló is derived from one of the contact languages while the lexicon is derived from the other. In other words, the grammar is Spanish grammar and the lexicon is Romani.

Focussing on the verbal paradigm, it is interesting to see that the intertwining of an L2 grammar with an L1 lexicon has given rise to verb forms with Romani roots and Spanish verbal endings. This mixture is illustrated in (9):

- (9) a. *para jal-a-r terel-a boque* (Boretzky & Iglá 1994: 55)
 to eat-cl1-inf have-cl1.3sg.prs hunger
 ‘in order to eat one has to be hungry’
- b. *mangue camél-a-ba* (Bakker 2003: 127)
 1sg.pro love-cl1-1sg.impf
 ‘I loved’

The Romani roots *jal-* ‘eat’ and *terel-* ‘have’, in (9a), and *camel-* ‘love’, in (9b), bear no resemblance to Spanish⁵. On the contrary, the verbal endings are clearly of Spanish origin. For example, the infinitive forms take the *-r* suffix (as in *jalar* ‘eat’), and the 1st person imperfect forms take the Spanish imperfect ending *-ba* (as in *camélaba* ‘I loved’).

Closer inspection of these verb forms also reveals that the inflected verb forms in (9) inflect according to the Spanish first conjugation. First, each one of the verb forms in (9) contains an intermediate *-a* stem which is typically associated to the first conjugation class. Second, verbs take first class verbal endings such as the imperfect marker *-ba*.

Table 1. Caló verbal template

ROOT	THEME VOWEL	VERBAL ENDINGS
<i>camel</i>	- <i>a</i>	- <i>ba</i>
love	- cl1	- 1SG.IMPF

Even though it seems that Caló has reduced the Spanish conjugation classes to one single class, the overall picture is slightly more complex. Very little data is unfortunately available about Caló, however evidence seems to suggest that Caló verbs borrow the theme vowel from another class for certain tense/aspect combinations. As noted by Bakker (2003), indefinite preterite forms use the theme vowel *-i*, as in (10), while the remaining paradigm uses the default theme vowel *-a*, as in (11):

- (10) a. *camel* -*i* (Bakker 2003)
 love -cl3.1sg.prt
 ‘I loved’
- b. *camel* -*i* -*ste*
 love -cl3 - 1sg.prt
 ‘s/he loved’
- (11) a. *camel* -*a* -*ba*
 ‘love- cl1 -1sg.impf’
 ‘I loved’
- b. *camel* -*a* -*ré*
 ‘love -cl1 -1sg.fut’
 ‘I will love’
- c. *camel* -*a* -*do*
 ‘love -cl1 -compl’
 ‘loved’

⁵ More Romani verb roots can be found in Rosensweig (1973: 38, 69, 110, ff).

Given (10) and (11), Caló lexemes draw their stems from the third conjugation class for the preterite forms⁶ and from the first class for all other tense/aspect combinations.

This phenomenon whereby lexical items take stems which belong to distinct inflection classes is known as heteroclisis. It can be found in a number of languages⁷: for example, in Latin the noun *balneum* 'bath' inflects as a second declension noun in the singular and as a first declension in the plural (Baerman 2007), and in some Romanian dialects, verbs which belonged originally to the third conjugation have acquired first conjugation morphology (Maiden 2008). The heteroclitic nature of lexemes raises interesting questions about the organisation of inflectional paradigms and has more recently attracted the attention of morphological theory (e.g., Stump 2006). For purpose of our paper, the presence of heteroclisis in Caló nicely shows that the kind of structure that characterizes contact varieties is by no means unusual but can be found in 'non-contact languages' as well.

Table 2. Indefinite preterite verb forms in Spanish

	conjugation class 1	conjugation class 3
1sg	<i>amé</i>	<i>dormí</i>
2sg	<i>amaste</i>	<i>dormiste</i>
3sg	<i>amó</i>	<i>dormió</i>
1pl	<i>amamos</i>	<i>dormimos</i>
2pl	<i>amasteis</i>	<i>dormisteis</i>
3pl	<i>amaron</i>	<i>durmieron</i>

As to the historical origin of heteroclisis, one can tentatively speculate about the forces that may have caused the *systematic* borrowing of third conjugation stems. To us, it seems plausible to assume, given the available data, that heteroclisis in Caló was driven by the paradigmatic uniformity of the third conjugation: preterite

⁶ Traditionally, the *i*-stem is classified as a third conjugation marker. However, the same stem is also found on second conjugation preterite verbs. This effectively means that the formal distinction between the second class and the third class has been neutralized in the Spanish preterite verb forms.

⁷ A searchable online database containing cross-linguistic data on heteroclisis can be found at www.cs.uky.edu/cgi-bin/cgiwrap/~gstump/heteroclisis.cgi

verb forms in the third conjugation take only one stem (namely the *i*-stem, as shown in Table 2), while preterite verb forms in the first class take at least *two* stems (namely, the default stem ending in *a*-, as in *amaste*, and the theme-less stem, as in *amé* and *amó*).

3.3. Retention of conjugation classes

In this section, we return to Portuguese-based creoles and examine the survival of conjugation classes in Indo-Portuguese⁸. In particular, we argue that the creoles of Korlai, Daman and Diu have preserved genuine theme vowels which function both as conjugation class markers and as stem-forming units (Luís 2008, forthcoming). The structure of Indo-Portuguese verbal paradigms is illustrated below (theme vowels are given in boldface):

Table 3. Daman Portuguese verbal paradigm (Clements and Koontz-Garboden 2002)

	conjugation class 1	conjugation class 2	conjugation class 3
unmarked forms	<i>kant-a</i> sing-cl1.unm 'sing'	<i>kum-e</i> eat-cl2.unm 'eat'	<i>sub-i</i> go up-cl3.unm 'go up'
past forms	<i>kant-o</i> sing-cl1-pst 'sang'	<i>kum-e-w</i> eat-cl2-pst 'ate'	<i>sub-i-w</i> go up-cl3-pst 'went up'
progressive forms	<i>kant-a-n</i> sing-cl1-prog 'singing'	<i>kum-e-n</i> eat-cl2-prog 'eating'	<i>sub-i-n</i> go up-cl3-prog 'going up'
completive forms	<i>kant-a-d</i> sing-cl1-compl 'sang'	<i>kum-i-d</i> eat-cl3-compl 'eaten'	<i>sub-i-d</i> go up-cl3-compl 'gone up'

By default, inflected verbs comprise three morphological units: a root, a theme vowel and a tense/aspect suffix. The verb root, which expresses the lexical meaning, is of Portuguese origin. Attached to the root is the theme vowel which has also been derived from Portuguese. Following the theme vowel, verb forms may also take a tense/aspect

⁸ The data in this section has been drawn from Clements (1996), Clements & Koontz-Garboden (2001), Cardoso (2006) and Schuchardt (1888).

marker which are all of Portuguese origin. The inventory of verbal endings is given in (12).

- (12) Indo-Portuguese tense/aspect suffixes
- a. Progressive suffix: *-n*
 - b. Completive suffix: *-d*
 - c. Past suffixes: *-o/ -w*

The progressive marker *-n* is derived from the Portuguese gerund marker *-ndo*, the completive marker *-d* from the Portuguese participle marker *-do*, and both past markers have been derived from the Portuguese 3rd person singular perfect endings. While the progressive suffix and the completive suffix are invariable, the past tense marker is sensitive to the conjugation class of the verb and exhibits two allomorphs⁹. These past allomorphs constitute an important piece of evidence in favour of our claim that conjugation classes have indeed survived in Indo-Portuguese. Note that the distribution of both past allomorphs is entirely determined by the conjugation class of the verbs: the *-w* suffix selects second and third conjugation stems, whereas the suffix *-o* selects first conjugation stems.

Underlying the paradigm in Table 3 is the ternary distinction between the first, second and third conjugation classes. In analogy to Portuguese, conjugation classes are identified inflectionally through a theme vowel. By default, class 1 stems take the theme vowel *-a*, class 2 stems take the *-e* vowel and class 3 stems take the *i-* vowel. As a result of the survival of theme vowels, verb forms such as *kumew* comprise three distinct morphological units: the root *kum-*, the stem *kume-* and the verbal ending *-w*. This means that, unlike in most Romance-based creoles, the Indo-Portuguese varieties spoken in Korlai, Daman and Diu make the distinction between roots, stems and inflected verb forms. This type of internal morphological structure is, to the best of our knowledge, unattested in other Portuguese-based creoles.

Further supporting our claim that Indo-Portuguese verb forms are organised into conjugation classes is the fact that verb stems undergo

⁹ In section 3.4, we show that the Indo-Portuguese creoles of Korlai and Daman have in effect one additional past allomorph which is associated to the fourth conjugation class.

allomorphic variation. As shown in Table 3, verbs belonging to the first and second conjugation classes take two stems, one primary stem and one secondary stem. In the first conjugation, past forms take a theme-less stem (e.g., *kant-o*), rather than the *a*-final stem (e.g., *kant-a-n*); in the second conjugation, participle verb forms take a third conjugation stem (i.e., *kum-i-d* not **kum-e-d*)¹⁰, rather than the default *e*-final stem (e.g., *kum-e-n* and *kum-e-w*). Stem-alternations such as these are not triggered by semantic or morphosyntactic factors, but merely by the conjugation distinctions underlying the Indo-Portuguese verbal paradigm.

Under the assumption that adult language learners play a crucial role in shaping the grammar of creoles (cf. section 3.1), recent studies have tried to account for the retention of purely formal inflectional distinctions in Indo-Portuguese creoles from a multidimensional perspective, combining untutored second language learning with both socio-cultural and attitudinal factors. Clements (1996) takes the view that favourable socio-cultural forces, especially, the isolation of substrate speakers from their native community, may have triggered a favourable attitude of substrate speakers towards the Target Language (TL), encouraging adult Marathi and Gujarati speakers to learn more than just a rudimentary version of it.

The idea that adult Second Language Acquisition can be influenced by the receptive attitude of the adult learners has also been observed by Siegel (2004) who argues that “limited second language acquisition” may also result from “the limits learners impose on themselves due to factors relating to identity or resistance”, rather than from the limited degree of exposure to the TL. So, adult language learners who refuse to identify with the TL community tend to learn an L2 variety for basic communicative purposes. In Klein & Perdue (1997), such a rudimentary interlanguage, also known as ‘Basic Variety’, is deprived of overt inflectional morphology. So, creoles with very little or no overt

¹⁰ Despite the remarkable extent of overt verbal inflection found in Indo-Portuguese, the inventory of verbal endings is still reduced when compared to Portuguese. No person/number suffixes have survived and several tense/mood/aspect markers have also been lost, as shown in Table 3.

inflection (such as Kriyol, see section 3.1) may have derived from a rudimentary or early stage of the adult L2 interlanguage.

On the contrary, as argued in Luís (forthcoming), Indo-Portuguese creoles derived from “Post-Basic Varieties” (Klein & Perdue 1997), i.e., elaborate interlanguages which already contained overt verbal inflection. Assuming that the resistance of substrate speakers towards the TL prevents adults from going past the ‘Basic Variety’, it seems plausible to assume that the favourable socio-cultural contexts addressed in Clements (1996) may have triggered the emergence of a more elaborate interlanguage in Indo-Portuguese.

3.4. Paradigm extension

In the final section of this paper, we look at the emergence of a new conjugation class. Returning to Indo-Portuguese creoles, we argue that an additional class has emerged in Korlai Portuguese and Daman Portuguese for verbs of substrate/adstrate origin (Clements 1996; Clements & Koontz-Garboden 2002).

In (13), verbs such as *badlu* ‘change’, *pislu* ‘get angry’ or *samdzu* ‘understand’, as given in (13), are taken over from Marathi, the substrate/adstrate language of Korlai Portuguese.

(13) (Clements 1996:137)

- | | | |
|-----------------|-----------------|------------------|
| a. <i>badlu</i> | b. <i>pislu</i> | c. <i>samdzu</i> |
| ‘change’ | ‘get angry’ | ‘understand’ |

Two morphological properties distinguish this new class from the classes which take verbs of Portuguese origin: a) they take an unmarked past form and b) they insert a ‘new’ theme vowel *-u* between the borrowed root and the verbal endings.

Starting with the verbal endings, Table 4 shows that there are four conjugation classes in Korlai: three classes of Portuguese origin and one ‘new’ class. All four classes share common verbal endings, namely, the completive suffix *-d*, the progressive suffix *-n* and the unmarked form.

Table 4. Verbal endings in Korlai and Daman Portuguese

	unmarked form	past form	progressive form	completive form
conjugation class 1/2/3	-	-o/-w	-n	-d
conjugation class 4	-	-	-n	-d

However, there are also important differences. The fourth conjugation class has no overt past marker. In other words, both the base form and the past form are syncretic, that is, they express different morphosyntactic features but share the same morphological form (see Table 5). This new past form, which is restricted to borrowed verbs, constitutes robust evidence in favour of our claim that there are in effect three past allomorphs and four conjugation classes in Korlai Portuguese and in Daman Portuguese.

Table 5. Marked and unmarked past forms in Korlai Portuguese (Clements 1996: 137)

	conjugation class 1	conjugation class 2	conjugation class 3	conjugation class 4
unmarked form	<i>kata</i> 'sing'	<i>kume</i> 'eat'	<i>subi</i> 'go up'	<i>tepu</i> 'heat up'
past form	<i>kato</i> 'sang'	<i>kumew</i> 'ate'	<i>subiw</i> 'went up'	<i>tepu</i> 'heated up'

While verbs of Portuguese origin take either the suffix *-o* or the suffix *-w* (cf. section 3.3), verbs of Marathi/Gujarati origin take a zero allomorph. Technically, it would be difficult to account for the distribution of the zero past allomorph without assuming that there is an additional conjugation class (Luís 2008).

As alluded to above, the fourth class bears its own class marker, namely the theme vowel *-u*, which occurs immediately after the root and before the verbal endings. The emergence of this theme vowel constitutes a clear case of paradigm-internal innovation and could not have been taken over the contributing languages: on the one hand, Portuguese does not have an *-u* theme vowel and, on the other hand, neither Marathi nor Gujarati have theme vowels (Masica 1993). The

question therefore remains as to why these creoles created an entirely new theme vowel. Studies on loanword adaptation have shown that languages tend to adopt different strategies to integrate borrowed lexical items: such integration strategies can be either marked or unmarked. In Portuguese, for example, loan verbs are integrated into the more productive *-a* conjugation class (e.g., *grafitar* 'to draw graffiti') in a clearly unmarked way.

Indo-Portuguese creoles, however, seem to have adopted a marked strategy, by developing a theme vowel which is exclusively attached to verbs of Marathi roots (for Korlai Portuguese) and to Gujarati roots (for Daman Portuguese)¹¹. In Luís (2008), we have argued that the *-u* vowel has been selected because it is phonetically distant from the already available *-a*, *-e* and *-i* vowels which have been taken over from Portuguese. The selection of *-u* ensures that each conjugation class has a perceptually distinct vowel (Crothers 1978). In these languages, then, a four-vowel system has emerged which preserves maximum acoustic differentiation between native (Portuguese) roots and borrowed roots.

4. Conclusion

Some domains of linguistic structure have been particularly short of empirical evidence and, hence, our understanding of their ability to undergo change has been somewhat limited. One such area of grammar has been examined in this paper, namely verbal inflection and, most especially, the organisation of verbal paradigms into conjugation classes. Despite the reduced amount of data, this paper has shown that Romance contact varieties provide a rare opportunity for linguists to investigate the fate of Romance conjugation classes under language contact.

Our study has revealed that conjugation classes may undergo different types of morphological change. In some contact-varieties, most frequently creoles, theme vowels have been integrated into

¹¹ Interestingly, verbs borrowed from other languages, including English, are not integrated into the creoles through the conjugation class system, but are accommodated syntactically through periphrastic constructions.

the phonological structure of the root, leading to the complete loss of conjugation classes. In Spanish-Romani, a mixed language, conjugations have suffered a significant reduction through paradigm levelling, however they have not been lexicalised. In Indo-Portuguese, on the contrary, Romance conjugation classes have not only completely survived but they have also been extended through morphological innovation. Overall, then, purely formal (and hence meaningless) inflections do not necessarily disappear under language contact.

References

- Aronoff, M. 1994. *Morphology by Itself*. Cambridge MA: The MIT Press.
- Baerman, M. 2007. Morphological typology of deponency. In: M. Baerman *et al.* (Eds.). *Deponency and morphological mismatches*. Oxford: Oxford University Press, 1-19.
- Bakker, P. 2003. Mixed languages as autonomous systems. In: Y. Matras; P. Bakker (Eds.). *The Mixed Language Debate*. Berlin: Mouton de Gruyter, 107-150.
- Baptista, M. 2003. Inflectional plural marking in pidgins and creoles: a comparative study. In: I. Plag (Ed.). *Phonology and Morphology of Creole Languages*. *Linguistische Arbeiten* 478. Tübingen: Max Niemeyer, 315-332.
- Baptista, M.; Mello, H.; Suzuki, M. 2007. Cape Verdean Creole and Guinea-Bissau Creole. In: J. Holm; P. Patrick (Eds.) *Comparative Creole Syntax*. London: Battlebridge, 53-82.
- Boretzky, N.; Iglá, B. 1994. Romani Mixed Dialects. In: P. Bakker; M. Maarten (Eds.). *Mixed Languages. 15 Case Studies in Language Intertwining*. Amsterdam: IFOTT, 35-68.
- Cardoso, H. 2006. Diu Indo-Portuguese: change and continuity since 1882. Paper presented at the meeting of the Associação de Crioulos de Base Lexical Portuguesa e Espanhola, University of Coimbra.
- Clements, J. C. 1996. *The Genesis of a Language: The formation and development of Korlai Portuguese*. Amsterdam/Philadelphia: John Benjamins.
- Clements, J. C.; Koontz-Garboden, A. 2002. Two Indo-Portuguese Creoles in Contrast. *Journal of Pidgin and Creole Languages*. **17(2)**. Amsterdam: John Benjamins, 191-236.
- Crothers, J. 1978. Typology and Universals of Vowel Systems. In: J. Greenberg (Ed.). *Universals of Human Language*. Stanford CA: Stanford University Press, Vol. 2, 93-152.
- DeGraff, M. 2005. Morphology and word order in 'creolization' and beyond. In: G. Cinque; R. S. Kayne (Eds.). *The Oxford Handbook of Comparative Syntax*. Oxford: OUP, 293-372.

- Kihm, A. 1994. *Kriyol Syntax. The Portuguese-based Creole Language of Guiné-Bissau*. Amsterdam: John Benjamins.
- Klein, W.; Perdue, C. 1997. The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research*. **13(4)**: 301-47.
- Luis, A. R. forthcoming. Tense marking and inflectional morphology in Indo-Portuguese creoles. In: S. Michaelis (Ed.). *Roots of creole structures: weighing the contribution of substrates and superstrates*. Amsterdam: John Benjamins.
- Luis, A. R. 2008. Conjugation classes in Portuguese-based creoles. Paper given at the meeting of the Associação de Crioulos de Base Lexical Portuguesa e Espanhola, Universidade Nova de Lisboa, July 4.
- Maiden, M. 2008. Why heteroclisis? Paper given at the International Conference on Continuity and Change in Grammar, University of Cambridge, March.
- Masica, C. P. 1993. *The Indo-Aryan Languages*. Cambridge: CUP.
- Matras, Y.; Bakker, P. (Eds., 2003). *The Mixed Language Debate*. Berlin: Mouton de Gruyter.
- McWhorter, J. H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology*. **5**: 125-166.
- Plag, I. 2008. Creoles as interlanguages: inflectional morphology. *Journal of Pidgin and Creole Languages*. **23(1)**: 114-131.
- Prévost, P.; White, L. 2000. Missing Surface Inflection or Impairment in Second Language Acquisition? Evidence from Tense and Agreement. *Second Language Research*. **16(2)**: 103-133.
- Rosenzweig, J. B. *Caló. Gutter Spanish*. New York: E. P. Dutton.
- Schuchardt, H. 1883. Kreolische Studien III. Ueber das Indoportugiesische von Diu. In: *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften zu Wien (philosophisch-historische Klasse)* 103, 3-18.
- Seuren, P.; Wekker, H. 1986. Semantic transparency as a factor in creole genesis. In: P. Muysken; N. Smith (Eds.) *Substrata versus universals in creole genesis*. Amsterdam: John Benjamins, 57-70.
- Siegel, J. 2004. Morphological simplicity in pidgins and creoles. *Journal of Pidgin and Creole Languages*. **19(1)**. Amsterdam: John Benjamins, 139-162.
- Stump, G. 2006. Heteroclisis and paradigm linkage. *Language*. **82(2)**: 279-322.
- Winford, D. 2003. *An Introduction to Contact Linguistics*. Oxford: Blackwell.
- Veenstra, T. 2003. What verbal morphology can tell us about creole genesis. In: I. Plag (Ed.). *Phonology and Morphology of Creole Languages*. Linguistische Arbeiten 478. Tübingen: Max Niemeyer, 293-313.

O purismo e a estandarización
como factores condicionantes da mudanza lingüística.
O caso dos plurais dos nomes polisílabos agudos
terminados en /-l/ no singular no galego contemporáneo

Ramón Mariño Paz
ramon.marino.paz@usc.es
Universidade de Santiago de Compostela (Galicia – España)

RESUMO. Neste traballo analízase como o purismo e a estandarización influíron no moderno desenvolvemento dunha variable morfolóxica da lingua galega: o plural dos nomes polisílabos agudos terminados en /-l/ no singular.

PALABRAS-CHAVE. Purismo lingüístico, estandarización, lingua galega.

ABSTRACT. In this paper I study how purism and standardization influenced in the modern development of a morphological variable of the Galician language: the plural of the polysyllabic acute nouns finished in /-l/ in their singular form.

KEY-WORDS. Linguistic purism, standardization, Galician language.

1. Obxecto de estudo e metodoloxía

No galego medieval estes plurais adoitaban formarse coa supresión da consoante lateral final e a adición dos alomorfos *-es* (*leal – leaes, cabedel – cabedeas...*) ou *-is* (*gentil – gentijs*), mentres que a innovación morfolóxica consistente en manter no plural a consoante final do singular e engadir o alomorfo *-es* (*catedral – catedrales, sotil – sotiles*) presentábase en xeral en porcentaxes residuais. Durante os séculos XVI, XVII e XVIII produciuse unha certa progresión dos plurais en terminados *-les* para o tipo de nomes polisílabos agudos que se estudan neste traballo, mais daquela aínda gozaban de grande extensión as solucións que resolveran mediante asimilación e coalescencia (ou só coalescencia) os antigos hiatos <aes>, <ees>, <iis> e <oes>: *reaes > reaas > reás, pichees > pichés, gentiis > gentís, españoos > españoos > españos* (Mariño Paz 2004).

O incremento do número de individuos bilingües con coñecemento activo do galego e do castelán que se produciu sobre todo no século XX propiciou o avance da variante desta variable morfolóxica que implicaba unha converxencia co único idioma oficial en España durante a maior parte da Idade Contemporánea (*animales, españoles...*), nomeadamente nos contextos vilegos e urbanos¹. Ao mesmo tempo, o cultivo rexenerador do galego sustentado polo galeguismo desde mediados do XIX supuxo a posta en práctica de actitudes puristas que intentaron descastelanizar o idioma de diversas maneiras. Aprobada a Constitución española de 1978, Galicia converteuse nunha Comunidade Autónoma con rango de histórica e o seu Estatuto de Autonomía de 1981 declarou o galego lingua cooficial no seu territorio, o que supuxo que se lle reservasen distintos usos de carácter formal: nas administracións públicas, no ensino, na vida política, nos medios de comunicación públicos e nalgúns privados, etc. En 1982 publicáronse as *Normas ortográficas e morfolóxicas do idioma galego* da Real Academia Galega (RAG) e o Instituto da Lingua Galega (ILG), axiña declaradas oficiais. O primeiro dos catro principios que se presentaron como inspiradores destas *Normas* era unha transparente declaración de purismo descastelanizador:

¹ Non me proponho abrir aquí unha discusión sobre se esta innovación se debe interpretar ou non como un préstamo estrutural que o galego tomou do castelán. Limitareime por tanto a afirmar que, se é certo que o préstamo de morfemas flexionais esixe entre as linguas doadora e receptora un contacto máis intenso ca o que pide o préstamo de morfemas derivativos ou de elementos fónicos e lexicais (McMahon 1994: 209-213), o contacto galego-castelán ten xa en Galicia antigüidade e intensidade abundas para producir transferencias de morfemas flexionais. Máis aínda, no último século produciuse un rápido aumento do número de individuos con competencia activa en ambas as linguas que creou condicións idóneas para que no galego xurdan e se espallen préstamos estruturais de distinto tipo tomados do castelán ou para que avancen con máis rapidez aquelas variantes lingüísticas que, aínda que fosen xeradas autonomamente polo galego, camiñan na dirección da converxencia co castelán. En todo caso, o que me interesa subliñar aquí é que ao longo dos séculos XIX e XX diversos observadores tiveron a percepción de que a variante *animales* se emprega en galego por castelanismo e que, en consecuencia, non foron poucos os que, como resposta a tal percepción, utilizaron ou chamaron a utilizar variantes deste plural sobre as que non pesase tal inconveniente.

A lingua normativa ten que estar ó servizo da cultura dun pobo real e concreto, e por tanto ha de ser necesariamente continuadora da lingua falada pola comunidade e ha de achegarse canto sexa posible a ela, a fin de enraizar sobre bases seguras e vivas. Agora ben, para que estas bases sexan efectivamente sólidas, a normativa debe acoller un galego fiel a si mesmo e limpo de canto de espúreo hai incrustado na fala viva pola presión do castelán. É dicir, a lingua común debe se-lo máis galega posible, establecida sen sometementos nin dependencias, con atención escrupulosa á estrutura e características lingüísticas do galego (Real Academia Galega & Instituto da Lingua Galega 1982: 7).

Para a variable morfolóxica estudada neste traballo, a selección da variante *animais*, minoritaria no conxunto das falas populares, como a propia do estándar galego configurado e socializado a partir da década de 1980 puxo en xogo un novidoso e importante factor sociolingüístico que parece estar a reorientar a secular marcha desta mudanza. Desde hai un tempo, a variante seleccionada para estes plurais no estándar galego vai sendo incorporada aos repertorios lingüísticos de certos falantes con nivel de estudos medio ou alto e formas de vida urbanas, de tal xeito que nalgúns medios estes plurais terminados en *-is* van entrando en variación contextual coa solución popular converxente coa do español (*animales*). Por outra parte, os plurais do estándar dominan case por completo no galego escrito, literario ou non, desde a década de 1980 ata hoxe.

Neste traballo vou estudar a evolución desta mudanza lingüística desde a época da entrada en escena do galeguismo (arredor de 1840) ata a actualidade. O meu propósito é o de explicar como, fronte ás tendencias populares, alzáronse durante este tempo dous factores que promocionaron solucións alternativas libres de toda sospeita de seren castelanismos. Refírome, en primeiro lugar, ao purismo lingüístico, máis ou menos acentuado segundo as épocas pero propio en xeral do cultivo literario do galego desde mediados do século XIX ata a actualidade, e, xunto a el, ao proceso de socialización do estándar galego a partir da década de 1980, que asumiu o purismo descastelanizador como o primeiro dos seus principios inspiradores. As fontes de estudo que sometín á análise foron, dunha parte, unha ampla representación de traballos dos séculos XIX, XX e XXI que, guiados pola intención de describir a lingua galega ou de prescribir sobre o seu correcto uso, achegan información de interese para quen queira

trazar a liña evolutiva desta variable morfolóxica galega nos últimos cento sesenta ou cento setenta anos, aproximadamente. Alén disto, analicei o comportamento desta variable en textos, literarios ou non, escritos entre 1840 e a actualidade.

2. Datos das fontes metalingüísticas

Entre as fontes que explorei figuran, por un lado, gramáticas sincrónicas e propostas normativas animadas por un prescristivismo e un purismo case sempre explícitos (Brea 1994; Freixeiro Mato 1998: 31-37; Sánchez Rei 2005; Santamarina 1974) e, por outro, gramáticas diacrónicas e modernos traballos de investigación. As variantes da desinencia para a formación destes plurais que detectei nestas fontes foron as seguintes: 1) *-s*, con supresión do *-l* final do singular; 2) *-is*, coa mesma supresión; 3) *-es*, coa mesma supresión; 4) *-es*, sen tal supresión; e 5) *-s*, sen tal supresión.

2.1. *-s*, con supresión do *-l* final do singular

A utilización da variante *-s* con supresión do *-l* final do singular (*animal* – *animás*, *papel* – *papés*, *funil* – *funís*, *español* – *españós*, *azul* – *azús*) é a única norma que proclaman Saco Arce (1868), Santiago y Gómez (1918), R. A. (1919) e Carballo Calero (1966), aínda que este último só ata a cuarta edición da súa gramática, de 1974. Valladares (1892), Cornu (1906) e Lugrís Freire (1931) tamén a recoñeceron, mais colocando ao seu carón outras posibilidades: a 3 nos casos de Valladares e Lugrís e a 3, a 4 e a 5 no de Cornu. Algúns dos gramáticos da segunda metade do século XX só contemplaron esta variante en excepcións (*cadriés*, *papés*, *perniés*, *reás*...) que se apartaban da regra que describiron para o que eles consideraban galego exemplar ou, despois de 1982, galego estándar. Do meu punto de vista, estas excepcións non son máis ca reliquias de sincronías pasadas en que este procedemento de formación do plural, caduco nas últimas décadas do novecientos, aínda gozaba entón de gran vitalidade.

Os datos recadados nos outros tipos de fontes exploradas conducen á conclusión de que esta variante gozou dun elevado nivel de uso na lingua literaria decimonónica que probablemente se correspondía coa súa conservación aínda moi difundida na lingua popular da Galicia

centro-occidental. As fontes escritas da primeira metade do oitocentos amosan que esta regra aínda era daquela produtiva cos neoloxismos (*constitucionás, servís* ‘partidarios do absolutismo’, etc.), ao tempo que con unidades lexicais como *reás*, plural de *real* ‘unidade monetaria’, e *papés* o seu mantemento maniféstase con particular firmeza en toda a documentación decimonónica conservada. No entanto, o *Atlas Lingüístico Galego* (ALGa 2: mapas 71, 72, 73, 74, 75, 76, 77), Francisco Fernández Rei (1990: 65) e Dubert García (1999: 96) mostran que no derradeiro cuartel do século XX esta variante estaba xa próxima a sucumbir ante o asoballante progreso que nas décadas anteriores viñera coñecendo *-les* na lingua popular. Quedaban entón, con todo, certas unidades lexicais en que os tradicionais plurais centro-occidentais rematados en *-ás, -és*, etc. aínda conservaban unha desigual frecuencia de uso: *currás, ladrás, reás, candís, cadrís, catedrás, fusís*... Como ben se sabe, isto é algo que se observa en moitos procesos de cambio lingüístico: é frecuente, mesmo habitual, que as distintas unidades susceptibles de resultaren afectadas por el non se transformen simultaneamente, senón con ritmos distintos e con desigual resistencia ao avance da innovación. As innovacións esténdense por medio de procesos de difusión léxica que non alcanzan nin transforman todas as palabras ao mesmo tempo nin coa mesma rapidez e que mesmo poden deixar que algunhas delas se libren do seu influxo. Neste caso, dá a impresión de que a variante tradicional resistiu máis firmemente o embate da innovación terminada en *-les* en dous casos: 1) co vocabulario elemental máis vinculado ás formas de vida tradicionais das clases populares: *animás, currás, ladrás, reás, papés, cadrís, candís, fusís*, etc.; 2) cos nomes utilizados a miúdo ou case en exclusiva en plural (*reás, papés*, etc.), con algúns dos cales mesmo puido chegar a quedar escurecida a existencia da base singular (caso, por exemplo, de *cadril* ‘cada un dos ósos da cadeira’ e de *ladral* < LATERALE ‘cada unha das táboas laterais do carro galego tradicional que serven para suxeitar a carga’).

Esta variante para a formación do plural é a que prescribiu o Seminario de Estudos Galegos (SEG) en 1933, e aínda a que a RAG utilizou na redacción do seu opúsculo ortográfico de 1970, no que esta institución non se pronunciou explicitamente verbo desta variable morfolóxica. Así mesmo, é a que o ILG recoñeceu no *Gallego 1*

(1971) e no *Gallego 2* (1972) só para os substantivos *cadrís* e *reás*, que presentaba como excepcións que escapaban á regra xeral de formar estes plurais coa adición do morfema *-es* e o mantemento do *-l* (*animales*). Non hai prescrición a favor desta variante 1 (*animás*) en ningunha das propostas publicadas posteriormente.

2.2. *-is, con supresión do -l final do singular*

Probablemente por ser minoritaria no uso popular, esta variante non foi declarada de regra por ningún gramático antes da cuarta edición da obra de Carballo Calero, datada en 1974 (Carballo Calero 1974: 160). Deste xeito, Carballo abandonaba a súa primixenia predilección polo plural con supresión de *-l* e adición de *-s* (*animás*) e apostaba por aquel que xa postulara nas normas morfolóxicas da Real Academia Galega de 1971, das que fora o principal inspirador. Sucedía isto nun momento avanzado do proceso de elaboración non planificada da variedade estándar do galego moderno, cando ao impulso purista que o viñera animando desde o século XIX se lle empezaba a unir, logo da incorporación da investigación sobre a lingua galega á Universidade de Santiago de Compostela na década de 1960, un maior coñecemento da súa variación diacrónica e diatópica. As gramáticas aparecidas despois da publicación das *Normas* da RAG e o ILG en 1982 reproducen a doutrina que a este respecto se oficializou naquel ano no relativo aos nomes oxítonos polisílabos: *animais*, *papeis*, *barrís*, *españóis* ou *azuis* son as formas canonizadas e prescritas en todas elas.

Esta variante fora xa a prescrita en 1971 pola RAG, institución que, no entanto, un ano antes empregara a solución *animás* na redacción do seu folleto ortográfico de 1970. ¿Cales foron as razóns que motivaron esta mudanza de orientación normativa, que non soamente rompía co que anteriormente fixera a mesma RAG e coas normas do SEG de 1933, senón tamén coa xeneralidade do discurso gramatical galego? Na argumentación que acompañaba esta proposta da Academia manexábanse os seguintes argumentos para motivala: a) *animais* é unha variante viva no galego moderno (característica das terras orientais, mais non “violenta” para ningún galegofalante); b) permite a harmonización co portugués; e c) ten sobrada tradición na documentación histórica. Aínda que isto non se explicitaba, desta

explicación inferíase que a solución *animás* xa non gozaba daquela do respaldo dun uso popular estendido e tamén que a súa adopción non propiciaba un achegamento ao portugués. A variante *animais* foi tamén propugnada nos textos prescritivos posteriores, coas únicas excepcións do *Gallego 1* (1971) e do *Gallego 2* (1972), dúas obras en que o galego que se ensinaba estaba moi achegado ao uso popular predominante e en que a intervención purista se aplicaba con moita medida. Nas *Bases* de 1977, ao argumento da harmonización co portugués agregáronse, para defender a solución rematada en *-is*, o de que era máis “arcaizante”, o de que estaba “plenamente recuperada” “no galego escrito dos últimos tempos” e o de que a terminada en *-s* e sen *-l* era xa unha opción en avanzado estado de decadencia. Finalmente, na argumentación das *Normas* da RAG e o ILG (1982: 52) insístiuse na idea de que, sendo tanto *animás* coma *animais* variantes galegas tradicionais, a primeira delas presentaba o inconveniente de se achar nun adiantado estado de abandono.

Logo de ser seleccionada para a variedade estándar nas *Normas* da RAG e do ILG, axiña difundidas a través de diversas vías ata o conxunto da poboación, cara ao final da década de 1990 esta variante parecía estar a comportarse xa como unha innovación propagada “desde arriba”. No seu estudo da fala do Santiago de Compostela finisecular Dubert García (1999: 94-95) encontrouse con “falantes con estudos e con abundantes trazos lingüísticos de fasquía urbana” que, situados nun contexto formal, utilizaban esta variante (*azuis, espirituais, municipais...*) en alternancia coa popular *-les*. Ao seu parecer, isto demostraba “que algúns procesos morfolóxicos da lingua estándar comeza[ba]n a estenderse nalgúns medios sociais e a entrar en variación contextual coas formas populares, de modo que o galego estándar comeza[ba] a compartir co castelán o papel de lingua teito e a competir con el na creación de formas de prestixio”. Para estes falantes, a utilización da variable *animales – animais* funcionaba xa como un marcador de status. Non hai motivos que nos induzan a pensar que o que Dubert sinalou para o dialecto de Compostela non se estivese a producir noutras tamén zonas do país. Cara a isto apuntan algunhas observacións realizadas nos últimos anos polo propio Dubert García (2002: 22-23) e por Regueira (1999: 865, 867), quen ademais subliña o esperable uso case exclusivo da variante normativa

animais na lingua escrita e publicada de 1980 en diante (Regueira 2003: 210; 2005: 81).

2.3. *-es, con supresión do -l final do singular*

A pesar de ser unha variante de moi escaso uso no moderno galego popular, Valladares (1892) e Carré Alvarellos (1967) chegaron a prescribir o seu uso, aínda que só en concorrencia con algunha outra opción. Con todo, o seu carácter “arcaico”, “histórico” ou antigo non pasou desapercibido para Lugrís Freire (1931), Carballo Calero (1966), García de Diego (1909) ou Ferreiro (1995), e as investigacións realizadas desde a década de 1970 confirmárona como fundamentalmente antiga e só conservada nalgunhas falas orientais do galego popular de hoxe². Nas gramáticas sincrónicas publicadas de 1986 en diante xa se omite toda referencia a ela. Non estraña, por tanto, que ningunha das propostas normativas exploradas apoiase o seu uso.

2.4. *-es, sen supresión do -l final do singular*

A prescrición a favor desta variante só a atopamos, con certas restricións, nas obras de Lugrís e Carré. Poderíamos dicir, por tanto, que semella aparecer nas gramáticas sincrónicas tardiamente, cando a súa difusión na lingua popular xa debía de estar bastante avanzada, e desaparece delas cando, constituído o galego como lingua cooficial de Galicia, instáurase un modelo normativo animado por un espírito descastelanizador que levou a optar por unha variante tradicional que non podía suscitar ningún tipo de receo entre os máis puristas. Foi precisamente esta prevención a razón pola que esta solución gozou de escaso apoio nas propostas normativas realizadas. O seu uso só o promoveron os métodos de aprendizaxe *Gallego 1* e *Gallego 2* do ILG, onde con todo se recoñeceron as excepcións de *reás*, *cadrís* e *adrais*, das que se aseguraba que conservaban unha alta frecuencia de uso popular; a decisión do ILG baseábase no feito de que a solución *-les* representaba “a tendencia moderna da lingua”, pois estaba “moi estendida popularmente” (García 1971: 139-140). A observación sobre

² Para o período final do século XX o ALGa 2 só a amosa nalgúns puntos do leste, case sempre en concorrencia coa variante *-is* con *-l* suprimido (*currais* – *curraes*, *catedrais* – *catedraes*, *reais* – *reaes*, etc.) ou coa variante con *-les* (*reales* – *reaes*).

a gran difusión popular desta variante reiterouse nas *Bases* (que a declararon alternativa admisible – mais non recomendable – de *-is*) e nas *Normas* de 1982 (que a rexeitaron por completo).

Debémolles a Carballo Calero (1966: 91-92, n. 6) a opinión, concordante cos datos que as investigacións dialectolóxicas nos proporcionarían moi pouco despois (ALGa 2; Fernández Rei 1990: 65), de que a mediados da década de 1960 a variante *-les* xa era maioritaria na lingua popular. Porén, os seus rexistros xa non son raros nalgunhas gramáticas do XIX ou do XX cuxos autores, aínda que non optasen explicitamente a prol dela, aduciron de vez en cando exemplos terminados en *-les*, que son especialmente frecuentes na terminoloxía gramatical que manexaron R. A. (*adverbiales, causales, condicionales...*) e Lugrís (*adverbiales, nasales...*)³. Os traballos de investigación que consultei permiten afirmar que, mediado o século XIX, non se chegara aínda ao punto de se producir unha rápida expansión desta variante, cousa que debeu de acontecer máis tardiamente, a son co veloz avance do coñecemento e o uso do castelán que se deu en Galicia sobre todo no século XX. O seu progreso debeu de producirse máis rapidamente cos neoloxismos (*faroles* en lugar de *farós*, *chavales* ‘rapaces’ no canto de *chavás...*), cos castelanismos flagrantes (*Madriles* por *Madrid*) e co léxico máis vencellado co mundo e a ideoloxía oficial do estado-nación español en proceso de construción desde a primeira metade do XIX (*españoles* en vez do antigo *españós*). Todo este léxico contou con axentes propagadores tan eficaces coma o ensino (cada vez máis estendido), a prensa escrita, os modernos medios de comunicación audiovisual ou o castelán falado nas vilas e cidades polas persoas influentes.

³ Os substantivos compostos que teñen como segundo lexema un monosílabo débátese entre a selección da variante *-es* con *-l*, esperable en virtude da súa análise como unidades bimembres cun segundo elemento monosilábico, e o seu comportamento como unidades polisilábicas, evidenciado polo emprego de solucións con eliminación da consoante lateral. A primeira opción rexístrase nas obras de Saco Arce (*pirifoles, pousa-foles*) e Carré Alvarellos (*zugameles, pousa-foles*), na sétima edición da de Carballo Calero (*pirifoles*) e nas gramáticas posteriores a esta que se ateen á letra das *Normas* da RAG e o ILG de 1982; a segunda, na de Valladares (*pirifòes, pirifòs*) e nas seis primeiras edicións da de Carballo Calero (*pirifòs*).

2.5. *-s, sen supresión do -l final do singular*

É unha variante infrecuente no galego escrito dos séculos XIX e XX, apenas mencionada nas gramáticas revisadas e non seleccionada en ningunha proposta normativa. Segundo o ALGa 2 e Fernández Rei (1990: 65), nos últimos tempos aínda se usa nalgúns falares populares do occidente, sobre todo en puntos da provincia da Coruña en que esporadicamente se oen *currals, catedrals*, etc.

3. Datos das fontes literarias

Se os datos expostos na sección anterior permiten recompoñer parcialmente a andadura desta mudanza lingüística no galego oral, sobre todo popular, e tamén no galego escrito de case os últimos dous séculos da historia de Galicia, é evidente que os que deseguido se van presentar achegan información sobre o comportamento desta variable morfolóxica no moderno galego escrito; ora ben, coido que colateralmente pode derivarse desta información algunha hipótese non demasiado desencamiñada a propósito do que a respecto disto debeu de ocorrer nos contextos – non moi numerosos antes da década de 1970 e certamente raros no século XIX e nos inicios do XX – en que se fixese un uso oral formal, esixente ou pulido da lingua galega. Aproximareime a estas cuestións por medio de oito figuras que amosan as porcentaxes de uso alcanzadas polas distintas variantes dos plurais de oito nomes polisílabos oxítonos no galego escrito entre 1840 e 2002. Como o que aquí nos importa é só o comportamento da variable morfolóxica que estamos a analizar, agrupei nunha soa entrada todas as variantes fonéticas de cada palabra que comparten a mesma solución para o morfema de plural: así, por exemplo, *reás, reiás, raás, raiás, riás* e *rás* presentareinas unificadamente baixo a entrada *reás*. Os datos necesarios para elaborar estas figuras recadeinos no *Tesouro Informatizado da Lingua Galega* (TILG).

Distinguín catro grandes etapas neste período de 163 anos. A primeira (1840-1915) correspóndese con toda a época anterior á irrupción en escena do nacionalismo galego. Nesta etapa o purismo lingüístico, aínda non sendo descoñecido, non chegou a atinxir a intensidade que alcanzaría entre 1916 e 1936, cando o cultivo do idioma galego, promovido en boa medida – mais non só – polo

nacente nacionalismo, adoitou facerse con esixencias de intelectua-
lización, desruralización e descastelanización. Nos anos posteriores
á instauración do franquismo en Galicia en xullo de 1936 foise
conseguido – moi devagar ata a década de 1950 – unha recupera-
ción do galego escrito, recuperación que a partir dos anos sesenta
se fixo xa en consonancia coas orientacións que naqueles tempos
foron proporcionando gramáticas e propostas normativas. Por fin, a
publicación das *Normas ortográficas e morfolóxicas do idioma galego*
(1982) marcou o inicio do proceso de estandarización planificado.

Figura 1. Plurais do substantivo *animal* no galego escrito entre 1840 e 2002.

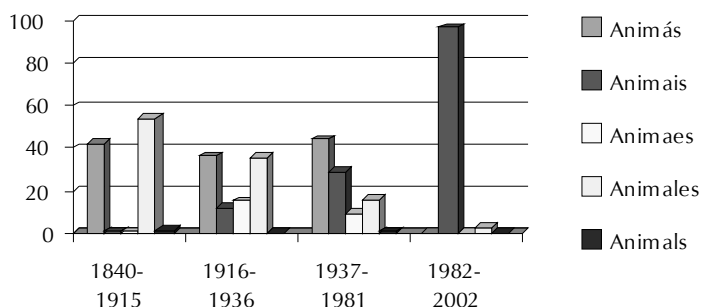
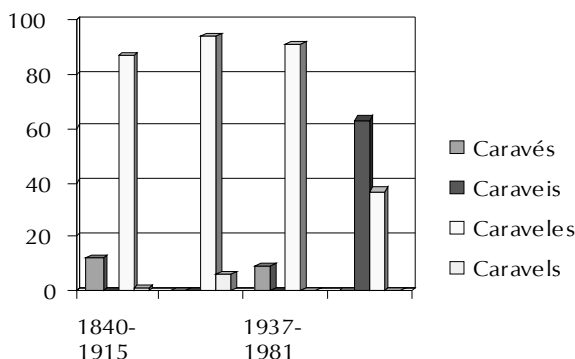


Figura 2. Plurais do substantivo *caravel* no galego escrito entre 1840 e 2002.

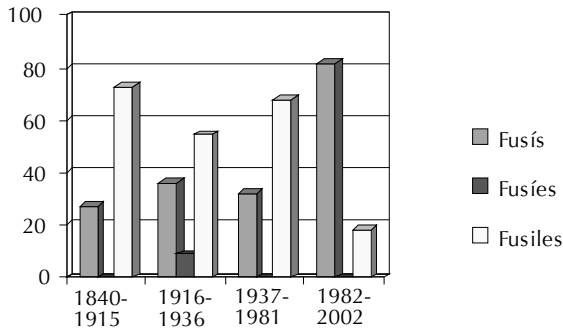


3.1. *-s, con supresión do -l final do singular*

É unha variante con desigual nivel de uso antes de 1982: moi
alto no caso das coñecidas “excepcións” *reás* e *cadriés* (FIGURAS 6

e 8) e algo menos alto no de *papés* (FIGURA 7); notable cos plurais de *animal* e *fusil* (FIGURAS 1 e 3); baixo cos de *caravel* e *azul* (FIGURAS 2 e 5); exiguo e en certos períodos nulo co de *español* (FIGURA 4).

Figura 3. Plurais do substantivo *fusil* no galego escrito entre 1840 e 2002.

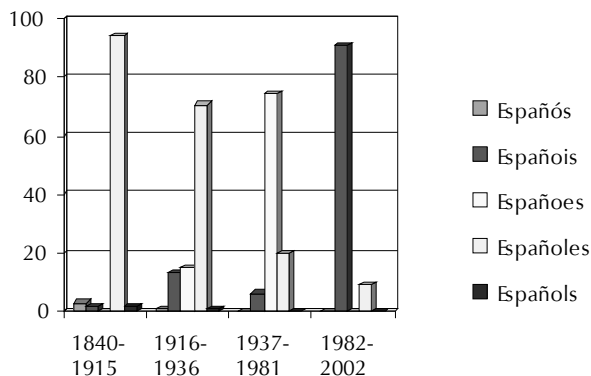


A meu ver, os datos tomados das fontes escritas concordan cos extraídos das fontes metalingüísticas no tocante ás tendencias observables no desenvolvemento desta opción e á identificación dalgunhas das unidades lexicais que máis resistencia opuxeron ao seu abandono. O esforzo purista por recuperar ou manter esta variante entre 1916 e 1981 apréciase ben nos casos de *animás*, *fusís* ou *azús*, aos que se lles podería engadir, por exemplo, *currás*, que atinxe unha representación do 30% entre 1916 e 1936 (3 rexistros de 10) e de nada menos ca o 58,8% entre 1937 e 1981 (10 rexistros de 17). Despois de 1982 o emprego desta variante só se mantivo con vigor cos nomes terminados en *-il* no singular (cos que é de regra no estándar) e mais coa reliquia *reás*, sobre a que as *Normas* da RAG e o ILG se pronunciaron dunha maneira que, a meu parecer, debeu de inducir a moitos a pensar que se trataba dunha excepción, se non promovida, si polo menos tolerada ou non mal vista no canto da normativa *reais*: “A solución *animás* debería ocupa-lo resto do territorio, pero só se oe en palabras illadas (como *reás* ou *cadrés*)” (RAG & ILG 1982: 52)⁴. A sentenza normativa a prol da variante *animais*,

⁴ Con certeza, con esta excepción resultou decisivo o feito de que na gramática de Álvarez, Regueira e Monteagudo (1986: 67) se dixese taxativamente que “o plural

caraveis, etc. significou o rápido eclipse de *animás*, *caravés*, etc. no galego escrito e/ou publicado despois de 1982.

Figura 4. Plurais do substantivo e adxectivo *español* no galego escrito entre 1840 e 2002.



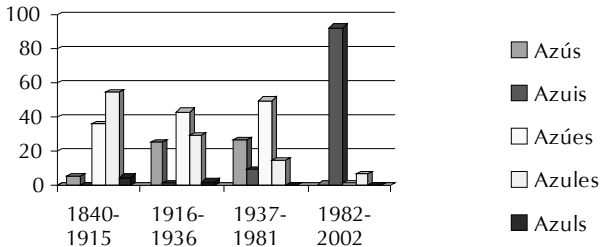
3.2. *-is, con supresión do -l final do singular*

En xeral, esta variante utilizouse pouco entre 1916 e 1982 e moi pouco antes de 1916. Ademais, os seus contados rexistros de entre 1840 e 1915 adoitan estar condicionados pola orixe oriental dos autores das obras en que se atopan: *españois*, *reais* e *tais* lense nos ensaios poéticos do berciano Fernández y Morales, do ano 1861, e *papeis* nunha entrega de 1913 da publicación fonsagrada *O Tio Pepe*. O ánimo purista que inspirou o cultivo do galego de non poucos autores despois de 1916 abonda para explicar o moderado progreso do uso escrito desta variante entre ese ano e 1981, progreso que se revela máis ostensible nos casos de *animais*, *españois* e *papeis* e algo menos no de *azuis*. Despois da Guerra Civil o seu gran salto adiante non empezou a producirse ata despois de 1971, é dicir, ata despois de que a RAG, na súas normas ortográficas e morfolóxicas

de *real* é *reás*" e, xa máis tarde, que Freixeiro Mato (2000: 49) lle dispensase certa acollida na súa obra ao afirmar que "o substantivo *real* mantén a solución patrimonial centro-occidental *reás* no plural na fala e na lingua literaria, en convivencia coa oriental e común *reais*".

dese ano, emitise unha prescrición favorable a ela. Despois de 1982 o esperable avance desta variante verificouse rapidamente.

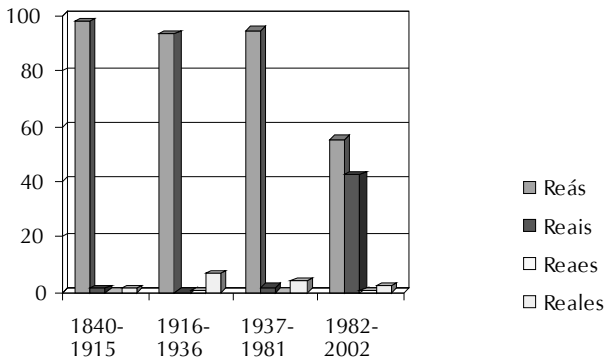
Figura 5. Plurais do adxectivo *azul* no galego escrito entre 1840 e 2002.



3.3. *-es, con supresión do -l final do singular*

É unha variante minoritaria que, con todo, antes de 1982 acadou unha discreta frecuencia de uso nos casos de *animaes*, *azúes* e *españoes*. Esta variante *españoes*, que antes da Guerra Civil foi unha solución da que gustaron algúns escritores, aparece como forma maioritaria no período posterior (cunha cota do 74,4%) só porque foi utilizada ata en 217 ocasións polo nacionalista Castelao no seu *Sempre en Galiza* (1944). Despois de 1982 os niveis de emprego desta variante do plural son insignificantes.

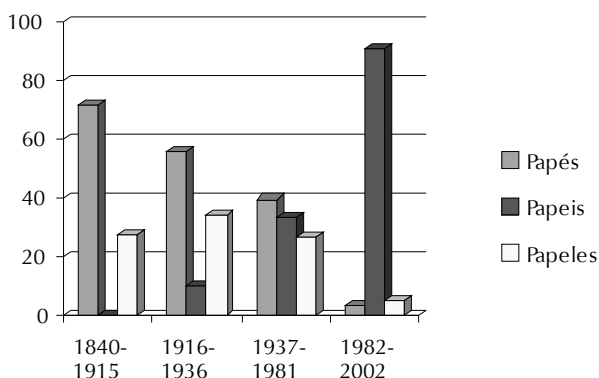
Figura 6. Plurais do substantivo *real* 'unidade monetaria' no galego escrito entre 1840 e 2002.



3.4. *–es, sen supresión do –l final do singular*

Antes de 1982 gozou dun elevado nivel de uso, nomeadamente nos casos de *caraveles*, *fusiles* e *españoles*⁵. Aínda que tamén importante, foi sensiblemente menos frecuente a súa selección coas unidades *animales*, *azules* e *papeles*. Finalmente, as súas porcentaxes de uso resultan moi baixas ou mesmo insignificantes nos casos dos substantivos *reales* e *cadriles*. Podemos, por tanto, reiterar aquí a idea de que esta variante converxente co castelán xa gozaba dunha notable difusión na segunda metade do século XIX e repetir a observación de que, debido a causas diversas ás que xa se aludiu, ao seu avance posterior opuxeron máis resistencia unhas unidades lexicais ca outras. No moderno galego escrito transparentáronse estas diferenzas entre os ritmos de difusión desta innovación ás distintas palabras susceptibles de adoptárense. Despois de 1982 o seu uso decaeu notoriamente.

Figura 7. Plurais do substantivo *papel* no galego escrito entre 1840 e 2002.

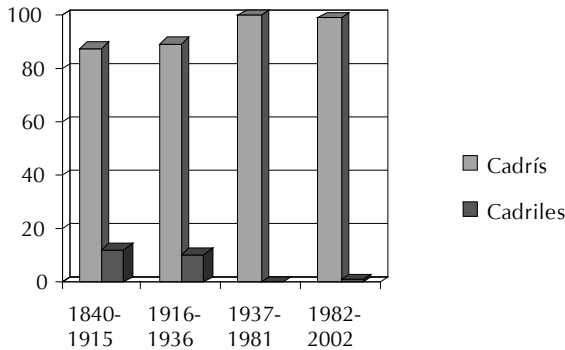


⁵ Intimamente vinculada coa ideoloxía segregada durante o proceso de construción do moderno estado español, a variante *españoles* utilizouse con contudente maioría entre 1840 e 1915 (no 94,3% dos casos rexistrados) e algo menos entre 1916 e 1936 (co 70,8%), cando comezou a ceder, sobre todo, ante dúas alternativas promovidas polos autores con actitudes máis puristas: *españóis* (13%) e *españoes* (14,6%). No período posterior chega incluso a facerse minoritaria (co 19,6%) debido ao sistemático e frecuentísimo uso que de *españoes* fixo Castelao no *Sempre en Galiza* (1944).

3.5. *-s, sen supresión do -l final do singular*

Na nosa moderna tradición escrita esta variante só atinxiu as residuais porcentaxes que cabía esperar da súa condición de solución moi minoritaria na lingua popular coetánea.

Figura 8. Plurais do substantivo *cadril* no galego escrito entre 1840 e 2002.



4. Conclusións

Polo que se refire á lingua da proximidade entre emisor e receptor, os datos tomados dos distintos tipos de fontes exploradas permítenos afirmar que a variante tradicional *animás* aínda debía de gozar dun notable nivel de uso na Galicia central e occidental a mediados do século XIX e que posteriormente, conforme foi aumentando o número de individuos con condutas bilingües galego-castelán, experimentou un rápido retroceso en beneficio de *animales* que a abocou á marxinalidade en que a presentan sumida os traballos de investigación realizados na segunda metade do XX. Como adoita suceder en toda mudanza lingüística, a innovación *animales* avanzou máis velozmente cunhas palabras ca con outras, e por distintas razóns algunhas unidades opuxéronlle particular resistencia (*reás, cadrís, papés...*). A variante *animais* conservouse ata hoxe como produtiva en moitos falares populares do centro e o sur do leste do territorio galegofalante. A meu ver, a súa selección como solución propia do estándar socializado a partir da década de 1980 está a propiciar a súa incorporación ao discurso oral

informal de persoas que tiveron como primeira lingua adquirida o español mais que por distintos motivos utilizan con maior ou menor frecuencia o galego, e tamén ao de individuos que, aínda que adquiriesen na infancia algunha variedade popular do galego con plurais terminados en *-les*, van introducindo solucións propias do estándar nas súas interaccións lingüísticas cotiás. Entre os aínda minoritarios grupos de persoas que tiveron como lingua inicial o galego estándar ou unha variedade próxima a el, *animais* é a solución empregada habitualmente tanto nos contextos formais coma nos informais. Pola súa parte, *animaes* e *animals* son variantes minoritarias que desde mediados do século XIX ata hoxe deberon de se manter con poucas alteracións nas contadas falas populares en que se utilizan, orientais no caso de *animaes* e occidentais no de *animals*.

No tocante á lingua da distancia, convén distinguir entre o uso escrito e o uso oral. As variantes do tipo de *animás* gozaron dun considerable nivel de uso na literatura decimonónica e, con desigualdades entre unhas palabras e outras, beneficiáronse do aprecio que non poucos escritores amosaron por elas entre 1916 e 1981. Foron ademais as que o Seminario de Estudos Galegos prescribiu en 1933 e aínda as que a Real Academia Galega utilizou na redacción do seu opúsculo ortográfico de 1970 e as que Carballo Calero apoiou ata a cuarta edición da súa gramática, de 1974. Non obstante, a decantación académica a prol da variante *animais* en 1971 acabaría por revelarse como unha reorientación destinada a perdurar, pois, logo de que os métodos de aprendizaxe do Instituto da Lingua Galega (favorables a *animales*) a contradixesen durante uns anos e despois de que as *Bases* de 1977 aínda lle recoñecesen *animales* como alternativa tolerable, foi consagrada polas *Normas* da RAG e o ILG de 1982. O que prevaleceu á hora de optar por esta prescripción foi o espírito depurador, descastelanizador e harmonizador co portugués que en maior ou menor medida xa se viñera revelando na tradición escrita anterior. Con escasas excepcións, as variantes do tipo de *animaes* e de *animals* foron pouco usadas no galego escrito moderno, especialmente as segundas.

Carecemos de fontes de estudo que nos permitan facer un seguimento da liña evolutiva da variable morfolóxica aquí estudada no galego utilizado oralmente en contextos formais desde mediados do

século XIX ata hoxe. Así e todo, paréceme sensato supoñer que esa liña evolutiva debeu de coñecer un trazado non moi distanciada do que acabo de sinalar para a lingua escrita. Antes de 1982, e quizais sobre todo entre os oradores máis comprometidos coa tarefa de descastelanizar o galego, debeu de predominar a preferencia polas variantes tradicionais sen mantemento de /l/ (*animás, animais, animaes*). Despois de 1982 os oradores máis esixentes e mellor preparados adoptan nos seus discursos a variante estándar *animais*, de modo que o uso de *animales*, sobre todo se é reiterado, considérase un sinal de descoido e de baixa competencia lingüística; as variantes *animás, animaes* e *animals* non se utilizan hoxe neste rexistro. Por outra banda, nos últimos tempos a variante *animais* comezou a competir con *animales* tamén na actividade lingüística de individuos con estudos e fasquía urbana que, situados en contextos de semiformalidade, intentan aproximarse no posible á variedade estándar.

A meu parecer, no caso do galego contemporáneo as innovacións lingüísticas que supoñen unha converxencia co castelán ou un achegamento a el tiveron na lingua da proximidade un ritmo de difusión máis rápido do que tiveron na lingua da distancia (oral ou escrita). Na lingua da proximidade o aumento da presión exercida do século XIX para acó polo idioma dominante encontrou poucas resistencias ao seu avance, cousa que semella lóxica se pensamos na indefensión en que ata hai moi pouco se achou o galego, percibido por moitos como un simple dialecto ou variedade baixa do castelán. Polo contrario, sobre a lingua da distancia actuou con rigor crecente unha reacción purista inserida no marco dun proceso de reemerxencia e elaboración lingüísticas contrario á dialectalización do galego. Logo da elaboración da variedade estándar e do inicio da súa socialización a través de institucións e organizacións públicas e privadas, as formas propias da lingua da distancia parecen ir infiltrándose en certas variedades da lingua da proximidade.

Podemos aseverar tamén que, aínda que o purismo se revelou como un factor decisivo no proceso de estandarización do galego moderno, a súa incidencia sobre a lingua escrita e sobre os crecentes espazos de uso oral que a esta lingua se lle foron abrindo no último século e medio da súa historia arrinca polo menos a mediados do século XIX, se ben se lle poden sinalar precursores coma o Padre Sarmiento (1695-1772).

Concordo, por tanto, con Brincat, Boeder e Stolz (2003: viii) cando aseguran que non existe unha interdependencia entre a emerxencia de tendencias puristas nunha comunidade lingüística e a creación planificada dun estándar no seo dela, é dicir, que o purismo non nace sempre e necesariamente como un atributo dos procesos de elaboración consciente dun estándar. Creo, con todo, que, aínda que o purismo non estea necesariamente conectado cos procesos de elaboración consciente de variedades estándar, si é polo menos un comportamento que tende a asociarse co nacionalismo cultural e político en sentido amplo (Thomas 1991). No caso galego pódese correlacionar o progresivo aumento da intensidade do purismo lingüístico desde mediados do século XIX ata hoxe co paralelo desenvolvemento cultural e político da consciencia da identidade galega durante esta época.

Finalmente, coido que, se é certo que o purismo exacerbado obstaculizou a revitalización de linguas ameazadas coma o tiwi australiano, o náhuatl ou o gaélico escocés (Dorian 1994), no caso do galego moderno está a prevalecer un purismo moderado que, sen abrir abismos entre a lingua da proximidade e a lingua da distancia, estase revelando como un factor capaz de expandir variantes diatopicamente pouco estendidas mais adornadas co atractivo de representaren unha tradición libre de toda sospeita de influencia castelanizante. Con certeza, a proximidade entre as estruturas do galego e do castelán esixe neste caso unha severa pero prudente actitude de vixilancia para impedir que o galego se desnaturalice por completo como consecuencia dun corremento cara ao castelán que na conciencia metalingüística predominante acabaría por convertelo nunha simple variante deturpada e empobrecida do idioma español.

Fontes exploradas

Gramáticas sincrónicas

- Álvarez, R.; Regueira, X. L.; Monteagudo, H. 1986. *Gramática galega*. Vigo: Galaxia.
- Álvarez, R.; Xove, X. 2002. *Gramática da lingua galega*. Vigo: Galaxia.
- Carballo Calero, R. 1966. *Gramática elemental del gallego común*. Vigo: Galaxia [266 pp.; 2ª ed., 1968, 286 pp.; 3ª ed., 1970, 338 pp.; 4ª ed., 1974, 330 pp.; 7ª ed., 1979, 346 pp.].

- Carré Alvarellos, L. 1967. *Gramática gallega*. A Coruña: Imprenta Moret.
- Cornu, J. 1906. *Gramática de la Lengua Portuguesa*. Apéndice: Morfología del gallego moderno. *Boletín de la Real Academia Gallega*. 1: 7-9; 2: 28-30; 3: 52-54; 4: 75-77; 5: 99-101.
- Costa Casas, X. X.; González Refoxo, M. A.; Morán Fraga, C. C.; Rábade Castiñeira, X. C. 1988. *Nova gramática para a aprendizaxe da lingua*. A Coruña: Vía Láctea.
- Cuveiro, J. 1868. *El habla gallega. Observaciones y datos sobre su origen y vicisitudes*. Pontevedra: Imp. de José A. Antúnez y C^ª.
- Freixeiro Mato, X. R. 2000. *Gramática da lingua gallega. II. Morfosintaxe*. Vigo: A Nosa Terra.
- Lugrís Freire, M. 1931. *Gramática do idioma galego*. 2^a ed. A Coruña: Imprenta Moret. [[1^a ed.: 1922; Ed. Facsímil, Santiago de Compostela: Consello da Cultura Galega, 2006].
- Mirás, F. 1864. *Compendio de gramática gallega-castellana*. Santiago: Establecimiento tipográfico de Manuel Mirás. [Ed. facsímil: Madrid: Akal, 1978].
- R. A. [atribuído a L. Carré Alvarellos] 1919. *Compendio de gramática galega*. A Coruña: Imprenta Nova.
- Saco Arce, J. A. 1868. *Gramática gallega*. Lugo: Soto Freire. [2^a ed.: Ourense: Gráficas Tanco, 1967].
- Santiago y Gómez, J. de 1918. *Filología de la lengua gallega*. Santiago de Compostela: Tip. de El Eco Franciscano.
- Valladares, M. 1970. *Elementos de gramática gallega*. Vigo: Galaxia. [Ms.: 1892].

Gramáticas diacrónicas

- Couceiro Freijomil, A. (s.d.). El idioma gallego. In: F. Carreras y Candi (Dir.). *Geografía general del Reino de Galicia*. Barcelona: Casa Editorial Alberto Martín, I, 161-327.
- Couceiro Freijomil, A. 1935. *El idioma gallego: historia, gramática, literatura*. Barcelona: Alberto Martín.
- García de Diego, V. 1909. *Elementos de gramática histórica gallega (fonética-morfología)*. Burgos: Hijos de Santiago Rodríguez. [Ed. facsímil: Anexo 23 de *Verba (Anuario Galego de Filoloxía)*. Universidade de Santiago de Compostela, 1984].
- Ferreiro, M. 1995. *Gramática histórica galega*. Santiago de Compostela: Laivento.

Propostas normativas

- Asociación Sócio-Pedagóxica Galega 1980. *Orientacións para a escrita do noso idioma*. A Coruña: Xistral.
- Associação Galega da Língua 1983. *Estudo crítico das Normas ortográficas e morfolóxicas do idioma galego*. A Coruña: AGAL. [2^a ed. corrigida e acrescentada, A Coruña: AGAL, 1989].

- Bases prá unificación das normas lingüísticas do galego (= Bases)*. Madrid: Anaya, 1977.
- Comisión de Lingüística – Xunta de Galicia 1980. *Normas ortográficas do idioma galego*. Santiago de Compostela: Xunta de Galicia. [Utilícei a seguinte edición: Xoán Carlos Rábade 1980. *Normas ortográficas do idioma galego. Texto íntegro. Guía e comentario*. A Coruña: La Voz de Galicia].
- Instituto de la Lengua Gallega 1971. *Gallego 1*. Universidade de Santiago de Compostela.
- Instituto de la Lengua Gallega 1972. *Gallego 2*. Universidade de Santiago de Compostela.
- Real Academia Galega; Instituto da Lingua Galega 1982. *Normas ortográficas e morfolóxicas do idioma galego*. Vigo: RAG / ILG. [13ª ed. revisada, 1995; 18ª ed. revisada: 2003].
- Real Academia Gallega 1970. *Normas ortográficas do idioma galego*. A Coruña: RAG.
- Real Academia Gallega 1971. *Normas ortográficas e morfolóxicas do idioma galego*. A Coruña: RAG.
- Seminario de Estudos Galegos 1933. *Algunhas normas pra a unificazón do idioma galego*. Santiago de Compostela: Nós.

Traballos de investigación

- ALGa 2: Instituto da Lingua Galega 1995. *Atlas lingüístico galego*. Vol. 2: *Morfoloxía non verbal*. A Coruña: Fundación Pedro Barrié de la Maza, Conde de Fenosa.
- Dubert García, F. 1999. *Aspectos do galego de Santiago de Compostela*. Anexo 44 de *Verba (Anuario Galego de Filoloxía)*. Universidade de Santiago de Compostela.
- Dubert García, F. 2002. Os sociolectos galegos. *Cadernos de Lingua*. **24**: 5-27.
- Fernández Rei, F. 1990. *Dialectoloxía da lingua galega*. Vigo: Edicións Xerais de Galicia.
- Freixeiro Mato, X. R.; Sánchez Rei, X. M.; Sanmartín Rei, G. 2005. *A lingua literaria galega no século XIX*. Universidade da Coruña.
- Mariño Paz, R. 2003. *O idioma galego no limiar da súa renacementa. Estudo lingüístico de textos pregaleguistas*. Monografía 2 da *Revista Galega de Filoloxía*. Universidade da Coruña.
- Regueira, X. L. 1999. Estándar oral e variación social da lingua galega. In: R. Álvarez; D. Vilavedra (Eds.). *Cinguidos por unha arela común. Homenaxe ó profesor Xesús Alonso Montero*. Universidade de Santiago de Compostela, I, 855-875.
- Regueira, X. L. 2003. Lingua falada e estándar escrito: O galego desde o Rexurdimento á modificación ortográfica de 2003. *Galicia dende Salamanca*. **4**: 185-220.

- Regueira, X. L. 2005. Estándar oral. In: R. Álvarez; H. Monteagudo (Eds.). *Norma lingüística e variación*. Santiago de Compostela: Instituto da Lingua Galega / Consello da Cultura Galega, 69-95.
- Sánchez Rei, X. M. 2006. *A lingua galega no cancionero de Pérez Ballesteros*. Santiago de Compostela: Laiovento.

Tesouros informatizados

- TILG: A. Santamarina (Dir.) 2003. *Tesouro Informatizado da Lingua Galega*. Instituto da Lingua Galega – Universidade de Santiago de Compostela (<http://www.ti.usc.es/TILG/>).

Referencias

- Brea, M. 1994. Galegisch: Grammatikographie und Lexikographie/Gramaticografía y lexicografía. In: G. Holtus; M. Metzeltin; Ch. Schmitt (Eds.). *Lexikon der Romanistischen Linguistik*. Vol. 6, 2: *Galegisch, Portugiesisch*. Tübingen: Max Niemeyer, 110-129.
- Brincat, J.; Boeder, W.; Stolz, Th. 2003. Preface. In: J. Brincat; W. Boeder; Th. Stolz (Eds.). *Purism in minor languages, endangered languages, regional languages, mixed languages. Papers from the conference on 'Purism in the Age of Globalisation' (Bremen, September 2001)*. Bochum: Universitätsverlag Dr. N. Brockmeyer, vii-xiv.
- Dorian, N. C. 1994. Purism vs. compromise in language revitalization and language revival. *Language in Society*. **23(4)**: 479-494.
- Freixeiro Mato, X. R. 1998. *Gramática da lingua galega. I. Fonética e fonoloxía*. Vigo: A Nosa Terra.
- García, C. 1971. Orixen e problemas do método de galego. *Grial*. (32): 129-141.
- Mariño Paz, R. 2004. Diacronía da Formación do Plural dos Nomes Oxítonos Galegos Terminados en // no Singular. In: A. M. Brito; O. Figueiredo; C. Barros (Orgs.). *Linguística Histórica e História da Língua Portuguesa. Actas do Encontro de Homenagem a Maria Helena Paiva (Faculdade de Letras da Universidade do Porto, 5-6 de novembro de 2003)*. Porto: Faculdade de Letras da Universidade do Porto, 169-195.
- McMahon, A. M. S. 1994. *Understanding Language Change*. Cambridge: Cambridge University Press.
- Sánchez Rei, X. M. 2005. Trazos xerais da tradición gramatical galega. *Revista Galega de Filoloxía*. **6**: 93-121.
- Santamarina, A. 1974. Gramática. In: *Gran Enciclopedia Gallega*. Xixón: Silverio Cañada, XVI, 191-195.
- Santamarina, A. 1994. Norma e estándar. In: G. Holtus; M. Metzeltin; Ch. Schmitt (Eds.). *Lexikon der Romanistischen Linguistik*. Vol. 6, 2: *Galegisch, Portugiesisch*. Tübingen: Max Niemeyer, 66-79.
- Thomas, G. 1991. *Linguistic purism*. London/New York: Longman.

Mudança e variação na terminologia açucareira actual do Brasil¹

Naidea Nunes Nunes
naidea@uma.pt
Universidade da Madeira (Portugal)

RESUMO. Este estudo comparativo pretende compreender a mudança e a variação na actual terminologia açucareira em diferentes estados brasileiros (Paraíba, Pernambuco, Baía, Minas Gerais, Santa Catarina and Rio Grande do Sul). O estudo, além da variação terminológica, permite-nos observar alguma variação histórica, geográfica e social. Podemos encontrar termos antigos e recentes, respectivamente em regiões conservadoras e inovadoras, explicados por factores históricos e geográficos. Simultaneamente, regista-se o uso de palavras populares e termos técnicos (variação social). Porém, a principal variação é a terminológica: formas sinónimas e polissémicas que nos mostram uma mudança em curso, como ocorre na língua comum.

PALAVRAS-CHAVE. Brasil, cana de açúcar, rapadura, cachaça, lexicologia, terminologia, dialectologia.

ABSTRACT. This comparative study intends to understand the change and variation in the actual terminology of sugar cane in different Brazilian States (Paraíba, Pernambuco, Baía, Minas Gerais, Santa Catarina and Rio Grande do Sul). This study, besides the terminological variation, allows us to observe some historic, geographic and social variation. We can find old and new terms, respectively in conservative and innovative regions, explained by historic and geographic factors. At the same time, there is the use of popular words and technical terms (social variation). But the main variation is the terminological one: synonymous and polysemic forms that show us a change in course, like it happens in common language.

KEY-WORDS. Brazil, sugar cane, rapadura, cachaça, lexicology, terminology, dialectology.

O estudo da terminologia açucareira histórica e actual do Mediterrâneo (incluindo a Sicília, Valência e Granada) ao Atlântico, iniciado

¹ Agradecemos ao Centro de Investigação e Tecnologia da Madeira (CITMA) a bolsa de pós-doutoramento, atribuída ao projecto de investigação «A terminologia açucareira no Atlântico: património linguístico-cultural madeirense», sem a qual este trabalho não teria sido realizado.

com uma tese de doutoramento na área da Linguística Românica, permite-nos contribuir para a compreensão da mudança e variação do léxico nas línguas românicas, pois, no âmbito da terminologia açucareira, registámos os mesmos processos de evolução lexical da língua corrente.

Dando continuidade ao estudo da terminologia açucareira actual, depois de ter realizado inquéritos linguístico-etnográficos sobre a produção açucareira, na ilha da Madeira e em Cabo Verde, avançámos para Canárias, S. Tomé e Príncipe e Brasil, procurando conhecer e compreender a mudança histórica e a variação denominativa existentes nas diferentes áreas geográficas estudadas, tendo em conta as condições sociolinguísticas e o contacto entre línguas, principalmente no Brasil, entre a língua portuguesa, as línguas indígenas (sobretudo as da família tupi-guarani) e a influência das línguas africanas.

Por razões de necessidade de limitação da comunicação e de aprofundamento da questão da mudança e variação, centraremos o nosso estudo na descrição e interpretação dos materiais linguísticos da terminologia açucareira actual do Brasil (Estados do Ceará, Paraíba, Pernambuco, Baía, Minas Gerais, Santa Catarina e Rio Grande do Sul), recolhidos no trabalho de campo realizado em Agosto e Setembro de 2006. Pretendemos, assim, partindo da primitiva terminologia açucareira madeirense, transplantada para o Brasil no início da sua colonização, contribuir para a compreensão dos mecanismos de mudança linguística, mais precisamente de mudança lexical, tendo em conta a variação histórica, mas também a variação geográfica e a variação social.

Weinreich, Labov e Herzog, em *Empirical Foundations for a Theory of Language Change*, publicado em 1968, defendem um modelo que busca descrever a heterogeneidade ordenada dentro da língua. A referida obra foi pioneira na perspectiva desenvolvida posteriormente, nomeadamente por William Labov, em muitas outras publicações. Como já referimos, nesta perspectiva, a mudança e variação linguísticas constituem a heterogeneidade ordenada da língua como fenómeno histórico, geográfico e social. Assim, pode ocorrer a coexistência de formas conservadoras e inovadoras num mesmo falante e numa mesma área geográfica. Os mesmos autores entendem a heterogeneidade ou variação integrada na competência linguística do falante e defendem a existência de forças motivadoras

na mudança linguística que podem passar por cima de qualquer tendência de preservar a função distintiva das unidades lexicais, à medida que haja elementos contextuais compensadores. Este fenómeno de variação e mudança linguística acontece tanto na linguagem corrente como nas linguagens de especialidade.

A terminologia tradicional de Eugen Wuster via os termos como unidades precisas e unívocas, que não aceitavam sinonímia nem polissemia. Com a revisão da Teoria Geral da Terminologia (TGT), realizada por Teresa Cabré (1993), na nova Teoria Comunicativa da Terminologia (TCT), as terminologias técnicas e científicas deixam de ser vistas como homogéneas e passam a ser vistas como realidades linguísticas heterogéneas, havendo variação terminológica, quanto ao grau de especialização: comunicação especializada (de especialista para especialista), com termos mais técnicos e maior densidade terminológica; comunicação de especialista para não especialista, em que há menor grau de especialização. Assim, o contexto comunicacional determina o grau de especialização da linguagem de especialidade.

1. Metodologia

O presente estudo da mudança e variação linguísticas na terminologia açucareira actual do Brasil resulta da aplicação de um questionário lexical, onomasiológico e terminológico, sobre a produção açucareira, estruturado por campos semânticos: cultivo e colheita da cana-de-açúcar, extracção do sumo da cana, fabrico do mel de cana, fabrico de rapadura e alfenim, fabrico de açúcar de cana e fabrico de aguardente de cana. A realização dos inquéritos terminológicos ocorreu nos Estados brasileiros com a mais antiga (tradicional) e actual produção açucareira artesanal e familiar, respectivamente na Baía, Pernambuco, Paraíba, Ceará, Minas Gerais, Santa Catarina e Rio Grande do Sul, excluindo a produção industrial das grandes fábricas e usinas do Rio de Janeiro e de S. Paulo, uma vez que estas, devido à sua grande industrialização, não apresentam interesse linguístico-terminológico, pois todos os processos de fabrico são mecanizados e os trabalhadores são simples operários de máquinas.

Para a selecção das localidades, dentro de cada Estado, num imenso território como é o Brasil, procedeu-se à procura dos produtos açucarei-

ros, nos mercados de produtos regionais, na cidade capital do Estado, com vista à recolha de informações sobre os engenhos e as localidades em que há produção açucareira artesanal e ao estabelecimento de contactos para deslocação, transporte e alojamento. Seguiu-se a deslocação às localidades no interior dos Estados e a visita aos engenhos para a realização de entrevistas ao dono ou, sempre que possível, ao trabalhador responsável por toda a produção do engenho.

Durante a visita ao engenho, depois da apresentação e explicação do objectivo da entrevista, a realização de um estudo linguístico-terminológico, solicitámos a disponibilidade de um informante que nos pudesse explicar todo o processo de produção açucareira, desde o cultivo da cana até a comercialização dos produtos resultantes da transformação da cana no engenho. Depois de encontrado o informante especialista sobre o tema, que, geralmente, como se trata de uma produção artesanal e familiar, domina tanto a parte agrícola como a parte industrial, certificámo-nos de que o informante é nativo da região, tendo vivido sempre na mesma, para que seja representativo da terminologia utilizada na área geográfica em estudo. Depois dos informantes concordarem em ser entrevistados, no âmbito do nosso estudo comparativo terminológico, realizamos um inquérito por informante, ou seja, aplicamos o questionário lexical previamente elaborado, como já referimos. As entrevistas têm a duração aproximada de 45 a 60 minutos.

Dada a especificidade do tema, os inquéritos foram efectuados em meios rurais, a trabalhadores de engenhos, especialistas da actividade açucareira. Assim, quanto ao tipo de informante, predominam os indivíduos com baixa escolarização ou nível de instrução básico; idosos, com mais de 55 anos (embora tenhamos entrevistado um jovem de 30 anos, responsável pelo engenho de S. Pedro, em Triunfo, Pernambuco); do sexo masculino, visto que se trata de uma actividade, geralmente, desenvolvida por homens. No entanto, excepcionalmente, entrevistámos duas mulheres, uma na Paraíba, que trabalha com o marido na produção açucareira, e outra viúva, ocupando o lugar do marido na terra e no engenho, em Santa Catarina. Trata-se de informantes de um nível sócio-económico e cultural baixo, uma vez que são estes que se dedicam à produção açucareira artesanal e familiar, dominando os termos e as técnicas de uma actividade tradicional em vias de extinção, que interessa conhecer e salvaguardar.

Todas as entrevistas foram realizadas pela autora, procurando controlar a influência do entrevistador sobre o entrevistado, designadamente tendo o cuidado de não referir os termos a ser recolhidos. Procurou-se também, sempre que possível, que as entrevistas fossem realizadas na ausência de terceiros e sem interrupções, evitando interferências indesejadas e ruídos na realização da entrevista e respectiva gravação. As entrevistas foram realizadas no ambiente familiar e de trabalho do entrevistado, preferencialmente no engenho, junto da moenda, do forno ou do tacho e do alambique, facilitando a identificação terminológica dos conceitos questionados.

As gravações foram feitas em minidisc recordable Sony de 80 minutos, com um portable minidisc recorder Sony MZ-R909 e um microfone exterior Sony, colocado junto do entrevistado, permitindo acompanhar a movimentação do entrevistado durante a entrevista, nomeadamente da sala da moenda para a sala de cozimento e para a sala do alambique. Posteriormente, todas as entrevistas foram transcritas para possibilitar o estudo lexical/terminológico comparativo entre os diferentes Estados brasileiros.

Embora a fundamentação teórica e o objectivo deste trabalho não sejam de índole dialectológica nem sociolinguística, mas sim terminológica, visando o estudo comparativo da terminologia açucareira actual dos diferentes Estados brasileiros estudados, como já referimos, podemos inferir algumas questões de variação geográfica entre Estados e de variação social, associada a factores como a idade, o sexo e a escolaridade, pois é importante dispensarmos alguma atenção aos potenciais factores de variação terminológica intra e inter Estados ou regiões açucareiras, verificando a existência de variação geográfica e sócio-cultural.

2. Cultivo e colheita da cana-de-açúcar

No campo semântico do cultivo da cana-de-açúcar, registámos apenas a denominação *cana-de-açúcar* para designar a planta cultivada, nos diferentes Estados brasileiros estudados, enquanto, na ilha da Madeira, encontramos, a par da referida denominação, a designação mais popular *cana doce* e a mais técnica e erudita *cana sacarina*.

Relativamente às denominações das diferentes variedades de cana-de-açúcar, registámos alguns termos comuns ao Brasil e à Madeira,

nomeadamente *cana branca*, *cana roxa* e *cana caiana*. No Brasil, encontrámos novas denominações referentes a novas variedades e realidades locais, por exemplo: *cana pitua*, na Paraíba; *cana três x* e *cana paulista*, em Pernambuco; *cana Java* ou *jabra*, *cana da embrapa* e *cana de Alagoa*, em Minas Gerais; *cana paulista*, *napa paulista* ou *toiçada* também em Santa Catarina e *napa paulista*, *cana mulata pelada* e *cana Tucuman* (da Argentina), no Estado do Rio Grande do Sul.

No que se refere às operações de cultivo e colheita da cana, encontrámos as mesmas denominações nos dois lados do Atlântico (Madeira, Canárias, Cabo Verde, S. Tomé e Príncipe e Brasil), designadamente *plantar* com a variante popular *prantar a cana*, *limpar a cana* e *cortar* ou *colher a cana*. Também as denominações das diferentes partes da cana e diferentes fases do seu crescimento são idênticas, nomeadamente *pé da cana*, *raiz* ou *soca*; *semente*, *muda* ou *planta da cana*, para denominar o pedaço de cana semeado; *gomo da cana*, a parte mole da cana que se chupa; *nó da cana*, a parte dura das divisões da cana em gomos; *olho da cana*, que denomina simultaneamente o rebento do nó da cana e a parte de cima da cana; *pendão (da cana)* e *bandeira*, a flor da cana. A par destas denominações, ocorrem outras unidades lexicais, como a forma *olhadura*, para denominar a ponta da cana utilizada como semente; a denominação *palmito* (em Minas Gerais) e *olha* (em Santa Catarina), para designar as folhas da parte de cima da cana; os nomes *gema*, *broto* e *nacedor*, respectivamente em Santa Catarina e na Baía, designando o rebento do nó da cana.

Registámos as unidades terminológicas: *limpar a roça*², *roçar o mato*, *carpinar* ou *carpir o mato*, em Pernambuco; *capinar*³, *carpinar*

² Segundo António Geraldo da Cunha, o termo *roça* (s. XVI), forma derivada regressiva de *roçar*, provém do latim **ruptiare*, de *ruptus*, participio de *rumpere*, romper (CUNHA 1992, p. 687). Segundo o *Dicionário Houaiss da língua portuguesa*, a primeira atestação da palavra *roça* data de 1552, com o significado de acção ou efeito de roçar; terreno em que se faz a roçada; terreno com muito mato, mato crescido e ainda terreno de lavoura, grande ou pequeno, e plantação ou plantio (1594) e *roçar* data do século XIV, com o significado de cortar, derrubar mato (*Dicionário Houaiss da língua portuguesa*, 2005, tomo xvi, p. 7053).

³ Segundo informações do *Dicionário Novo Aurélio século XXI*, o termo *capinar* (de *capim* + *-ar*) é utilizado no Brasil, Angola, Cabo Verde, Guiné, Moçambique e S. Tomé e Príncipe, com o significado de limpar as plantas, uma plantação ou um terreno, apresentando entre outros o sinónimo *carpir* (*Dicionário Novo Aurélio*

ou *carpir o mato*, a par de *roçar o mato*, em Minas Gerais, e também *capinar*, como sinónimo de limpar as ervas da plantação de cana, em Santa Catarina e no Rio Grande do Sul, sendo que o termo *roçar* é a forma mais antiga e *capinar* com a variante *carpinar* e a forma sinónima *carpir* são termos mais recentes e que surgem no português africano e brasileiro. Uma plantação de cana-de-açúcar é designada *plantio* em Pernambuco; *roça de cana* na Baía; *canavial*, *roça* e *moita de cana* em Minas Gerais; *roça de cana* em Santa Catarina; *roça de cana*, *plantio (de cana)*, *canavial (de cana)* e *prantação de cana*, no Rio Grande do Sul. O termo *roça (de cana)* denomina uma pequena plantação de cana, a par do termo *canavial*, grande plantação de cana, em Minas Gerais e no Rio Grande do Sul.

Salientamos ainda a conservação dos termos: *cortador de cana*; *olho da cana*, *soca da cana* e *fi(lh)o*, *fi(li)ação* e *fi(lh)ozinho de cana* (denominações do rebento da cana cortada), que, na Madeira, nas Canárias, em Cabo Verde e em S. Tomé e Príncipe também se denominam *filho* ou *herdeiro*; *pranta* e *cana nova* (a cana do primeiro corte); *soca* e *ressoca* ou *soqueira*, respectivamente as canas do segundo e terceiro corte; *feixe de cana* ou *molho de cana* (conjunto de canas reunidas para serem transportadas para o engenho), a par das palavras *embolar* (Pernambuco) e *bandeirar* e *bandeira* (Minas Gerais), respectivamente para denominar o amarrar da cana e o molho de canas.

3. Extracção do sumo da cana

As instalações de transformação artesanal ou semi-industrial da cana-de-açúcar são denominadas *engenho*, em todos os Estados brasileiros, tal como na ilha da Madeira, enquanto nas Canárias (nomeadamente na ilha de La Palma), em Cabo Verde e em S. Tomé e Príncipe, encontrámos ainda o termo mais antigo *trapiche*, que, no

século XXI, 1999, p. 398). O termo *capinar* surge datado de 1871, no *Dicionário Etimológico Nova Fronteira da Língua Portuguesa* (CUNHA, 1992, p. 150), enquanto, no *Dicionário Houaiss da língua portuguesa*, é datado de 1836, com o significado de retirar capim, usando enxada ou afim, limpar terreno de mato ou ervas daninhas (*Dicionário Houaiss da língua portuguesa*, tomo v, p. 1749).

Brasil, adquiriu um novo significado, armazém e porto de embarque de mercadorias (primitivamente associado à exportação de açúcar para a Europa).

O moinho de extracção do sumo da cana, que antigamente funcionava a tracção animal (*almanjarra* ou *manjarra de besta e engenho a boi*) e manual (*engenhoca*, *arrebenta-peito*, *moenda de pau* ou *engenho de pau*), meio artesanal que ainda hoje encontramos em Cabo Verde e em S. Tomé e Príncipe, praticamente desapareceu em todos os Estados brasileiros, tendo sido substituído por pequenos moinhos ou engenhos a motor diesel e eléctricos. Em Minas Gerais, encontrámos ainda um *engenho de água*, com *roda de água*, *roda de madeira* ou *cubos*, que recebem a água, fazendo movimentar a moenda, e alguns engenhos de bolandeira e pau de almanjarra, movidos por animais.

A época de corte da cana e laboração do engenho denomina-se *safra* ou *tempo de moagem*, em praticamente todos os Estados brasileiros, tal como na ilha da Madeira, nas Canárias, em Cabo Verde e em S. Tomé e Príncipe, sendo que a época de plantio e crescimento da cana, em que pode haver produção de cachaça, a partir do concentrado ou melado de cana, é designada *entre-safra*.

A máquina de moer a cana é denominada *moenda* (*do engenho*) em todos os Estados brasileiros estudados, sendo que, em Pernambuco, apresenta a variação denominativa *moagem*; em Minas Gerais, *máquina de moer* e, no Rio Grande do Sul, *engenho de moer*. Os três cilindros de madeira (antigos) ou de ferro (actuais), na posição vertical (os mais antigos encontrados em Minas Gerais e em Santa Catarina) e horizontal (os mais modernos em triângulo) também são denominados *moenda(s)*, em todos os Estados brasileiros estudados, apresentando as seguintes variações denominativas: *terno de moenda*, na Paraíba; *tambores* ou *eixos*, no Ceará; *tambores da cana* e *rolos* em Pernambuco; *moenda de ferro* e *moenda da cana*, em Minas Gerais; e *moenda de ferro*, *peão de moenda* e *rolo(s) de ferro*, em Santa Catarina. Como podemos observar, estamos perante uma variação mais terminológica do que geográfica.

Registámos o termo *moer a cana*, *espremer* e *esmagar*, praticamente, em todos os Estados brasileiros estudados, tal como na Madeira, nas Canárias, em Cabo Verde e em S. Tomé e Príncipe, onde também ocorre

o termo *trapichar* (de *trapiche*), para denominar o mesmo conceito. Apenas encontrámos a unidade terminológica *capitar a cana* no Estado de Minas Gerais, onde ocorre também a denominação *capitadeira da cana*, designando o primeiro cilindro que esmaga a cana, sendo sinónimo de *moenda quebradeira*. O termo *moenda mestre* denomina o cilindro do meio, que movimenta os outros dois, e *espremedeira* é o termo que designa o cilindro que esmaga pela segunda vez a cana, sendo denominados respectivamente *moenda do meio* e *moenda de lateral*, no Rio Grande do Sul. Ainda no que diz respeito à moenda, em Santa Catarina, o termo *macaco*, como denominação da parte de trás da moenda antiga de três cilindros verticais, é uma particularidade desta moenda, que serve para empurrar a cana, dispensando a necessidade de um trabalhador na parte de trás da mesma, para voltar a introduzir a cana, o que acontece em Cabo Verde e em S. Tomé e Príncipe. O trabalhador que introduz a cana na moenda denomina-se *cevador de cana*, na Paraíba e na Baía; *tombador* e *moendeiro*, em Pernambuco; e *engenheiro* ou *moedor (de cana)* em Minas Gerais, sendo que esta última denominação é também o termo que ocorre em Santa Catarina e no Rio Grande do Sul.

O sumo da cana extraído na moenda é denominado *caldo de cana*, em todos os Estados brasileiros, sendo esta a designação mais popular, uma vez que *caldo* denomina vulgarmente o sumo de um fruto, como por exemplo *caldo de limão*, daí a necessidade do complemento determinativo *de cana*, em *caldo de cana*. Registámos o termo mais técnico *garapa* em Pernambuco, em Minas Gerais, em Santa Catarina e no Rio Grande do Sul, com a variante *guarapa* (possivelmente por influência da forma espanhola *guarapo*, devido à proximidade com o Uruguai e com a Argentina). Importa sublinhar que *garapa* é também o termo técnico utilizado, actualmente, na ilha da Madeira, para denominar o mesmo conceito, sendo controversa a origem deste termo. Encontrámos também a forma espanhola *guarapo* nas Canárias, onde também denomina, na ilha La Gomera, o mel de palma, parecendo ser um termo muito antigo, talvez de origem guanche. Como curiosidade, acrescentamos que, em Cabo Verde e em S. Tomé e Príncipe (que recebeu a terminologia açucareira cabo-verdiana), registámos apenas a forma crioula *calda*, para designar o sumo da cana-de-açúcar.

O recipiente que recebe o sumo da cana que sai da moenda é designado *parol*, termo muito antigo, do latim *pariolum*, com a forma *perol* no castelhano, sendo um catalanismo de uso rústico e local para designar um recipiente ou tina, em português *parol*⁴. No Brasil, actualmente, o termo *parol* ainda denomina o recipiente em que se recolhe o sumo da cana da moenda, no Ceará, em Pernambuco e na Baía, enquanto nos restantes Estados é denominado por nomes mais comuns: *depósito (de madeira)*, *cocho*, *tanque*, *tina de madeira* e *decantador* (termo mais técnico que ocorre em Santa Catarina). Os resíduos da cana esmagada são designados *bagaço (da cana)*, em todos os Estados brasileiros, tal como na Madeira, nas Canárias, em Cabo Verde e em S. Tomé e Príncipe. Apenas, na Baía, encontramos os termos sinónimos *tirador de bagaço* e *puxador do bagaço*, para denominar o trabalhador que retira o bagaço da moenda. Apesar da tendência conservadora da terminologia açucareira atlântica, registámos, no Brasil, a ocorrência do termo de origem africana *banguê*, que denomina a padiola de cipós trançados, na qual se leva o bagaço à *bagaceira*.

4. Fabrico do mel de cana

No fabrico do mel de cana, tal como no fabrico de rapadura, de alfenim e de açúcar de cana, o sumo da cana passa por um processo de cozimento, em recipientes denominados *tachas* e *tachos*, sendo que as *tachas* são maiores do que os *tachos*. Em Santa Catarina, registámos a denominação *tachado*, a par dos termos *tacho (de melado)*, *tacho de cobre* e *forno (de melado)*, para designar o mesmo conceito, sendo que estes últimos termos também ocorrem no Rio Grande do Sul. Verificámos que, no caso do último termo, *forno (de melado)*, por contiguidade, ocorre uma extensão do significado da palavra *forno*, sinónimo de *fornalha*, local onde se coloca o fogo, denominando também o tacho de cobre em que se coze o mel.

⁴ Cf. Naidea Nunes, *Palavras doces. Terminologia e tecnologia históricas e actuais da cultura açucareira: do Mediterrâneo ao Atlântico*, Governo Regional da Madeira, Secretaria Regional do Turismo e Cultura, Centro de Estudos de História do Atlântico, 2003, p.504.

O primeiro tacho que recebe o sumo da cana da moenda chama-se *recebedeira* na Paraíba; *recebedor* no Ceará; *tacha de caldo frio* em Pernambuco, e *caldo frio* ou *primeira tacha* na Baía, sendo que não registámos nenhuma denominação para este conceito nos restantes Estados brasileiros, acontecendo o mesmo com a denominação das restantes caldeiras, pois o cozimento do sumo da cana, muitas vezes, é reduzido apenas a duas tachas. A segunda tacha em que o sumo da cana ferve é denominada *caldeira*, na Paraíba, no Ceará e em Pernambuco, onde registámos ainda a denominação sinonímica *tacha caldo quente*, enquanto, na Baía, é designada *segunda tacha*. A terceira tacha, em que o sumo da cana é concentrado, denomina-se *caldeirote*, nos Estados de Paraíba, Ceará e Pernambuco, enquanto, na Baía, é denominada *terceira tacha*. A quarta tacha, em que se apura o caldo, chama-se *apurador*, na Paraíba; *cruzeta*, em Pernambuco, e *quarta tacha* na Baía. O último tacho que fica na boca da fomalha chama-se *boca* na Paraíba, *tacha da boca* em Pernambuco e *tacho do ponto* na Baía, sendo denominado *tacho* e *fomalha*, no Rio Grande do Sul. Depois de retirado do último tacho, com uma *passadeira* ou *cuia*, o mel passa para um recipiente para arrefecer, denominado *resfriadeira*, na Paraíba e Ceará, com a variante *esfriadeira* em Minas Gerais, onde ocorre a variação denominativa *cama fria* ou *fria* e *cocho (de madeira)*, enquanto, em Pernambuco, registámos o termo *desafogador*, para denominar o mesmo conceito. Muitos destes termos parecem constituir inovações lexicais brasileiras, uma vez que não ocorrem na documentação histórica nem actualmente nas outras regiões açucareiras do Atlântico.

O termo *ponto do mel* é comum às várias regiões açucareiras do Atlântico, apresentando as variantes *ponto do melado* em Minas Gerais, Santa Catarina e Rio Grande do Sul e o sinónimo *mel apurado*, na Paraíba. O produto resultante da concentração do sumo da cana é denominado: *mel (de cana)*, na Paraíba, Ceará, Pernambuco e Baía; *mel de engenho*, em Pernambuco e na Baía; *melado (de cana)*, na Paraíba, na Baía, em Minas Gerais, em Santa Catarina e no Rio Grande do Sul e ainda *melaço*, forma mais popular, em Pernambuco e na Baía (termo que também denomina o subproduto do açúcar). O trabalhador responsável pelo fabrico do mel é denominado *fornalheiro* e *mestre do cozimento*, na Paraíba; *mestre*, no Ceará; *caldeireiro*,

em Pernambuco; *cozinhador*, na Baía; *tacheiro*, em Minas Gerais; e *meladeiro* (forma derivada de *melado* com o sufixo *-eiro*), no Rio Grande do Sul.

5. Fabrico de rapadura e alfenim

O produto açucareiro feito a partir do mel, em moldes de madeira quadrados denomina-se *rapadura* em todos os Estados brasileiros. Na Baía, a par do nome *rapadura*, ocorrem as denominações *rapadurinha* (*rapadura* feita em moldes pequenos), *rapadura serenta* (de má qualidade), *rapadura tradicional*, com a variante *raspadura* (de *raspa*) e *rapadura puxa* (tipo de *rapadura* puxada, ficando um pouco mais clara), denominações que também ocorrem em Minas Gerais, em Santa Catarina e no Rio Grande do Sul, onde encontrámos a variante *rapadura de melado* e a denominação *quebra-queixo*, como sinónimo de *rapadura puxa*, a par da designação *puxa-puxa* (*rapadura* mais mole), denominação que também ocorre em Santa Catarina.

O nome *rapadura*, primitivamente, denominava os restos caramelizadas do mel que se rapavam do fundo do tacho e que serviam de alimento para os trabalhadores dos engenhos. Esta denominação passou da ilha da Madeira, onde ocorre a primeira atestação, na língua portuguesa, *rapaduras*, em 1523, para Canárias, onde ainda hoje existe na ilha de La Palma (doce de forma cónica feito de mel de cana com diferentes ingredientes), para Cabo Verde e S. Tomé e Príncipe (onde ainda conserva o seu significado original) e para o Brasil, onde a palavra adquiriu um novo significado, designando uma espécie de açúcar bruto, misturado com amendoim, coco, goiaba, tangerina, mamão, limão, abacaxi, gengibre, abóbora, etc.

O trabalhador responsável pelo fabrico de *rapadura* é denominado *mestre da rapadura*, na Paraíba e em Pernambuco; *mestre*, no Ceará; *cozinhador*, na Baía, e *rapadureiro*, no Rio Grande do Sul. O grau de concentração do *melado* para fazer *rapadura* é designado *ponto da rapadura*, em praticamente todos os Estados, apresentando as variações denominativas: *ponto do mel*, em Pernambuco, *trempe* e *tempre*, em Pernambuco e na Baía, e *puxa*, em Minas Gerais. Os moldes quadrados de madeira em que se enforma a *rapadura* são denominados: *formas de rapadura*, nos vários Estados brasileiros,

sendo que, em Pernambuco, encontrámos as variantes *foima*, *formazinha* e *caixinha de madeira*, enquanto, em Minas Gerais, registámos as variantes *formas de madeira*, *forminha* e *grade(s) da rapadura*. O conjunto das formas é designado *tabletezinhas*, no Ceará; o trabalhador que põe a *rapadura* nas formas é denominado *botador de rapadura*, na Baía; o local ou mesa em que se colocam as formas é o *tendal da rapadura*, na Paraíba e no Ceará.

No Ceará e em Pernambuco, encontrámos um tipo de rapadura puxada que fica mais clara, denominada *alfenim* ou *alfinim*, com as formas sinonímicas *melaço puxado* e *puxa-puxa*, em Pernambuco. O termo de origem árabe *alfenim* é muito antigo, tendo praticamente desaparecido do Português Europeu, conservando-se apenas nos Açores, nas festas do divino Espírito Santo. O termo foi conservado no Brasil, nomeadamente no interior do Ceará e de Pernambuco, onde se conservou uma produção açucareira antiga e artesanal. Em Pernambuco, tal como na Baía e em Minas Gerais, encontrámos também a denominada *rapadura batida* ou *batida*, feita com o melado em ponto de açúcar, ou seja, mais apurada e com cravo, canela e erva doce. No Rio Grande do Sul, nomeadamente em Santo António da Patrulha, entrevistámos um informante que faz *rapadura batida* (numa *batedeira*), chamada *raspa (do tacho)* ou *raspinha do tacho*, denominando-se *raspeiro* o seu fazedor.

6. Fabrico do açúcar de cana

O fabrico do açúcar de cana exige uma maior limpeza do caldo da cana, utilizando-se para isso vários produtos, nomeadamente *água de cal*, *semente de carrapateira*, *mamona*, *azeite de dende* ou *cinza*. Para denominar este produto utilizado para limpar o sumo da cana na fervura, registámos o termo *decoada*, no Rio Grande do Sul, termo muito antigo, atestado na documentação histórica da Madeira. As impurezas contidas no caldo da cana são denominadas *impurezas* e *sujeira*, na Paraíba, *sujeira* e *escuma* no Ceará e na Baía, *bascuio*, *impurezas*, *tiborna* e *borra(s)*, em Pernambuco, *escuma do caldo* e *impureza da cana* em Minas Gerais, *cisco* em Santa Catarina e *escuma preta* no Rio Grande do Sul. O instrumento utilizado para retirar as impurezas do caldo da cana, *limpar o caldo* ou *escumar*, é

denominado *passadeira* na Paraíba, no Ceará e em Pernambuco, onde também ocorrem as denominações *peneira de arame* e *escumadeira*. Na Baía, registámos os termos *passador* e *balde*, enquanto, em Minas Gerais, encontrámos as denominações (*a*)*panhador da impureza*, *cuia de cobre*, *canecão grande*, *colher furada* e *escumadeira*, termo que também ocorre em Santa Catarina e no Rio Grande do Sul.

O grau de concentração do açúcar é denominado *ponto do açúcar* em praticamente todos os Estados, tendo sido registados ainda os termos *arrocho*⁵ em Pernambuco, *tempre* na Baía e *ponto da forma* em Minas Gerais. O melado, depois de atingir o ponto de açúcar é retirado do tacho para um *cocho*, chamado *resfriadeira*, para *esfriar*, para isso é mexido com um instrumento denominado *palheta grande* na Paraíba e *rominhol* ou *reminhol* em Minas Gerais. *Rominhol* ou *reminhol* é um termo muito antigo, provavelmente do latim *remus* ou *rame* (cobre), que denomina uma colher de cobre com um cabo comprido. As primeiras atestações do termo ocorrem na Sicília e em Valência, na primeira metade do século XV, respectivamente com as formas *ramiols* e *romiols*. Na primeira metade do século XVI, mais propriamente em 1535, registámos a forma *raminhois* na documentação madeirense⁶.

O processo de formação de cristais de açúcar é denominado *cristalizar*, na Paraíba e no Rio Grande do Sul, com a variação denominativa *açucarar*, termo que também ocorre em Pernambuco, com a variante *açucar*, e em Minas Gerais com os sinónimos *embolar*, *granar* e *coagular*. O utensílio, geralmente de madeira, em que se cristaliza o açúcar é denominado *forma (de açúcar)*, em praticamente todos

⁵ Encontrámos a palavra *arrocho* (pedaço de madeira), associada às formas *arrochar* (1553), fixar utilizando o *arrocho*, por extensão apertar(-se) muito e *arrochado* (1553), apertado com *arrocho*, por extensão fortemente apertado, no *Dicionário Houaiss da Língua Portuguesa* (tomo iii, p. 872). No contexto da terminologia açucareira, *arrocho* denomina o ponto do açúcar, por este ser o ponto mais apertado ou de maior concentração e cozedura do sumo da cana. Na ilha da Madeira, nas regiões rurais mais isoladas, ainda ouvimos a forma *arrochiado* como sinónimo de apertado.

⁶ Cf. Naidea Nunes, *Palavras doces. Terminologia e tecnologia históricas e actuais da cultura açucareira: do Mediterrâneo ao Atlântico*, Governo Regional da Madeira, Secretaria Regional do Turismo e Cultura, Centro de Estudos de História do Atlântico, 2003, p.540.

os Estados, ocorrendo a variação *foimas de tauba*, em Pernambuco, e *forma de madeira* e *sinozinho*, em Minas Gerais.

O mel que escorre das formas de açúcar é denominado *mel de furo*, na Paraíba e em Pernambuco; *melado do tanque* em Minas Gerais e no Rio Grande do Sul, sendo que, em Minas Gerais, ocorrem também os termos *melado de açúcar*, *melado do pingo* e *melaço*, para denominar o mesmo conceito, e, no Rio Grande do Sul, registámos também o termo *melaço* e ainda *melado escorrido*. Como podemos verificar, o termo *melaço*, que denomina este subproduto do açúcar, confunde-se com o termo *melado*, surgindo este com diferentes especificações, *de açúcar*, *do pingo* e *escorrido*, para que não se confunda com o melado de cana.

Em Minas Gerais, Santa Catarina e Rio Grande do Sul, encontramos um açúcar com melado muito grosseiro, denominado respectivamente *açúcar batido*, *açúcar esfregado* ou *açucão*; *açúcar escura* e *açúcar grossa*; e *açúcar embolado*, *açúcar grosso* e *açúcar esfregado*. Esta denominação deve-se ao processo de se esfregar o açúcar, formando cristais grossos. O açúcar com menos melado é denominado *açúcar mascavo* em todos os Estados, apresentando grande variação terminológica, nomeadamente *açúcar bruto* e *açúcar vermelho* (Pernambuco); *açúcar mulatinho*, *açúcar batido* e *açúcar quebrado* (Baía); *açúcar vermelho*, *açúcar moreno* e *açúcar preto* (Minas Gerais); *açúcar vermelho* e *açúcar amarelinho* (Rio Grande do Sul). O açúcar mais claro, purgado, que é separado do mel, é designado *açúcar branco* em todos os Estados. Registámos o termo *açúcar cristal* na Paraíba e em Minas Gerais, a par dos termos *açúcar filtrado* e *açúcar escorrido*, também registado no Rio Grande do Sul. O trabalhador responsável pelo fabrico do açúcar é denominado *mestre de açúcar*, na Paraíba e em Pernambuco, enquanto, na Baía, registámos o termo *cozinheiro*, em Minas Gerais *engenheiro* e, no Rio Grande do Sul, *açucareiro*.

7. Fabrico de aguardente de cana

A bebida alcoólica feita do sumo da cana ou do mel é denominada *cachaça* em todos os Estados brasileiros. Na Paraíba, registámos, a par da denominação *cachaça*, os termos *aguardente* e *pinga* para

denominar o mesmo conceito. Encontrámos também a denominação *aguardente* em Pernambuco e a designação *pinga de melado*, *pinga de cana* e *pinga (de garapa)* em Minas Gerais, a par dos termos *cachaça de melado* e *cachaça de garapa*. Na Baía e em Santa Catarina, apenas encontrámos o termo *cachaça (de caldo)*, enquanto, no Rio Grande do Sul, o termo simples *cachaça* surge a par das denominações *cachaça de ouro* (envelhecida em pipas de carvalho) e *cachaça prata*. A aguardente feita do mel escorrido das formas do açúcar, em Pernambuco, é denominada *aguardente de mel de furo*.

O fabrico de aguardente ou cachaça exige a fermentação do sumo da cana, realizada em *dornas de fermentação*, termo registado na Paraíba, em Pernambuco, Baía e no Rio Grande do Sul, a par do termo *cuba*, na Paraíba e em Pernambuco, enquanto, em Minas Gerais, encontrámos as denominações *vasilha de fermentação* e *tambor*. A preparação do *fermento* denomina-se *pé de cuba*, termo registado na Paraíba e no Rio Grande do Sul. O sumo da cana fermentado, geralmente, é denominado *caldo fermentado* ou *garapa fermentada*, apresentando alguma variação terminológica, nomeadamente *vinho da cana*, na Paraíba; *caldo de cana azedo*, em Pernambuco; *vinho*, na Baía; *calda* e *ponto de lambicar*, em Minas Gerais; e *guarapa azeda* e *vinhada*, no Rio Grande do Sul.

O aparelho em que se fabrica a cachaça é denominado *alambique* ou *lambique*, em todos os Estados, sendo também denominado *panela de cobre*, na Paraíba, em Pernambuco e em Santa Catarina; e *forno*, *coluna* e *destilador*, no Rio Grande do Sul. O processo de fabricar aguardente denomina-se *destilar* em todos os Estados, apresentando a variante *estilar*, em Pernambuco e no Rio Grande do Sul, e os sinónimos *ferver* em Minas Gerais e *lambicar*, no Rio Grande do Sul. A aguardente muito forte ou muito fraca pode voltar ao alambique para *redestilar* ou *restilar*, é a chamada *cachaça bidestilada*. A primeira aguardente, muito forte, que sai do alambique é denominada *cabeça*, praticamente em todos os Estados. Na Paraíba, o termo *cachaça de cabeça* é sinónimo de *cana de cabeça*, enquanto, em Pernambuco, *cachaça de cabeça* é sinónimo de *aguardente de cabeça*. Em Minas Gerais, *cabeça* é sinónimo de *restilo* (porque esta aguardente tem de ser *restilada*) e de *cabeceira*, enquanto, em Santa Catarina e no Rio Grande do Sul, *cabeça* é sinónimo da denominação *cabeça de*

cachaça, em vez de *cachaça de cabeça*. A segunda aguardente que sai do alambique é denominada *cachaça do coração*, *coração da cachaça* ou simplesmente *coração*, na Paraíba, em Pernambuco, em Santa Catarina e no Rio Grande do Sul, enquanto, na Baía, apenas registámos o termo *cachaça* e em Minas Gerais *cachaça* e *pinga*. A última aguardente que sai do alambique, muito fraca, é designada *cauda*, na Paraíba e em Santa Catarina, onde também é denominada *rabo* e *água-fraca*. Encontrámos também a denominação *rabo* no Rio Grande do Sul e *água-fraca* em Minas Gerais, enquanto, em Pernambuco, registámos o termo popular *caixixi* e na Baía a designação *cachaço*, forma masculina de *cachaça*, com valor depreciativo. O trabalhador responsável pelo fabrico da cachaça é denominado *mestre da cachaça*, na Paraíba; *cachaceiro*, em Pernambuco, Minas Gerais, Santa Catarina e Rio Grande do Sul (embora este nome também designe alguém que bebe muita cachaça), a par do nome (*a*) *lambiqueiro*, em Minas Gerais e no Rio Grande do Sul.

A variação terminológica ocorre, muitas vezes, nas entrevistas realizadas aos especialistas, agricultores e produtores de rapadura e alfenim, de melado, de açúcar mascavo e de cachaça, nos vários Estados brasileiros representados neste estudo. São frequentes os sinónimos no discurso de um mesmo informante, como por exemplo: *mel de engenho*, *melado* e *melaço*, na Baía; *água-fraca*, *cauda* e *rabo*, em Santa Catarina. A polissemia também é frequente, por exemplo nos termos: *melaço* (subproduto do açúcar sem sacarose e mel de cana) e *forno* (fornalha e tacho de cozer o mel). É interessante verificar que os termos polissémicos resultam dum processo de extensão semântica, geralmente por contiguidade, provocando variação terminológica, perdendo o valor distintivo e conduzindo à mudança linguística. Assim, muitas vezes, a variação terminológica, sinónimos e termos polissémicos, é indicação de que está a ocorrer uma mudança linguística, tendendo a desaparecer o termo mais antigo e mais especializado ou técnico, por exemplo: *parol* tende a ser substituído por *depósito de madeira*, *tanque*, *cocho*, *decantador* e *tina de madeira*; *escumadeira* tende a ser substituída por *passadeira*, *passador*, *peneira de arame*, *balde*, *cuia de cobre*, *canecão grande*, *colher furada* e (*a*) *panhador da impureza*.

No que se refere à variação histórica e geográfica, verificamos a existência de regiões mais conservadoras, as primeiras regiões de

produção açucareira (Nordeste Brasileiro) e a região de conservação da tecnologia açucareira mais antiga (Minas Gerais), onde ocorrem os termos mais antigos, por exemplo: *mestre (de açúcar)* na Paraíba; *alfenim* ou *alfinim* no Ceará e Pernambuco; *parol* na Baía; *reminhol* ou *rominhol* em Minas Gerais e ausência destes termos mais antigos em Santa Catarina e no Rio Grande do Sul, regiões mais inovadoras, onde ocorrem termos mais recentes, por razões histórico-geográficas. No entanto, as regiões mais conservadoras, a par dos termos mais antigos, apresentam algumas inovações lexicais ou terminológicas, por exemplo: *capinar* ou *carpinar* (limpar o mato das plantações de cana), a par do termo mais antigo *roçar*, em Pernambuco e Minas Gerais, tal como em Santa Catarina e no Rio Grande do Sul, regiões onde apenas ocorre o termo *capinar*. Em Minas Gerais, a par da palavra *moer*, ocorre o termo *capitar a cana* e, em Minas Gerais e Pernambuco, tal como no Rio Grande do Sul, ocorrem as formas *açucarar*, *coagular*, *embolar* e *granar*, como sinónimos de cristalizar o açúcar.

Relativamente à variação social, na variação entre os sexos, dado o menor número de mulheres entrevistadas (apenas duas), não é possível apurar a existência de diferenças significativas, tal como entre os informantes mais idosos e os mais jovens (apenas um). O facto de se tratar de uma área de actividade especializada, com terminologia técnica, pode explicar a menor variação social e geográfica. Assim, registámos uma grande uniformidade terminológica intra e inter Estados, sendo que as variações dialectais e sociolinguísticas não são muito evidentes. No entanto, registámos algumas variações sociais e terminológicas entre denominações mais populares e termos mais técnicos, por exemplo: *caldo (de cana)* denominação mais popular e *garapa* designação mais técnica do sumo da cana; *melaço* denominação mais popular e *melado* designação mais técnica do mel de cana.

Verificamos que as novas realidades e necessidades sociais conduzem ao alargamento vocabular da área terminológica, neste caso da terminologia açucareira, através da introdução de novas palavras ou termos mais recentes e, geralmente, menos especializados, como já referimos anteriormente. Assim, a variação sincrónica entre duas ou mais formas concorrentes ou alternativas, em que uma delas acaba

por prevalecer em detrimento das outras, por pressão ou prestígio social, é um indicador de mudanças linguísticas em curso, como já mostraram os estudos de geolinguística e sociolinguística.

Através deste estudo da terminologia açucareira actual do Brasil, podemos concluir que os fenómenos de mudança e variação nas linguagens de especialidade estão sujeitos a percursos de mudança e a factores que a condicionam do mesmo tipo que aqueles que ocorrem na língua comum ou corrente. Também podemos verificar que os fenómenos de variação e mudança encontrados no Português do Brasil são semelhantes aos que ocorrem nas outras regiões açucareiras estudadas, nomeadamente na Madeira, Canárias, Cabo Verde e S. Tomé e Príncipe, havendo grande proximidade entre as duas margens do Atlântico.

Referências

- Cabré, M. T. 1993. *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Editorial Antártica/Empúries.
- Nunes, N. N. 2003. *Palavras Doces. Terminologia e tecnologia históricas e actuais da cultura açucareira: do Mediterrâneo ao Atlântico*. Funchal: Centro de Estudos de História do Atlântico, Secretaria Regional do Turismo e Cultura.
- Weinreich, U.; Labov, W.; Herzog, M. I. 1968. Trad. port. de M. Bagno. *Fundamentos empíricos para uma teoria da mudança linguística*. S. Paulo: Parábola, 2006.

Dicionários

- Cunha, A. G. 1992. *Dicionário Etimológico Nova Fronteira da Língua Portuguesa*. Rio de Janeiro: Nova Fronteira (2ª edição revista e acrescida de 124 páginas).
- Dicionário Houaiss da Língua Portuguesa*. Lisboa: Temas e Debates, 2005, 18 vols.
- Novo Aurélio século XXI. O Dicionário da Língua Portuguesa*. Rio de Janeiro: Nova Fronteira, 1999.

A descrição do vocalismo átono quinhentista: linhas e entrelinhas nos textos metalinguísticos coevos

Maria Helena Paiva

mhelenapaiva@netcabo.pt

Faculdade de Letras da Universidade do Porto (Portugal)

RESUMO. Com base no tratamento informático dos textos de gramáticos e ortógrafos quinhentistas, analisam-se as descrições de fones e as relações entre estas e factos sincrónicos e diacrónicos igualmente descritos; as ocorrências das palavras seleccionadas como objecto de descrição, ou ilustração dos factos descritos são comparadas com as ocorrências não metalinguísticas no mesmo autor e em diferentes autores. Sendo a elevação de <a> átono, o único ponto que é incontroverso na matéria, dada a clareza da descrição do primeiro gramático, Fernão d'Oliveira, a informação permite testar a fiabilidade dos outros testemunhos. A elevação do restante vocalismo átono continua a ser um tema em debate que apresenta uma vasta área de questões comuns, começando pela cronologia. Tendo em conta que o contributo mais importante é o de Fernão d'Oliveira, analisaram-se os conceitos básicos («letra», [vogal] «grande» e «pequena»), definiu-se a perspectiva dominante como sendo a da escrita, constatou-se a maior incidência das descrições nos fones representados por <e> e <i>, <o> e <u>, em consonância com as afirmações da proximidade entre eles, e observaram-se contrastes entre a doutrina expressa e a informação contida nas palavras-testemunho, bem como entre as grafias preconizadas e as grafias genuínas do Autor. A carência de uma terminologia adequada, evidente ainda em Soares Barbosa (1822), projecta-se em dificuldades de conceitualização, também patentes nos restantes metalinguistas, cuja informação complementar é sintetizada. A variação <e> ~ <i> e <o> ~ <u> origina um número elevado de referências; as interpretações e relacionações nem sempre adequadas, não deixam por isso de ser elucidativas, quando comparadas com mudanças por assimilação, dissimilação e analogia, quer ocorridas no passado quer em curso na época em estudo. No conjunto, a exploração das descrições conduz parcialmente à caracterização, no plano do vocalismo átono, do estado de língua sobre que virá a desenvolver-se o movimento de relatinização.

PALAVRAS-CHAVE. Vocalismo átono quinhentista. Fonética histórica. Tratamento informático. História do Português. História da Gramática.

ABSTRACT. Based on the computer processing of the 16th century grammarians, we have analyzed the definitions of phones and the relationships between these and synchronic and diachronic facts, which are also defined; the forms of the words that have been selected as the object to be described or as the illustration of the described facts are compared with the forms of the same words employed in non metalinguistic uses. The raising of atonic <a> being the only point that is undisputed in this matter (this is due to the clearness of the description of the first grammarian, Fernão d'Oliveira), this information enables us to test the reliability of the other testimonies. The raising of the remaining atonic vowels continues to be a subject of debate, and it presents us with a vast area of common questions, starting with the chronology. The contribution of Fernão d'Oliveira being the most important, we have analysed the basic concepts («letter», [vowel] «pequena» = small, «grande» = big); we have defined the dominant perspective as being that of writing; we have revealed a larger incidence of descriptions in the phones represented by <e> and <i>, <o> and <u>, in accordance with the affirmations of proximity between them, and we have observed contrasts between the expressed doctrine and information conveyed by testimony-words, as well as between the praised writings and the genuine writings of the Author. The lack of an adjusted terminology, which is still evident in Soares Barbosa (1822) is reflected in the difficulties in conceptualization, also clear in the other Authors, whose complementary information we have synthesised. The variation <e> ~ <i> and <o> ~ <u> give rise to a large number of references; the interpretations and connections are not always appropriate, but this does not make them less informative, at least when they are compared with changes by assimilation, dissimilation and analogy, whether these changes occurred in the past or are still in progress at that time. As a whole, the exploration of the descriptions leads partly to the characterization, in the plane of the atonic vowels, of the state of the language on which the movement of relatinisation will spread.

KEY-WORDS. 16th century atonic vowels. Historical Phonetics. Computer processing. History of Portuguese. History of Grammar.

1. Introdução

O português é sem dúvida uma das línguas românicas em que é maior o contraste entre um vocalismo tónico extremamente conservador e claramente delineado na sua quase totalidade desde os primeiros textos graficamente consistentes, e um vocalismo átono não só de difícil apreensão, mas em evolução ininterrupta até um presente caracterizado pela crescente consonantização, ou seja pela

não articulação, em contextos determinados, de determinadas vogais átonas. No percurso traçado, três momentos merecem particular atenção: o período inicial, porque constitui a matriz donde todo o resto deriva; o século XVI, porque desse período datam a fixação do padrão linguístico e os primeiros testemunhos de gramáticos e ortógrafos; e finalmente o século XVIII, por surgirem então afirmações peremptórias sobre a elevação de [e] e [o] átonos. Não admira por isso que o assunto seja um tema obrigatório nas obras sobre história da língua portuguesa e tenha suscitado uma ampla reflexão expressa numa vasta bibliografia.

Duas coordenadas configuram este estudo: primeiro, a análise das descrições linguísticas incluídas nos textos; segundo, o tratamento informático do *corpus* metalinguístico quinhentista (a partir de agora, *CMQ*, descrito no anexo) que, reunindo obras situadas entre 1536 e 1606, permite quantificar a variação gráfica nas suas relações com a mudança.

2. <a> átono

A elevação de <a> átono constitui o único ponto sobre que existe unanimidade.

O primeiro testemunho é o de Fernão d'Oliveira, inequívoco, e por isso frequentemente citado: «Temos a grãde como almada & α pequeno como alemanha» (GR 12 23-24)¹. Ao preconizar o uso da

¹ Trancrevo todas as citações das obras constantes do *CMQ*, de acordo com as pré-edições, aligeirando os critérios então retidos (Paiva, 2002, I, p. 70-79). As principais alterações são as seguintes: 1) uso de <i> e <u>, <j> e <v> de acordo com as convenções gráficas actuais; 2) substituição de <ç> por <c> antes de <e> e <i>; 3) substituição de *s longo* por *s de dupla curva*, mas conservação da sua função de demarcação da palavra; 4) introdução de acentos gráficos apenas quanto estes são indispensáveis à inteligibilidade do texto, casos em que a vogal alterada aparece em itálico; 5) A abreviatura que pode ser desdobrada como *pera* ou *para*, variantes que alternam em Oliveira, é representada como p^a. De entre os traços originários conservados, destaco: 1) o uso do til, generalizadamente considerado explicitamente pelos vários gramáticos-ortógrafos como símbolo da nasalidade das vogais, mais adequado do que <m> e <n>, letras que representam propriamente consoantes; 2) a geral conservação da palavra gráfica, sendo a alteração das delimitações originárias assinalada pelas convenções seguintes: [] – reunião de formas separadas: *alproveitão*;

letra grega <α> para representar “a pequeno”, Oliveira considera a distinção suficientemente importante para ser indicada por um sinal gráfico que facilite a sua identificação; e embora, no exemplo que dá, de “αλεμᾶηα”, a letra grega assinale também a vogal tónica fechada antes de nasal, é este o único caso em que é proposto um símbolo para identificar uma vogal átona, visto que o recurso às letras gregas visa distinguir vogais tónicas: <ε> para “e grande” como em *fēsta*, diferente de *festo*, <ω> para “o grande”, como em *fermōsos*, distinto de *fermoso* (12 24-26). Exceptuando a vogal tónica de “αλεμᾶηα”, e a vogal inicial de “almada”, o que a selecção de exemplos acusa é a elevação da vogal átona inicial e final; outros testemunhos corroboram estes: ilustrando o facto de «vogaes [que] se trocãõ»: «αμαρῶ [pretérito perfeito] & αμαρῶ [futuro] » (25 24-26) e, enumerando os «casos» dos «artigos»: « α . δα . [...] .αs. das . » (62 2-3).

Na esteira de Oliveira, Barros distingue a *grande*, que representa por <á>, como em *más* e *á*, contracção da preposição com o artigo feminino, e a *pequeno* como em *mas*, conjunção, a artigo feminino e pronome (nesse caso representado por <â>) e ainda vogal átona final, como em *bõa conciencia* (43r 24 – 44r 1).

Em 1576, na *Ortografia*, a relação entre [a] tónico aberto e [ɐ] átono fechado integra-se já numa regra de derivação, incidentalmente apresentada por Leão como argumento rejeitado contra a sua tese de que «não há mais que hum .a.» (2v 11):

«nas vogaes nenhũa differença teemos dos Latinos, de quem teem origem a nossa lingua. E a razão que faz parecer que são dous .aa. hum grande, & um pequeno, he a pronunciação varia, que se causa dos accentos, ou das letras, a que se ajunta esta vogal. Porque quando teem o accento agudo, parece grande, como em *prato*, & quando graue, parece pequeno, como em *prateleiro*.” (2v 21 – 3r 4).

Esclareça-se que “acento agudo” equivale em Leão a acento tónico e “acento grave” a atonicidade. (2v 20-21). Mas o que merece principalmente destaque é o contraste entre a doutrina, que é falsa, e o exemplo, que é verdadeiro.

Do conjunto de informações depreende-se que a elevação de <a> átono no português padrão do século XVI tem carácter geral e

1) [– separação de formas reunidas: *o][qual*; ” – separação de formas reunidas quando há elisão ou crase: *d”agua*»; 3) a geral não intervenção em matéria de pontuação.

está consolidada em posição pretónica em inicial absoluto de palavra (*alemanha*, *amarão* [pretérito perfeito] ou não (*amarão* [futuro]), como postónica final (*alemanha*, *almada*, *bõa consciencia*) e nas enclíticas (*a* e *as*, artigo, *da* e *das*, contracção da preposição com os artigos, *mas*, conjunção). Não há referência à vogal postónica não final.

A informação veiculada pela aplicação do código gráfico adoptado por Barros permitirá avaliar a extensão do fenómeno.

O exame não exaustivo, mas bastante amplo, da obra de Barros, efectuado a partir do *Índice de Vocábulos* (Paiva, 2002), revela falhas de sistematicidade na aplicação do código, decorrentes de uma insuficiente capacidade de reconhecimento do traço a representar e da dificuldade de integração do particular no geral: assim, Barros escreve *sálvaçám*, mas, em contraste com o exemplo dado por Oliveira, *almada*, o sinal de abertura é completamente omitido nos vocábulos iniciados por *al-*, como *alcançar*, *algum*, *almirante*, *alteza*. O mesmo nos vocábulos em que a vogal inicial se manteve aberta, como em *acerca*, *além*, *avante*.

A muito larga maioria das notações comprova que a redução da vogal átona é um fenómeno que se tornou regular, como o atesta a derivação: *fácil* – *facilidade*, *gráve* – *gravidade*, *liberál* – *liberalidade*, *particulár* – *particularidade*, *prática* – *praticar*, *trabálho* – *trabalhó-sos*.

A conservação da vogal átona aberta regista-se em duas situações: a primeira, como no português padrão contemporâneo, quando a vogal resulta da crase de dois <aa> em hiato devido à queda de consoante intervocálica como em *sádios* (10v 8) (Teyssier 1966: 144); a segunda, em latinismos recentes, cuja origem se mantém desse modo patente: é o que acontece nos derivados de *bárbaro* < lat. BARBĀRUS < gr., vocábulo atestado desde os séculos XIV-XV (Machado, 1990 s.v., Cunha 1996, s.v.), que são provavelmente neologismos criados por Barros *bárbarismo* (3 ocs) e *bárbarizar* (1 oc.).

Pode assim concluir-se que a elevação de <a> átono no português padrão do séc. XVI tem carácter geral e está consolidada, na medida em que não se registam factos de variação, o que retira consistência aos argumentos que, com base no português do Brasil, onde o fenómeno não se dá, são aduzidos para a reconstituição do vocalismo átono quinhentista.

3. O restante vocalismo átono.

3.1. A área comum da problemática.

3.1.1. O teor dos argumentos.

Antes de mais, no plano metodológico, e seleccionando apenas a informação mais frequentemente utilizada, não retirei, como argumentos susceptíveis de caracterizarem o português quinhentista, identidades que têm sido apontadas entre o traço a reconstituir e traços do português do Brasil ou dos crioulos por quatro razões: primeiro, porque a não coincidência entre o português quinhentista e o português do Brasil ficou claramente comprovada relativamente à redução de <a> átono e, relativamente às outras vogais, os resultados no Brasil não são uniformes; segundo, porque a identificação de um fonema por um aloglota é sempre condicionada pelo sistema da sua língua nativa, e por isso, mediante a aplicação de testes específicos a estrangeiros, M^a. R. Delgado Martins obteve «a confirmação de que a percepção dum traço [se faz] em função da sua pertinência na língua falada pelo sujeito» (1986: 150, traduzido); terceiro, porque não é tido em conta o conceito de comunidade linguística, definida como grupo que partilha «o mesmo conjunto de normas» (Labov, 1984: 158, traduzido), o que conduz à superação da variação por expansão de determinadas variantes em detrimento de outras; quarto, porque de um determinado estado do sistema numa variedade linguística em sentido estrito, ou seja, a partir de «germes» de mudança, para adoptar a expressão de Herculano de Carvalho, que a usa no sentido de «condições de desenvolvimentos futuros» (1969 a): 45, traduzido), podem decorrer evoluções simétricas ou assimétricas noutras variedades linguísticas.

Pelo contrário, consideram-se da maior relevância as informações que se apoiam sobre textos portugueses anteriores ou coevos.

3.1.2. A questão cronológica.

Uma acentuada disparidade de opiniões marca o delineamento da mudança, quer no que se refere à sincronia ou assincronia da elevação das vogais átonas representadas por <e> ~ <i> e por <o> ~ <u>, quer no que se refere à localização do facto no tempo.

Limitando a definição do estado da questão às *Histórias da Língua Portuguesa* publicadas entre 1952 e 1991, sintetizamos no *Quadro 1* as concepções assumidas; em vez de fonemas, foram representados fones, por se considerar que, só depois de se saber que um fone existe, se pode discutir se tem ou não função distintiva.

Segundo Teyssier (1982: 58-62), relativamente ao vocalismo átono final, [e] evolui «de acordo com o seguinte esquema: [e] > [i] > [ĩ]», «que deve ter aparecido na segunda metade do séc. XVIII », datando a evolução [o] > [u], aproximadamente da mesma época. Os argumentos mais decisivos são colhidos em obras didáticas sobre o Português destinadas a estrangeiros, publicadas entre 1682 e 1788, a que se acrescenta «em 1767, o *Compendio de Orthografia* de Luís do Monte Carmelo [que] traz listas de ‘erros’» como *cutovelo, murar (morar), purtagem, tucar, xuver*».

Quadro 1. Vocalismo átono – séc. XVI

	final		pretónico	
	<e> ~ <i>	<o ~ <u>	<e> ~ <i>	<o> ~ <u>
S.S.Neto, 1952:483	[i]	[u]	[e]	[o]
P. Teyssier, 1982:56-63	[e]	[o]	[e]	[o]
I. Castro, 1991:251-252	[i]	[u]	[e]	[o]

I. Castro (2006: 195), reproduz as datações propostas por Teyssier (1982), mas considera haver indícios de «uma elevação [o > u] na primeira metade do séc. XVII» e «de uma paralela elevação [e > i] no mesmo século», concluindo estar-se «diante de um problema que merece mais estudo».

3.1.3. A crítica da informação veiculada pelo CMQ.

Dos metalinguistas quinhentistas, Fernão d’Oliveira é aquele cuja informação tem sido justificadamente mais explorada; a sua concepção de que tudo se encontra em perpétuo movimento, “quanto mais as falas que sempre se conformão cõ os conceitos ou entenderes, juizos & tratos dos homens” (50 4-5) suscita o interesse em captar

essa mobilidade, focando a variação, desenvolvendo a acuidade na observação, e elegendo o presente directamente observável como área privilegiada da descrição. Por isso, são as suas concepções que analisamos em primeiro lugar e predominantemente. Não obstante, Oliveira não pode deixar de ser integrado no seu tempo, como os restantes gramáticos e ortógrafos que, quando descrevem a língua, têm de ser lidos com a consciência da distância temporal convertida em progresso no estudo da linguagem.

3.1.3.1. O conceito de «letra».

Talvez nada ilustre melhor este tipo de questões do que o conteúdo do termo «letra» ou concretamente de um dos grafemas, <e> ou <o>, enquanto signos da metalíngua. A perspectiva pode ser gráfica, quando visa a leitura ou a escrita, e especifica o respectivo sinal ou o seu uso: «& se a tal penultima assi de vogaes puras não tiver o acênto não na escreveremos cõ .i. se não cõ .e.» (25 9-10). Também pode ser fonética ou fonológica quando se integra em descrição articulatória ou envolve menção de traço articulatório, quando ilustra factos de fonética combinatória ou sintáctica, quando tem carácter distintivo implícito ou explícito, quando se insere em confronto com fone ou fonema próximo, a nível do sistema: «antre .i. que é letra delgada aguda & viva & antre .e. grande soa na nossa lingua hũa outra voz mais escura: & não mais que hũa: & a este chamamos .e. pequeno» (32 24-27). Finalmente, os dois planos ou não podem ser discriminados ou são simultaneamente activados, como quando a letra é usada pelo fone de que constitui a representação: «Verdade é que depois de g quãdo logo vem .e. ou .i. escrevemos no meio .u.» (20 23-24).

3.1.3.1.1. A exploração da informação quantitativa.

Explorando o *Índice de Vocábulos* (Paiva, 2002) a partir das referências às «letras» vogais na *Gramática*, e tendo em conta que cada lema engloba as variantes gráficas contabilizadas a partir dos grafemas utilizados, constata-se a assimetria do panorama de conjunto, assim delineado, por ordem crescente do número de ocorrências: <a> – 22 ocs.; <u> – 25 ocs.; <i> – 31 ocs.; <o> – 32 ocs.; <e> – 42 ocs.

A análise interna da informação relativa aos grafemas mais frequentemente referidos revela, quanto a <e>, que a «letra» agrupa as seguintes especificações, assim quantificadas: indiscriminadas – 8; [ɛ] – 12; [e] – 20; [ẽ] – 1; [j] – 1.

O pico relativo a [e] suscitou a integração em contexto e conduziu à constatação de que, nesse total de 20 ocorrências, <e> aparece associado a <i> 9 vezes, quer se trate de distinguir quando deve ser usada uma das letras, quer haja referência a particularidades que aproximam ou afastam os fonemas representados. Além destas, distribuem-se por [ɛ] 2 ocs. que situam [e] entre [ɛ] e [i] e uma 3ª, correspondente a [j], em que a letra é referida como alternando com <i>, na representação de ditongos decrescentes.

Quanto a <o>, constatou-se que as 32 ocs. se distribuem assim: indiscriminadas – 5; [ɔ] – 11; [o] – 15; [õ] – 1. No total das 15 ocs. de [o], a letra é associada a <u> em 5 ocs., semelhantes àquelas em que <e> é associado a <i>.

3.1.3.1.2. *A explicitação dos problemas de identificação*

Além dos indícios que os paralelismos numéricos veiculam, outro paralelismo não menos elucidativo é a referência à problemática coincidente, quer referida conjuntamente, quer em separado: «Não pareça a alguém que nós confundimos .i. pequeno cõ .e. pequeno: nem .o. pequeno com .u. pequeno» (32 21-22); «.i. pequeno & .e. pequeno são muy vezinhos» (69 18), «das vogaes antre u & o pequeno há tanta vezinhença que quasi nos confundimos» (25 1-2).

3.1.3.2. «Grande» e «pequeno».

A segunda questão prévia inevitável é a dos conteúdos dos termos *grande* e *pequeno* que adjectivam as vogais em Oliveira. Herculano de Carvalho (1969 c): 81-88) depreendeu oposições no plano da quantidade (*vogal grande* é identificada com vogal longa e *vogal pequena* com vogal breve), no plano da abertura (*vogal grande* é a vogal aberta e *vogal pequena*, a vogal fechada), e no plano do acento (*vogal grande* é a tónica e *vogal pequena*, a átona). A polissemia é um factor permanente de dificuldade de interpretação dos numerosos passos em que os vocábulos são usados; a meu ver, a diferença de quantidade, que Oliveira refere provavelmente por influência da

descrição das línguas clássicas, é a menos relevante, quer porque nunca se reporta sobre o plano do conteúdo, quer porque é sempre perspectivada como decorrente do contexto fonético, quer ainda porque é associada às diferenças de abertura e principalmente à distinção entre tónicas e átonas. Por outro lado, a oposição de abertura só pode ser validada para [a]:[ɐ], [ɛ]:[e] e [ɔ]:[o], mas é incompatível com a afirmação de Oliveira de que <i> e <u>. «sempre são grandes» (32 8-9) De facto, relativamente a <e> e a <o>, a oposição essencial é entre vogal tónica e vogal átona, sendo que as vogais tónicas têm maior duração que as átonas, e o grau de abertura pode estar relacionado com o acento tónico. A polissemia dos termos, um dos traços que justifica a caracterização do saber linguístico quinhentista como pré-científico, tem como consequência que os respectivos conteúdos só possam ser esclarecidos por integração no contexto.

3.1.4. O texto enquanto confluência de nexos.

É o que pode ser ilustrado pelo passo transcrito a seguir, que é justificadamente um dos mais frequentemente citados, em que Oliveira combina o plano da duração com o da intensidade, transitando do «tempo» da articulação, dependente dos fonemas com que as vogais se combinam, para a identificação de *i grande* e *u grande* como tónicos e de *i pequeno* e *u pequeno* como átonos.

.i. & .u. letras vogaes também segundo mais ou menos consoantes de que vierẽ acõpanhadas assi gastarão mais ou menos tempo: mas ellas em .si. sempre são de hũa mesma quantidade & a mi me parece que sempre são grandes como *ouvido* . *escudo* . & em lugar de .i. pequeno serve .e. pequeno como *memorea*, *hostea*, *necessareo* *reverêcea*: nas penultimas : das quaes partes & outras semelhantes eu nũa escreveria .i. se não .e. porque eu tenho que a penultima pura ou ultima qualquer que se escreve cõ .i. sempre tem o acento da dição como *Maria*. *Ouvir*. & as que nam tẽ esse acento da dição escrevense com .e. pequeno como já dissemos. Outro tanto dizemos de .u. vogal como dissemos do .i. o qual .u. vogal sempre é grãde: como *gorgulho*. *arguyo*: e em lugar de .u. pequeno escrevemos .o. pequeno: como *argoyr* *continoar*. Onde se estevera .u. poséramos o acento na penultima como *concluyo*. (32 5-20)

A determinação da perspectiva adoptada pelo gramático condiciona a interpretação do passo: tendo em conta que a *gramática* é tradicionalmente a arte de bem escrever, o objectivo é antes ou

também a definição de um critério depreensível através da observação, mediante a qual a escrita possa ser a correcta.

Que a perspectiva é predominantemente gráfica, esclarecem-no as formulações «em lugar de .i. pequeno *serve* .e. pequeno», «eu nũa escreveria .i.», «ou ultima qualquer que se escreve cõ .i.», «escrevense com .e. pequeno», «em lugar de .u. pequeno *escrevemos* .o. pequeno». Os parâmetros em que a escrita assenta são explicitados: quando a vogal é tónica, ela não pode ser senão representada pelas letras <i> e <u>, que por isso «sempre são grandes como *ouvido. escudo*»; <i> na «penúltima», como em *Maria* ou na última como em *ouvir* «sempre tem o acento da dição»; «outro tanto» [u]: quando tónica, a vogal só pode ser representada pela letra <u> como *gorgulho, arguyo*; como átona, a vogal é representada por <o>: *argoyr, continoar*. O único caso em que a perspectiva é de reconhecimento de uma forma gráfica, portanto a da leitura, é a que se projecta na última frase: «Onde se estivera .u. poséramos o acento na penultima como *concluyo*. »

3.1.5. Vogal em hiato versus semivogal.

Toda a exemplificação que, no passo anteriormente comentado, dá apoio ao critério gráfico, opõe, para a selecção de <e>, palavras cuja forma preconizada termina em <ea> ou <eo> e, para a selecção de <o>, palavras em que a vogal se encontra em hiato – *argoyr, continoar* – que contrastam com formas em que [u] tónico figura num tritongo: *arguyo, concluyo*.

A incidência da pesquisa sobre as palavras-testemunho, ou seja um determinado tipo de ocorrências metalinguísticas, permite obter as primeiras conclusões.

Nas ocorrências metalinguísticas inseridas neste passo relativas a <e> ~ <i>, *hostea* e *reverêcea* são ocorrências únicas, mas *memorea* e *necessareo*, a que acrescentamos *gloria* ~ *glorea*, que é associado à doutrina e a *memoria* em 3 ocs. (25 5 : 2 ocs, 25 11) apresentam variação interpretável.

Das 13 ocorrências do lema *memória*, 10 terminam em <ea>; destas, 8 são metalinguísticas, incluindo as 3 que são preconizadas ou consideradas preferíveis. Só há portanto 2 formas não metalinguísticas em <ea> : «dom João da boa *memorea*» (49 8); «& nós dos nossos faremos *memorea* a seu tẽpo» (73 16). Das restantes

3 ocs., em – <ia> , 1 é metalinguística e, quanto às 2 restantes, nada as distingue das ocs em –<ea>, a não ser o facto de surgirem antes da primeira formulação da doutrina, no fólio 25: «& os socessores deste edificarão em *memoria* & honra de seu capitão» (5 25-26); «& ficará com mayor eternidade a *memoria* delle (10 5-6).

Das 10 ocs. do lema *glória*, 6 são metalinguísticas e destas, 3 ilustram a relação entre <e> e <i> postónicos, de acordo com a doutrina expressa; 1 das restantes, *grorea*, é irrelevante para o problema em análise, visto que constitui uma referência à forma antiga da palavra; as duas outras, exemplos de consoantes líquidas, são grafadas com <i> (21 7 e 21 26). As 4 ocs. restantes, não metalinguísticas, apresentam todas <i>, o que prova que só na zona estrita em que a doutrina é formulada, Oliveira usou a grafia que preconizou.

Finalmente, o lema *necessário*, com 19 ocs., apresenta apenas 2 ocs. metalinguísticas (32 10 e 33 23), ambas grafadas com <e>, declarando-se, na segunda, essa grafia como preferível. Representámos no *Quadro 2* a quantificação da variação, criando três zonas: a primeira corresponde aos totais das terminações <io> e <eo> na *Gramática* , a segunda e a terceira incluem, respectivamente, o número de ocs. antes e depois do fólio 25, onde pela primeira vez a doutrina é formulada, embora só no fólio 32 a palavra figure como exemplo.

O quadro torna claro o sentido da variação: no total da *Gramática*, a grafia com <i> predomina sobre a grafia com <e> na relação de 63, 15 % para 36,84 %. Mas antes da formulação da doutrina, a percentagem de ocorrências com <i> é de 80 % e depois desce para 44,44 %, enquanto a grafia com <e> apresentava antes a percentagem de 20 % e depois sobe para 55,55 %. Não pode deixar de concluir-se que a forma espontânea ou pelo menos dominante em Oliveira é a forma com <i>, e que é a preocupação de coerência entre a norma preconizada e a escrita posta em prática que inverte a tendência, sem contudo a eliminar completamente.

Quadro 2. Necessário

	Na <i>Gramática</i>		antes de fl. 25		depois de fl. 25	
	T.	%	T.	%	T.	%
-<io>	12	63,15	8	80	4	44,44
-<eo>	7	36,84	2	20	5	55,55

A distribuição das variantes de *memória* e de *glória*, anteriormente referida, confirma este juízo. De facto, «a reflexão metalinguística [é] factor simultâneo de homogeneidade e de heterogeneidade» (Paiva, no prelo, § 2.2.), mas as fracturas que deixa no texto são fortemente esclarecedoras, tanto de um pensamento que vai amadurecendo à medida que se exerce, como da realidade linguística que é objecto de exame.

Neste caso, a doutrina que, embora distinga «i grande» e «i pequeno», restringe a letra <i> à representação da vogal tónica e torna «i pequeno» uma realidade virtual, átona, representada por <e> nas terminações <ea>, <eo>, contrasta com as grafias que podem reputar-se de genuínas e é por isso um factor de flutuação. *Memória*, *glória*, *necessário* são formas sem dúvida articuladas com [i], ou com [j], mas o facto de Oliveira não conceituar ditongo crescente não contribui para o esclarecimento da questão. Relativamente a estes vocábulos os restantes textos apresentam grafias constantes com <i>.

Reaproximando-nos do passo que tem vindo a ser comentado, para nos fixarmos na relação entre <o> e <u>, constata-se, antes de mais, a acentuada simetria no plano da doutrina, que é maximamente clara quando o contraste se estabelece no âmbito da flexão verbal: no mesmo verbo, as formas rizotónicas escrevem-se com <u> como *arguyo* enquanto as formas arrizotónicas se escrevem com <o>, como *argoyr*. A indagação sobre as restantes ocorrências das palavras-testemunho na *Gramática* é pouco ou nada informativa (visto que *argoyr* e *concluyo* são ocorrências únicas e de *arguyo* há só outra ocorrência (29 24), também metalinguística, à excepção de *continuar*: das 6 ocs. do verbo, 5 estão de acordo com a doutrina expressa, visto que, sendo arrizotónicas, são grafadas com <o> – *continoar* (2 ocs), *continoando* (2), *continoadas* (1) – mas a restante, rizotónica, na 3ª pessoa do indicativo presente – «não se *continoa* logo outra vogal» (33 26), apresenta [u] tónico representado por <o>, o que só é possível porque a letra <o> tinha passado a representar o fone [u], quando [o] átono tinha evoluído para [u], embora se mantivesse na grafia como <o>, como acontece ainda hoje; havia assim o hábito de fazer corresponder ao fone [u], quando átono, a letra <o>, e esse hábito alastra à representação da vogal quando esta é tónica. É de esperar contudo que casos como este sejam raros numa *Gramática*,

visto que são verdadeiros erros de ortografia, e como tais não passam o filtro da autocrítica. Quando a vogal é átona, a componente conservadora da escrita mantém a letra <o>, pelo menos no âmbito de hábitos individuais, embora o fone representado seja, mais ou menos generalizadamente, [u]. Este juízo é corroborado pela comparação das grafias de *contínuo*, adjectivo, em Oliveira e em Barros: num total de 5 ocs, Oliveira representa a vogal átona por <o> (1 oc.), de acordo com a doutrina, enquanto Barros a representa por <u> (4 ocs.), em simetria com a grafia que adopta para o verbo, não documentado no resto do CMQ, de que há 2 ocs. metalinguísticas na *Gramática*: *continuásse* (20v 15), 1ª pessoa do imperfeito do conjuntivo.

3.1.6. A representação das semivogaes nos ditongos decrescentes.

Dado que uma percentagem elevada de especificações dos fones representados por <e> ~ <i> e por <o> ~ <u> situa as letras antes de outra vogal, podendo <i> e <u> representar uma vogal em hiato ou um ditongo crescente, importa indagar se as descrições e as grafias adoptadas para os ditongos decrescentes, únicos conceituados como ditongos, pode esclarecer a questão.

Oliveira inventaria 16 ditongos, orais ou nasais (26 22-24):

/j/ é representado por <i> ~ <y> em: *ay* [leia-se *ãÿ*] (*mãÿ*), *ei* (*tomei*), *oi* (*caracois*), *oi* (*boi*), *ui* (*fui*); por <e> em: *ae* (*tomae*), *ãe* (*pães*), *oe* (*soe* [verbo *soer*]), *õe* (*põe*).

/w/ é representado por <u> em: *eu* (*meu*), *ou* (*dou*); por <o> em: *ao* (*pao*), *ão* (*pão*), *eo* (*ceo*), *eo* (*deos*), *io* (*fugio*).

Relativamente a <e> ~ <i>, no capítulo sobre a formação do plural, Oliveira informa-nos de que na representação do mesmo fonema intervêm critérios de distinção gráfica baseados na morfologia: /j/ é representado por <i> ~ <y> no plural dos nomes, mas por <e> nos verbos: tendo estabelecido a regra de formação do plural dos «nomes acabados em .ol.», «como *caracol caracois*» (69 3-4), escreve a 2ª pessoa do indicativo com <e>: «*soyo. soes. por acostumar, e royo roes. por roer.* (69 14-15). A dificuldade em dissociar o sinal acústico e o sinal visual manifesta-se no assumir de um eventual erro, que justifica pela proximidade entre os fones [i] e [e]:

Dey a estes nomes no plural estes ditongos .ay. e .oy. cõ .i. & não com .e. porque as minhas orelhas assi o julgão: & não é muito enganarme pois .i. & .e. pequeno são muy vezinhos: mas com tudo os verbos se escreverão com .e. assi soes . roes. tomaes. tomaes. andaes. (69 15-20)

Passando à representação de [w], é notório o predomínio de <o> (5 ditongos), sobre <u> (2 ditongos), o que não deixa dúvidas de que neste contexto, <o> representa um fone, [w], muito próximo de [u]: Oliveira *ouve* a semivogal como «o pequeno», como ele próprio diz, falando das «líquidas»:

Algũas letras se fazem liquidas. Quer dizer liquido aqui brando, ou diminuido de sua força [;] das vogaes nós fazemos .u. liquido algũas vezes despoys de .g. & .q. como *quando*: & *lingua* mas se o meu sentir he acertado eu sinto nos taes lugares .o. pequeno & não já u & assi o escreveria se me atrevesse desta maneyra *lingoa*. *qoando*. porque assi me soa a mi nas minhas orelhas: & se outra cousa fazem por imitar a os latinos não é nosso o que seguẽ. (20 15-23)

Forçoso é constatar que muito poucas vezes se *atreveu* Oliveira e escrever estas palavras como acha que deviam ser escritas, visto que, no total de 130 ocs. do vocábulo *língua*, 126 são grafadas com <u>, e 4 com <o>, entre as quais a que consta do excerto transcrito e, no total de 63 ocs. de *quando*, todas são grafadas com <u>, à excepção da que foi citada acima, e de outra, também metalinguística: «Bem podẽ escusar essa letra .q. como *cadeira*. *coando*. *começo*» (24 21-22).

Clarinda Maia (1986: 426), menciona vários casos de representação de [w] por <o> em documentos dos séculos XIV-XVI, como *agoa*, *agoardar*, *mengoa* o que comprova a antiguidade da tradição gráfica, e a equivalência, em contexto com <g>, entre <o> e a semivogal posterior. Mas a afirmação de Oliveira é inequívoca: para ele, o fone que, em *língua* e *quando*, identifica como «u liquido», é «sentido» como «o pequeno».

3.1.7. Ainda a terminologia.

Também a vogal final de *Marcos* e *Domingos* é identificada como «o pequeno», como diz no capítulo sobre as relações entre terminação e género: «em os. cõ .o. *pequeno*: & em ws com .ω. grande são

masculinos como *marcos domingos, cós, retrós*» (65 9-11). O artigo masculino, quer no singular, quer no plural, é referido, na «declinação» do artigo, como tendo a forma de «o pequeno»: «os artigos masculinos acabão ã .o. pequeno no singular» (62 15-16), «tambê nesse [genitivo] acabão em vogais pequenas os artigos, o masculino ã .o. & o feminino ã .a. (62 17-20), « & no derradeiro [caso] a que os latinos chamão accusativo [...]: acabão em .o. pequeno: os masculinos» (62 22-24), «no plural todos estes acabão nesta letra .s. accretada sobre o seu singular» (62 24-26).

A formulação de Oliveira pode ser esclarecida por comparação com grafias correntes no *Testamento de Afonso II*, manuscrito de Toledo, em que a forma *u* do artigo definido e do pronome demonstrativo, não interpretável como latinismo gráfico, bem como *ou* e *ous* (provenientes da contracção com a preposição *a-*), além de *susu*, preposição < SŪRSŪ, conduziram A. M. Martins (1985: 10-15; 20), a concluir que «*susu* e *u* terão de ser tomadas como testemunho de uma realização –[u] para a vogal átona final, quando esta tem na sua base um – Ū latino»; «em *ou* e *ous* do ms de Toledo, temos a representação de um ditongo em que a primeira vogal foi velarizada por assimilação à semivogal» (1985: 120). Estes juízos adensam a hipótese de ter também carácter fonético a grafia *us*, que foi anteriormente considerada como podendo «interpretar-se como um caso de ‘inércia’ ao nível do processo de selecção grafemática», visto que em contexto com *u*, *ou* e *ous*. (1985 : 20).

Na descodificação do termo «o pequeno», e também no de «e pequeno», não pode deixar de ter-se em conta o conteúdo dessas designações ainda em Jerónimo Soares Barbosa (1737-1816): numa época em que não pode pôr-se em dúvida que a vogal átona representada na escrita por <o> seja [u] e que <e> gráfico átono corresponda a [i], Barbosa teve em conta o facto quando, na *Gramática Filosófica da Língua Portuguesa*, publicada postumamente em 1822, enumerou «vinte vozes» (entenda-se vogais); mas isso não o impediu de se lhes referir deste modo: «E *Pequeno*, como *Se*, *Conjunccção*», «O *Pequeno*, como *O*, artigo masculino». Daí a necessidade de adoptar para [o] e [e], tónicos, respectivamente as designações de «Ê *Grande Fechado*, como *Sê*, verbo» e «Ô *Grande Fechado*, como no Substantivo *Avô*, masculino» (Barbosa 1822: 3).

Muito mais que a persistência dos termos, o conteúdo do passo seguinte, escrito quase dois séculos depois da *Gramática* de Oliveira, acusa a sobrevalorização do escrito em detrimento do oral, e a consequente impossibilidade de perspectivar a escrita como secundária, relativamente à oralidade:

A lingua Portuguesa porêm toca mais dois pontos ou vozes na sua corda vocal; huma entre o E *Pequeno*² e o I *Commum*; e outra entre o O *Pequeno* e o U *Commum*, as quaes, por serem surdas e pouco distinctas, se podem chamar *Ambiguas*, e por isso não tem signal Litteral proprio, e se notão na escriptura, a primeira ja com e ja com *i*, e a segunda ja com o ja com *u*. Taes são as que mal se percebem, quando estas mesmas vogaes se achão em qualquer palavra, ou antes de alguma voz grande immediata, ou depois da mesma nos Diphthongos, e no fim das palavras. Assim e parece ter o mesmo som que *i* nas palavras *Cear*, e *Ciar* (ter zelos) e nos diphthongos destas *Paes*, *Pai*; e pelo mesmo modo o tem o mesmo som confuso que *u* nas finaes de *Paulo*, *Justo*, *Amo*, e nas palavras *Soar*, e *Suar*, e nos Diphthongos, como em *Pao Paulo*, *Seo Seu*. (Barbosa 1822/ 2005, p. 4/ 60)

O facto de a realidade fonética a descrever nos ser conhecida torna patente a sobreposição da grafia sobre a fonia, de que resulta a inversão da relação entre os dois planos: de facto, não são as «vozes» que são pouco audíveis e que por isso são consideradas «surdas», «pouco distinctas», motivo por que «se podem chamar *Ambiguas*», mas a escrita que não dispõe de um grafema que as identifique claramente; não é por serem *ambiguas*, que «não tem signal Litteral proprio», como não é por serem «as que mal se percebem» que tal acontece, como o ilustram de resto os exemplos apresentados, mas porque a realidade fonética evoluiu sem que a escrita tivesse acompanhado a evolução.

O que a interpretação do passo quer dizer, não é evidentemente que «o pequeno», quando átono, equivale em Oliveira a [u], mas sim que não equivale necessariamente a [o] e que, à semelhança da letra que continua a ser a mesma quer o fonema seja tónico ou átono, há passos em que «o pequeno», átono, designa [u] ou [w], não porque, parafraseando-o, entre [o] e [u] haja tanta semelhança «que quase nos confundimos» (25 1-2), mas porque está em curso a elevação da vogal átona.

² Todos os sublinhados do excerto são do Autor.

Por maioria de razão, quando os restantes metalinguistas quinhentistas, menos vocacionados para a descrição da língua oral, falam de «e pequeno» e «o pequeno», isso não quer dizer que estejam a referir-se a [e] e a [o] átonos.

3.1.8. *As restantes descrições gerais.*

João de Barros mantém as designações de «grande» e «pequeno» para <e> e <o>.

Um dos «ofícios» de «e pequeno» é «servir» de conjunção copulativa, como é ilustrado a seguir: «tu e eu e os amigos da pátria louvamos a nóssa linguágem.» (44r 9-11). Como observou Jacinto do Prado Coelho (1946: 221), que cita o exemplo apresentado por Barros, «antes de vogais, já no *Cancioneiro Geral* surge y em vez de e: ‘y elas sem mays ouvir’ (11, 17, 19). Tratando-se de uma proclítica, o facto, no âmbito da fonética sintáctica, deve ser relacionado com a evolução da terminação átona <ea>, <eo>, que comentámos a partir das palavras-testemunho usadas por Oliveira, tendo então concluído pela genuinidade das formas oliveirianas com [i] ou [j] (Ver 3.1.5).

Dos três «ofícios» de «o pequeno» – «artigo masculino», «relativo masculino» e «composiçám das dições» (45r 1-5), o terceiro não chega a ser esclarecedor porque faltam exemplos ilustrativos, o segundo, de pronome – «este livro sempre andaré limpo se ô guardárem bẽm» (45r 4-5) – não se distingue no plano fónico do primeiro que, referente ao artigo, coincide com a informação dada por Oliveira e anteriormente comentada (Ver 3.1.7.).

Como Gândavo não se refere ao vocalismo, restam as descrições de Duarte Nunes de Leão, que enfermam das mesmas características que assinalámos a propósito da negação da existência de [ɐ] (Cf. 2.).

O preconceito etimológico, o predomínio do visual sobre o acústico e o articulatório, o pendor conservador, prejudicam as descrições de Leão que se esforça por comprovar industriosamente a existência não só de um único <a>, mas também de únicos <e>, <o>, quer tónicos, quer átonos:

E hé letra vogal simplez, & não de duas maneiras, como algũs cuidão, que fazem .e. pequeno como em besta por animal, & .e. grande como em bésta per arma, & instrumêto de tirar: o que não há. Porque na pronunçiação dessa letra nenhũa differença teemos dos latinos. E a differença, que vai desse .e. que aos vulgares parece lõgo, ao outro, a que erradamênte chamão breve, notamos com accêto agudo ou circumflexo, ou grave. (LRT 6r 15-23)

Muitos homêes mui doctos, & curiosos da lingoa Hespanhol cuidarão, que acerca de nós havia duas maneiras de .o. hum grande, & outro pequeno, como acerca dos Gregos. Mas, como teemos dicto do .a. assi como não teem mais que hũa figura, assi não teem mais que hũa natureza: que ser longo, ou breve, he accidente, como nas outras vogaes.” (LRT 14r 12-18).

Como os acentos gráficos são raramente usados por Leão, a questão não fica esclarecida. Por vezes, a insuficiência informativa das descrições e das práticas escriturais é suprida pelo que nas entrelinhas se apreende, como acontece no passo seguinte, em que, contrariamente às afirmações constantes acima documentadas, caracteriza a semivogal posterior do ditongo «ão», como tendo «um sabor de .o.», inexistente na sílaba inicial de *campo*, o que não deixa dúvidas quanto à equivalência entre <o> e o fone mais semelhante a [w], ou seja, [u]:

O IIII diphthongo he .ão. [...] sobre que há mais opiniões, & duvida, em que lugares se há de usar. Porque hũs indistinctamente o usão, & o confundem com esta terminação .am. não fazendo de hum a outro differença algũa. O que hé erro manifesto. Porque no fim das palavras, que acabamos com esta pronunçiação, achamos um sabor de .o. que não achamos no fim da primeira syllaba desta palavra, *campo*. (LRT 27v 22 – 28r 7)

3.1.9. *Variação fonética versus analogia.*

Numa época em que estão em curso importantes fenómenos de analogia no âmbito do verbo, importa distinguir os casos em que a variação dos grafemas <o> ~ <u> e <e> ~ <i> é uma consequência dessa dinâmica e aqueles em que pode estar relacionada com a redução das vogais átonas.

Assim, quando Oliveira afirma «das vogaes antre u e o pequeno há tanta vezinhença que quasi nos confundimos dizendo hũs *somir* e outros *sumir*: & *dormir* ou *durmir*, & *bolir* ou *bulir* & outras muitas partes semelhantes.» (25 1-4), nos pares de infinitivos que seleccionou são contrapostas formas de dois tipos: primeiro, aquelas em que <o>

ou é etimológico (*dormir* < DORMĪRE , *bolir* < BŮLLĪRE) ou, no caso de *somir* < SŪMĚRE, resulta da alteração da vogal do radical por integração analógica no grupo de verbos com o no latim clássico, donde SŪMĪT > *some* (Williams, 1975: 218) ; segundo, as formas em que, a partir da 1ª pessoa do indicativo *durmo* < DORMĪO e *bulo* < BŮLLĪO, cujo resultado fonético regular inclui [u] , por inflexão vocálica, é criado analogicamente um infinitivo com [u].

Portanto, a não uniformidade de usos a que Oliveira faz referência não é, como diz, consequência da semelhança dos fones, o não é tão fechado que *quase se confunde com u*, mas a variação que é inerente à expansão de qualquer mudança; a sua interpretação, não sendo adequada, porque o fenómeno de analogia em causa está fora da sua área de conhecimento (embora haja na Gramática a projecção do conceito enquanto princípio de regularidade), não deixa por isso de comprovar a existência de um fenómeno de variação entre [o] e [u] átonos, do mesmo tipo do que é referido a seguir : «E outro tanto antre .i. & .e. pequeno como *memoria* e *memorea*, *gloria*: ou *glorea*.” (OGR 25 4-6).

Também Barros, ao falar da formação do indicativo a partir do infinitivo, estabelece relações em que uma das formas é sempre analógica: «Os verbos da terceira cōjugaã, terminã o infinitivo, ã, *ir*, e fórmam o seu presēte [...] poēdo ã lugar de , *ir*, esta letera , o , e fica formádo *firo*, de *firir* , *durmo* de *durmir*, *sento* de *sētir* , *ubro*, de *ubrir*. » (27v 15- 19).

À excepção de «*sento* de *sentir*», a relação postulada inverte a relação histórica, do mesmo tipo da que foi referida para *durmir*, como é o caso de *firir* (FĚRĪO > *feito*, substituído por *firo*, por integração no esquema SERUĪO > *sirvo* (Williams, 1975: 218) e de *ubrir*, construído a partir de *ubro* < COOPERĪO. *Sento* é igualmente uma forma analógica, mas construída com base em *sentir* < SENTĪRE, uma vez que SENTĪO evoluiu regularmente para *senço*, e para *sinto*, segundo o esquema SĚRUĪO > *sirvo*, acima referido (Williams, 1975: 219).

A procura de simetria, que é uma das características da atitude metalinguística de Barros, pode ter influenciado algum tanto a selecção de formas, visto que Barros usa *ferir* – «o ofício de *ferir* nas outras vogáes» (45r 20), em consonância com Oliveira (*ferir*: 1 oc.), *dormir* – «Os que acábam em , *ir* , sam da terceira: como ouvir, *ir*, *dormir*»

(21r 24-25); mas o significado da existência de <u> em *cubrir* é corroborado por 2 ocs. de *cuberto* (26v 15 e 55r 20), grafia amplamente documentada no período medieval. Quanto à relação *sento* – *sentir*, a forma de 1ª pessoa do indicativo é uma ocorrência única, em Barros, mas há 2 ocs. de *sinto* em Oliveira (20 19, 23 18).

Afectando as vogais representadas por <e> ~ <i> e <o> ~ <u>, estes fenómenos de analogia, de natureza diferente da dos fenómenos fonéticos, contribuíram contudo para gerar a impressão de terreno movediço que reflectem as descrições de Oliveira e de Leão.

4. Conclusão

A clareza da descrição de Oliveira, retomada por Barros, não deixa dúvidas sobre a elevação do fonema átono representado por <a>; negado por Leão, o facto é comprovado pelo próprio exemplo seleccionado para o contestar e pela prática escritural de Barros, que atesta o facto como consumado.

Relativamente ao restante vocalismo átono, não há descrições inequívocas e torna-se necessário ler nas entrelinhas.

Uma finalidade prática não é estranha à produção de gramáticas e ortografias. Sendo assim, a perspectiva dominante não é a da leitura, assente no reconhecimento do significante gráfico como equivalente do significante acústico memorizado, mas a da escrita, inerente implicitamente à actividade do gramático, tal como ele é perspectivado na época, e explicitamente à do ortógrafo. Mesmo o mais arguto observador da língua oral, Fernão d'Oliveira, não escapa a prolongar, enquanto gramático, a sua experiência de mestre de meninos nobres (Mendonça, 1898: 4, 111). O reconhecimento da mudança [a] > [e], mas a não explicitação das mudanças simétricas [o] > [u] e [e] > [i] não prova que estas não se tenham dado ou não se encontrem, pelo menos, em curso. No início do século XIX, para Soares Barbosa, não é perturbador que exista um fonema [e], sempre representado por <a>; mas que um fonema [i] não tenha em caso algum uma representação gráfica específica, e que [u] possa ser representado por <u> e por <o>, suscita-lhe ainda a justificação da grafia, pelo carácter «confuso» ou «ambíguo» dos próprios fones (1822: 4).

Não é por isso de admirar que a explicitude plena se encontre mais facilmente em obras didácticas que contrapõem duas línguas: Luís Caetano de Lima, em 1736, na *Ortographia da Lingua Portuguesa*, considera ainda que as vogais finais representadas pelos grafemas <e> e <o> são respectivamente /e/ e /o/ (Carvalho, 1969 c): 92); mas dois anos antes, em 1734, na *Gramática Italiana*, tinha escrito que os portugueses “mudam quasi o e final em i” (p. 10) e que, quanto à “vogal o [...] no final das palavras” lhe dão “som [...] fechado ou escuro [...] equivocando-o com u” (p. 15).

Assim sendo, e não podendo descurar-se a exploração da dimensão descritiva das obras metalinguísticas, a via indirecta da informação permite colher, em primeiro lugar, indícios que se concentram em determinadas direcções.

O que é inexplicável torna-se claro, se se admitir que a vogal átona representada por <e> não é nem [e] nem [i], mas [ɨ]: a existência dessa vogal é prenunciada pelos picos quantitativos incidentes em <e> e em [e], associado a [i]; ela é objecto de preceitos gráficos, assentes na oposição entre tónica e átona que reservam <i> para a representação da tónica, e, por coerência com o preceito, restringem <e> à representação de «e pequeno», identificado com [i] ou [j], o que contrasta com a escrita genuína do Autor e com o quase indiferente uso de <e> e <i> em ditongos decrescentes. Dadas as restrições de carácter heurístico acima referidas, será necessário corroborar por outros meios o que é mais do que uma simples hipótese, a existência desta «outra vogal mais escura» (OGR 32 26) .

A problemática da identificação da vogal átona representada por <o> é em grande parte comum à que se refere a <e>, mas a redução de [o] átono transparece no uso de <u>, em vez de <o>, no uso de <o> com clara correspondência a [u] ou a [w], e em factos de variação.

Assim, em Oliveira, a inequívoca representação de [u] átono por <o>, nas formas arrizotónicas *argoyr* ou *continoar*, em contraste com as formas rizotónicas, grafadas necessariamente com <u>, mas a conservação abusiva da equivalência [u] = <o>, quando a vogal é tónica, em *continoa*, 3ª pessoa do indicativo presente de *continuar*; a variação <o> ~ <u> nas formas arrizotónicas de *continuar* e em *contínuo*, adjectivo, sendo constantes as grafias com <o> em Oliveira

e com <u> em Barros; a identificação da semivogal posterior de ditongos crescentes com o fone que mais se lhe assemelha, [u], representado por <o> nas grafias *lingoa* e *qoando*, preconizadas mas raramente adoptadas, e a notação indiferente de ditongos decrescentes enumerados como *meu*, *dou* mas também como *pao* (*pau*), *pão*, *ceo*, *deos*, *fugio*. Em todos estes casos, «o pequeno» designa efectivamente [u] ou [w] e é a primeira destas correspondências que é estabelecida relativamente ao final de *Marcos* e às formas masculinas do artigo definido, encontrando-se ainda em Leão a identificação entre o fone por que termina o ditongo *ão* e aquele por que termina a palavra *campo*. Por outro lado, a variação entre *dormir* e *durmir* e outras formas análogas, que Oliveira descreve como ilustrativas da proximidade entre «u e o pequeno», podem resultar de se encontrar em expansão o processo de analogia em que todas se integram, ou de <o> em posição átona representar já [u], como o comprova a relação estabelecida por Barros que, invertendo a relação histórica, é por isso mesmo esclarecedora, no plano fónico: «*durmo*, de *durmir*». Os factos referidos indicam que está em curso a elevação de [o] átono, movimento cuja amplitude só poderá contudo ser avaliada por exploração de outros meios de informação.

Referências

- Barbosa, J. Soares. 1822. *Gramática Filosófica da Língua Portuguesa*. Edição anastática de Amadeu Torres. Lisboa: Academia das Ciências, 2004.
- Carvalho, J. G. Herculano de. 1969a. Sincronia e diacronia nos sistemas vocálicos do crioulo caboverdiano. *Estudos linguísticos*. Coimbra: Atlântida Editora: II, 5-31.
- Carvalho, J. G. Herculano de. 1969b. Le vocalisme atone des parlers créoles du Cap Vert. *Estudos linguísticos*. Coimbra: Atlântida Editora, II, 33-45.
- Carvalho, J. G. Herculano de. 1969c. Nota sobre o vocalismo antigo português: valor dos grafemas e e o em sílaba átona. *Estudos linguísticos*. Coimbra: Atlântida Editora, II 75 -103.
- Castro, I. 1991. *Curso de História da Língua Portuguesa*. Lisboa: Universidade Aberta.
- Castro, I. 2006. *Introdução à História do Português*. 2ª ed. Lisboa: Edições Colibri.
- Coelho, J. do Prado. 1946. Para o estudo da pronúncia do Português medieval. *Revista de Portugal, série A: Língua Portuguesa*. X: 217-221

- Cunha, A. Geraldo da. 1996. *Dicionário Etimológico Nova Fronteira da Língua Portuguesa*. 2ª ed., reimp. Rio de Janeiro: Editora Nova Fronteira.
- Labov, W. 1984. *Sociolinguistic Patterns*. [1ªed.: 1972]. 3ª ed. Filadélfia: University of Pennsylvania Press.
- Lima, L. Caetano de. 1734. *Grammatica Italiana*. Lisboa: Of. da Congregação do Oratório.
- Lima, L. Caetano de. 1736. *Orthographia da Língua Portuguesa*. Lisboa: Of. António Isidoro.
- Maia, C. de Azevedo. 1986. *História do Galego-Português. Estado linguístico da Galiza e do Noroeste de Portugal desde o século XIII ao século XVI*. Coimbra: I.N.I.C.
- Machado, J. Pedro. 1990. *Dicionário Etimológico da Língua Portuguesa*. 6ª ed. Lisboa: Livros Horizonte.
- Marquilhas, R. 2000. *A Faculdade das Letras. Leitura e Escrita em Portugal no séc. XVII*. Lisboa: Imprensa Nacional.
- Martins, A. Maria. 1985. *Elementos para um Comentário Linguístico do Testamento de Afonso II*. Reprodução facsimilada. Lisboa: Faculdade de Letras da Universidade de Lisboa.
- Martins, Mª.Raquel Delgado. 1986. *Sept Études sur la Perception*. Lisboa: I.N.I.C.
- Mendonça, H. Lopes de. 1898. *O Padre Fernando Oliveira e a sua Obra Náutica*. Lisboa: Academia Real das Ciências.
- Neto, S. da Silva. 1952. *História da Língua Portuguesa*. Rio de Janeiro: Livros de Portugal.
- Nunes, J. Joaquim. 1956. *Compêndio de Gramática Histórica Portuguesa (Fonética e Morfologia)*. 5ª ed. Lisboa: Livraria Clássica Editora.
- Paiva, M. H. 2002. *Os Gramáticos Portugueses Quinhentistas e a Fixação do Padrão Linguístico. Contribuição da Informática para o estudo das relações entre funcionamento, variação e mudança*. Vol I: *Objecto e Método*; Vol. II: *Pré-edições*; Vol. III : *Índice Geral de Vocábulos, Índices Alfabéticos de Formas de Outras Línguas*; Vol. IV: *Conclusões*. Dissertação de doutoramento. Porto: Faculdade de Letras da Universidade do Porto.
- Paiva, M. H. No prelo. [Fernando Oliveira :] A língua da *Gramática* à luz dos outros textos». In: C. Morais (Org.). *Fernando Oliveira, Um Humanista Genial*. Aveiro: Universidade de Aveiro.
- Silva, R. V. Mattos e. 1989. *Estruturas Trecentistas. Elementos para uma gramática do Português Arcaico*. Lisboa: Imprensa Nacional.
- Teysier, P. 1966. La prononciation des voyelles portugaises au XVI siècle d'après le système orthographique de João de Barros. *Annali* [Nápoles: Instituto Universitario Orientale, VII,1: 127-198.
- Teysier, P. 1982. *História da Língua Portuguesa*. Tradução de Celso Cunha. Lisboa: Sá da Costa [1ª ed. francesa: Paris: P.U.F., 1980].
- Williams, E. B. 1975. *Do Latim ao Português*. Trad. 3ª ed. Rio de Janeiro: Tempo Brasileiro [1ª ed ingl.: 1938].

Descrição do corpus

Primeira sincronia (tratada exaustivamente)

1. Fernão d'Oliveira

Grammatica da lingoagem portuguesa (1536) – (OGR).....23 538 oc... 35,7 %

2. João de Barros

Texto contínuo da “*Cartinha*” (1539) – (BCA)

Grammatica da lingua portuguesa (1540) – (BGR)

Diálogo em louvor da nossa linguagem” (1540) – (BDL).....24 549 oc.... 37,3 %

Total da primeira sincronia.....48 087 oc.... 73,1%

Segunda sincronia (tratada por amostra aleatória de ¼ de cada texto)

3. Pêro de Magalhães de Gândavo

Regras que ensinam a maneira de escrever e

Orthographia da lingua Portuguesa [...] (1574) – (GRE)

“*Dialogo em defesaõ da lingua Portuguesa*” (1574) – (GDD)..2 097 oc..... 3,1 %

4. Duarte Nunes de Leão

Orthographia da Lingoa Portuguesa (1576) – (LRT).....8 698 oc..... 3,5 %

5. Duarte Nunes de Leão

Origem da Lingoa Portuguesa (1606) – (LRI).....6 684 oc... 10,1 %

Total da segunda sincronia.....17 679 oc.... 26,8%

Total do corpus.....65 766 oc

Mudança genolexical: teoria e realidade

Graça Maria Rio-Torto
gracart@gmail.com

Faculdade de Letras da Universidade de Coimbra (Portugal)

RESUMO. As teorias e as tipologias de mudança necessitam de ser validadas pelos dados históricos das alterações efectivamente ocorridas.

As teorias sequencialistas sobre mudança têm dificuldade em contemplar a enorme rede factores que dominam as alterações de valores das unidades linguísticas. A análise das mudanças semânticas no uso de *-eir-* requer evidencia a necessidade de adoptar um quadro teórico que não dicotomize lexicalização e gramaticalização, mas que incorpore fenómenos de subjectivização e de pragmatização, até porque estes são concomitantes com os demais.

PALAVRAS-CHAVE. Derivação, mudança semântica, subjectivização, sufixo, língua portuguesa

ABSTRACT. Theories and typologies of change need to be supported and validated by historical evidence which can establish the changes that effectively occurred.

Sequentialist theories of change encounter difficulties in accommodating the vast number of factors that determine the changes suffered by the value of linguistic units. An analysis of the semantic changes suffered by *-eir-* reveals the need for a theoretical framework which does not dichotomize lexicalization and grammaticalization, but which takes into account factors such as subjectivization and pragmatization, which after all coexist with the other ones.

KEY-WORDS. Derivation, semantic change, subjectification, suffix, Portuguese language.

«linguistic typology which in the twentieth century was reduced to a synchronic discipline, is here conceived as evolutive typology. Consequently, the theory of grammaticalization is tied, from the very start, to evolutive typology» (Lehmann 2002: 2).

1. Introdução

Constitui objectivo desta comunicação reflectir sobre a relação entre teorias e realidade da mudança linguística, indagar em que medida os esquemas de mutação formulados no âmbito das “teorias da mudança” se compaginam ou não com a efectiva evolução empírica que os usos imprimem a um dado domínio da realidade linguística. Tomamos como objecto de estudo um sector genolexical específico, o que envolve os sufixos *-eir-*, por se prestar a uma reanálise à luz dos objectivos enunciados, uma vez que envolve léxico, semântica e subjectivização, três dimensões sempre interpelantes na análise da linguagem e na teorização sobre mudança linguística.

Os dados empíricos são extraídos (i) do corpus do português organizado por Michael Ferreira e por Mike David (www.corpusdoportugues.org), envolvendo fontes documentais dos séculos XIII a XX¹, (ii) de fontes documentais analisadas em Coelho (2005), representativas do português arcaico (galego-português e médio), e ainda (iii) dos estudos levados a cabo por Viaro (2002; 2007).

2. Breve conspecto teórico

Conhecer o modo como a teoria da mudança se compagina com a efectiva mutação da realidade lexical e semântica que as línguas acusam no seu devir histórico leva-nos forçosamente a reflectir sobre os modelos de representação que, sobre esta matéria, têm sido disponibilizados pela ciência da linguagem. Incontornáveis são, neste âmbito, os modelos de gramaticalização e de lexicalização, tão

¹ As pesquisas realizadas em www.corpusdoportugues.org reportam-se à data de 03-04-2008, em que se procedeu à reconfirmação dos dados anteriormente recolhidos.

recorrentemente invocados para enquadrar as mudanças no interior dos sistemas linguísticos.

Os primeiros modelos de gramaticalização e de lexicalização disponíveis são tipicamente unidirecionalistas.

Para Givón (1979: 208-209), as fases de um processo de gramaticalização percorrem um trilho do seguinte tipo: DISCURSO > SINTAXE > MORFOLOGIA > MORFOFONÉMICA > ZERO.

Neste esquema não é claro o lugar do léxico e da semântica e, por consequência, dos processos de (des)lexicalização e de (des)semanticização.

Para quem dissocia de forma radicalmente antagónica léxico de gramática, a gramaticalização implica deslexicalização, dessemanticização referencial, des-referencialização, um maior grau de funcionalidade morfo-sintáctico, rumo a uma textualização/discursivização e/ou pragmatização acrescida; ao invés, a lexicalização traduz-se por desgramaticalização, acréscimo de semanticização referencial ou de referencialização.

É sensivelmente este o ponto de vista já há muito defendido por Jerzy Kurylowicz, para quem “grammaticalization consists in the increase of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status, eg. from a derivative formant to an inflectional one” (Kurylowicz 1975: 52). A esta concepção de gramaticalização contrapõe o autor a de lexicalização.

Em Heine *et al.* (1991: 213) elencam-se alguns efeitos das diversas modalidades de gramaticalização, traduzidos do seguinte modo nos diferentes níveis de organização da língua:

Quadro I. Modalidades de gramaticalização (Heine *et al.* 1991)

Semantic	Concrete meaning > abstract meaning Lexical content > grammatical content
Pragmatic	Pragmatic function > syntactic function
Morphological	Free form > clitic > bound form Compounding > derivation > inflection
Phonological	Full form > reduced form > loss in segmental status

Neste modelo, pragmatização contrapõe-se a gramaticalização, na medida em que esta se manifesta por maior índice de sintactização e correspondente menor funcionalidade gramática. Também a

gramaticalização se contrapõe a lexicalização, uma vez que a presença acrescida de um conteúdo gramatical corresponde a perda de conteúdo lexical. E não é igualmente linear que a transição de um sentido concreto para um abstracto corresponda necessariamente a gramaticalização.

Os modelos unidireccionalistas são necessariamente redutores, porquanto não permitem uma articulação intermodal entre diferentes componentes da língua.

Uma concepção arquitectural de gramática, como a jackendoffiana, abre portas para outros modos, já não unilineares, mas preferencialmente modulares e interactivos² de encarar a gramaticalização, a lexicalização e bem assim outros processo de alteração do estatuto ou dos valores semióticos das unidades linguísticas.

Um problema crucial para a descrição dos dados em análise consiste no facto de o funcionamento das línguas não se limitar a estes dois pólos de ancoragem – a gramática e o léxico – mas envolver outros níveis de organização, como o semântico e o pragmático, que importa incorporar na modelização da mudança. Como se constata neste trabalho, através da observação dos dados derivacionais (cf. 2), a tradicional visão dicotómica da gramaticalização e da lexicalização não contempla, pelo menos sem algum desconforto ou até contra-dição interna, os processos de discursivização, de pragmatização e de subjectivização (Athanasiadou *et al.* 2006).

Acresce que os esquemas unidireccionais e sequencialistas de mudança dificilmente conseguem explicar fenómenos de coexistência e/ou de sobreposição de conteúdos subjectivos, de tipo avaliativo e/ou de modalização, com conteúdos denotacionais mais específicos, como os locativo, colectivo, instrumental, agentivo.

Nos anos mais recentes, Ataliba de Castilho, baseando-se numa concepção dinâmica e multissistémica da língua, em que não há hierarquização entre componentes ou módulos destas, mas interacção entre eles, propõe que a mudança linguística compreende quatro tipos de

² Tenha-se em conta que as expressões linguísticas exibem simultaneamente propriedades lexicais, gramaticais, semânticas e discursivas. A mente humana, ao processar as informações extralinguísticas e linguísticas, opera simultaneamente sobre o conjunto das categorias lexicais, gramaticais, semânticas e discursivas codificadas nos respectivos sistemas.

processos: gramaticalização; lexicalização (entendida como o processo de criação de itens lexicais); semanticização e discursivização.

Vamos deter-nos sobre o que este autor entende por semanticização, por ser um aspecto crucial para a análise dos dados que nos propomos levar a acabo.

Para Castilho, a Semântica é a componente de criação de significados “baseada em estratégias cognitivas tais como o emolduramento da cena, a hierarquização de seus participantes, a organização do campo visual, a movimentação real ou fictícia dos participantes, sua reconstrução através da metáfora e da metonímia, etc. Daqui resultam as categorias semânticas de dêixis, referenciação, predicação, foricidade e conexidade” (Castilho 2006: 251). Como produto da semanticização temos “os sentidos das palavras (Semântica lexical), as significações composicionais das expressões multivocabulares (Semântica sintática) e os significados inferenciais (Semântica pragmática)” (Castilho 2006: 265).

Como este salienta, uma teoria sobre mudança deve ser multifactorial e interactiva entre condicionantes internas e externas da mudança; deve centrar-se não apenas em fenómenos de gramaticalização e de lexicalização mas também em outros processos como a subjectivização e a pragmatização, que muitas vezes coexistem com os demais, e integrá-los harmonicamente para a explicação dos factos.

O esquema de Heine *et al.* (1991: 157) PERSON-OBJECT-ACTIVITY/PROCESS-SPACE-TIME-QUALITY, se bem que sequencialista, reflecte alguns dos valores associados aos derivados em *-eir-*. Mas onde incluir os valores de intensidade, de avaliação e de expressividade de alguns produtos em *-eir-* (*bigodeira*, *discurseira*, *snoqueira*), e que representam uma fonte de renovação genolexical singular no transcurso dos paradigmas derivacionais do português? Como incorporar em modelos unidireccionais os casos explícitos de valores que se desenvolvem de forma não sucedânea, mas simultânea, ainda que eventualmente com um estatuto ou um peso diferentes? Os trabalhos de Traugott, adiante mencionados, trazem respostas plausíveis para estas interrogações.

2. Análises dos dados

2.1. Uma análise cognitivista

Almeida & Gonçalves (2005) propõem-se descrever, tendo por quadro de referência a relação fillmoriana entre a “cena” de evento e os seus actantes (quem faz o quê, onde, como, com quê, com que finalidade), a rede que explica a extensão polissémica dos diferentes valores que o sufixo *-eiro* manifesta. No centro dessa rede encontra-se o valor agentivo e dele se derivam os valores locativo e intensificador.

O centro básico prototípico dos agentivos é preenchido pelo conjunto dos denotadores de agentes profissionais. Por extensão metafórica, deste grupo derivam-se dois outros: os denotadores de agentes habituais e naturais³. Em ambos os casos está em jogo um mesmo processo cognitivo: a conceptualização de um domínio em termos de outro.

Os locativos e os modais implicam uma alteração de foco nos elementos da cena. “No primeiro caso, focaliza-se o local pelo agente; no segundo, o modo ou atributo da ação” (Almeida & Gonçalves 2005: 244).

Os intensificadores são, no entender dos autores, derivados dos locativos, havendo neles um traço de multiplicidade que é refocalizado como excesso de algo.

Faltaria salientar que os adjectivos que preenchem a classe de Modo/atributo da acção também são marcados por um traço de intensidade e de excessividade, e até não raro de depreciação.

Este constructo reflecte, em parte, o esquema de gramaticalização PERSON-OBJECT-SPACE-TIME-PROCESS-QUALITY.

Mas sobretudo ele postula uma ordem de deriva semântico-derivacional (cf. Quadro II) cuja fundamentação carece de evidência conceptual e empírica, e que não coincide inteiramente com os trilhos de evolução histórica dos valores semânticos efectivamente atestados de *-eir-*, do século XIII até ao presente.

³ “As extensões parecem se dar do seguinte modo: se 1. Trabalhar é agir (agente profissional), então, 2. Praticar algo habitualmente é agir (agente habitual) e 3. Produzir naturalmente é agir (agente natural)” (Almeida & Gonçalves 2005: 244).

Quadro II. Evolução da rede polissêmica de -eiro (Almeida & Gonçalves 2005)

- 1 AGENTE (*pedreiro*)
 1a AGENTE HABITUAL (*fofoqueiro*), 1b AGENTE NATURAL (*coqueiro*)
 2 LOCAL (*cinzeiro, galinheiro*)
 2b. INTENSIFICAÇÃO/EXCESSO (*nevoeiro, lamaceiro*)
 3 MODO/ATRIBUTO da ACÇÃO (*grosseiro, certoiro*)

Para observarmos os dados de *-eir-* ao longo dos séculos recorreremos aos trabalhos de Coelho (2004), sobre o português arcaico em suas duas fases (2.2.), e depois aos trabalhos de Viaro (2007) sobre a evolução deste sufixo e ao www.corpusdoportuguês.org/, disponível em linha (2.3.).

2.2. O sufixo -eir- no português arcaico

Segundo Coelho (2004), são os seguintes os tipos semântico-derivacionais de produtos em *-eir-*, nas duas fases do português arcaico.

Quadro III. Tipos semânticos de produtos em *-eir-* no português arcaico (Coelho 2004)

Tipos derivacionais Coelho 2004	Port. Arcaico 1ª fase Sec. XIII-1385: alguns exemplos	Port. Arcaico 2ª fase 1385-1536: alguns exemplos
(i). Nomes de agente humano	albergueiro, armeiro, barbeiro, besteiro, çapateiro, cavaleiro, conselheiro, despenseiro, escudeiro, guerreiro, justiceiro, mercadeiro, messejeiro, ouvelleiro, pousadeiro, quinhoeiro	archeiro, beesteiro, camareiro, conselheiro, despenseiro, porteiro, pregoeiro, soldadeiro, thesoureiro, trombeiro
(ii).(Nomes) adjetivos	braceiro, certoiro, falseyro, mentireiro, tenreiro, uerdadeyro, usureiro, praceiro	estrangeiro, praceiro, postumeiro, viandeiro
(iii). Nomes de instrumento	cevadeira, esmolneira, geolheira, goteyra, lumeeira, sombreyro	barreira, candieiro, fogueira, topeteira, traseira
(iv).nomes locais	cabeceira, cativeiro, celeiro, galinheiro	catyveiro, galinheiro
(v). Nomes de árvore	figueira, maceeyra, oliveira	Ø ocorrências

Em função das fontes compulsadas e dos dados nela recolhidos⁴, verifica-se que os nomes de agente humano são os mais numerosos, com percentagens que se situam acima dos 50%, seguindo-se os adjectivos, muitas vezes usados como predicadores de ser humano, com percentagens acima dos 20%. Abaixo situam-se os instrumentais e os locativos, com valores entre 8% e 14% e, no fim da tabela, os nomes de árvores, com apenas 2,3%, e \emptyset na segunda fase.

⁴ Os dados recolhidos por Coelho são bastante diversificados, no que diz respeito à tipologia das fontes compulsadas – fontes secundárias, para o galego-português, e fontes primárias, para o português médio –, no que diz respeito à natureza tipológica dos textos analisados, em ambos os períodos, e no que diz respeito ao volume de dados tratados de cada uma das fases do português arcaico.

As fontes seleccionadas para o galego-português são de três tipos:

- (i) documentação poética: glossário da edição crítica do *Cancioneiro da Ajuda* (Carolina Michaëlis de Vasconcelos 1990. *Cancioneiro da Ajuda*. Vol. I (*Glossário do Cancioneiro da Ajuda*). Revista Lusitana, vol. XXIII, 1920. Lisboa: Imprensa Nacional-Casa da Moeda); glossário da edição crítica das *Cantigas de escárnio e de mal dizer* (M. Rodrigues Lapa 1965. *Cantigas d'escarnho e de mal dizer*. Coimbra: Galaxia); e glossário da edição crítica das *Cantigas de Santa Maria* (Walter Mettmam 1972. *Cantigas de Santa Maria*. Vol. IV (*glossário*). Coimbra: Universitatis Conimbrigensis)
- (ii) documentação em prosa não literária (documentos de natureza judicial): glossário da edição crítica do *Foro Real* (José de Azevedo Ferreira 1987. *Foro Real*, vol. II (*glossário*). Lisboa: INIC).
- (iii) documentação em prosa literária, traduzida (*Índice geral de palavras lexicais dos Diálogos de São Gregório*: Rosa Virgínia Mattos e Silva 1971. *A mais antiga versão portuguesa dos quarto livros dos Diálogos de São Gregório*. Vol. IV (*índice geral das palavras lexicais*). Tese de Doutoramento. São Paulo: Universidade de São Paulo), ou escrita em português (Bertil Maler 1964. *Orto do Esposo*. Vol III (*glossário*). Stockholm: Almqvist & Wiksell).

Para o português médio, a recolha de dados recai sobre as 100 primeiras páginas de cada uma das seguintes obras: *Crónica de D. Pedro de Menezes*, de Gomes Eanes de Zurara. Edição e Estudo de Maria Teresa Brocardo 1997. Lisboa: FCG e JNICT; *Crónica de D. Pedro*, de Fernão Lopes (edizone critica, com introduziona e glosario por Giuliano Macchi 1966. Roma: Ateneo); *Vida e feitos de Júlio César* (ed. Maria Helena Mira Mateus 1971. Lisboa: FCG); *Vida e feitos de D. João II*, de Garcia de Resende (ed. Evelina Verdelho 1994. In: *Livro das obras de Garcia de Resende*. Lisboa: FCG).

Quadro IV. Ocorrências e percentagens de produtos em *-eir-* no português arcaico

Tipos derivacionais (Coelho 2004)	Port. Arcaico 1ª fase Sec. XIII-1385: 126 ocorrências*: 100%	Port. Arcaico 2ª fase 1385-1536: 77 ocorrências: 100%
(i). Nomes de agente humano	66 ocorrências: 52,4%	44 ocorrências: 57,1%
(ii). adjectivos	32 ocorrências: 25,4%	16 ocorrências: 20,8%
(iii). Nomes de instrumento	13 ocorrências: 10,3%	6 ocorrências: 7,8%
(iv). nomes locativos	12 ocorrências: 9,52%	11 ocorrências: 14,3%
(v). Nomes de árvore	3 ocorrências: 2,38%	∅ ocorrências

* Em cada coluna as percentagens reportam-se ao total de ocorrências de cada fase.

2.3. O sufixo *-eir-* ao longo da história

Para termos uma panorâmica mais enriquecida do comportamento de *-eir-* ao longo dos séculos cruzaremos as informações de Coelho (2004) e de Viaro (2007)⁵ com a nossa observação dos dados contextualizados extraídos de www.corpusdoportugues.org⁶. A pesquisa por nós realizada nesta base de dados contempla os derivados em *-eir-* doravante mencionados, em vista a uma confirmação ou infirmação do valor semântico que lhes tem sido atribuído extra-contextualmente, e da interpretação do mesmo em função da fonte textual em que ocorre.

⁵ O estudo de Viaro (2007) baseia-se nos dados empíricos coligidos no Houaiss (2001), posteriormente filtrados por uma criteriosa análise etimológica, que permite expurgar do *corpus* as palavras portadoras de *-eir-* não derivadas. Neste trabalho o autor elenca vinte e cinco significações associadas aos derivados em *-eir-*, que agrupa em sete séries derivativas, cada uma das quais comporta diversas classes semântico-derivacionais.

⁶ A necessidade de observar em contexto os dados em análise torna-se imperiosa, sob pena de podermos incorrer em erros de interpretação. A par com *galinheiro* locativo registado no século XVI, encontrámos (cf. www.corpusdoportugues.org) num texto vicentino *galinheiro* com valor adjectival-atributivo, funcionando como aposto que remete para a actividade profissional de Valejo: “sospirando entre cantar e cantar entrará enha sobrinha e Costança das Ortigas que em todo Val das Corigas nem na vila mui asinha nam jazem tais raparigas. E como entrar sairá a bailar Valejo o *galinheiro* qu’em Tomar chamava ao coelho conejo esse mesmo há de bailar e por fêsta a Ramalhoa bailarà com Pero Luz vestido no seu capuz”.

Antes do século XIII, os produtos em *-eir-* são essencialmente adjetivos (*dianteiro*, *traseiro*) e nomes de agentes de actividades profissionais (*vaqueiro*), dando continuidade ao valor latino de *-ARIU-*. A construção destes dois tipos de derivados nunca foi desactivada pela língua, até ao presente (cf. *boavisteiro* '(simpatizante, adepto) do Boavista', *doleiro* 'vendedor de dólares'), mas ao longo da história constata-se uma reatualização dos sentidos concretos, dominantes nas primeiras fases da língua, em favor de sentidos mais abstractos, fenómeno a que alguns autores identificam como 'bleaching' (Booij 2005: 266).

No século XIII começam a ter maior expressão numérica os nomes locativos, os nomes de "objecto relacionado com x" e os nomes de árvores e arbustivas. A formação destes tipos de nomes em *-eir-* está consolidada no século seguinte, como se depreende da amostragem patente no Quadro III e nas respectivas percentagens documentadas no Quadro IV.

No século XV surgem nomes de estado (*cegueira*)⁷ e nomes de quantidade/intensidade, como *cabeleira* e *nevoeiro* (*neuoeyro*, segundo Cunha (1997), já no século XIV).

No entender de Viaro os séculos XVI e XVII são períodos de grandes novidades no que à produção de derivados em *-eir-* diz respeito. São destes séculos palavras que denotam:

- (i) adjetivos modalizados como *aventureiro*, *casamenteiro*, *grosseiro*, *lisonjeiro* (século XVI: Cunha (1997) e www.corpusdoportugues.org), *gazeteiro*, *mexeriqueiro* (sec. XVII: www.corpusdoportugues.org) que, uma vez nominalizados, denotam "pessoa que tem,

⁷ Nas Cantigas de Santa Maria os nomes atitudinais e/ou de qualidade são portadores de *-dade* (*deslealdade*, *falssidade*, *maldade*, *piadade*), *-eza* (*braveza*, *crueza*, *fereza*, *vileza*), *-ia* (*aleivosia*, *cortesia*, *covardia*, *ousadia*, *sobervia*), *-ura* (*queixadura*, *usura*). As formas *-ice*, *-ícia* ou *-ície* estão ausentes deste corpus. Tenha-se igualmente em conta que, no *Leal Conselheiro*, de D. Duarte, datado de 1437/1438 (cf. Caetano Mocho, 2003), obra de uma época em que a criação lexical de denominações abstractas sofreu um impulso tão significativo, os nomes de propriedade são portadores dos sufixos *-eza* (*avareza*, *blandezza*, *curteza*, *firmeza*, *fraqueza*, *grandeza*, *graveza*, *igualleza*, *largueza*, *limpeza*, *malleza*, *naturalleza*, *pequeneza*, *simpleza*, *riqueza*, *tristeza*), *-dade* (*averssydade*, *castidade*, *contrariedade*, *cujidade*, *famylyardade*, *graciosidade*, *infielidade*, *levydade*), *-ice* (*gargantuyce*, *golsice*, *ledice*, *sandice*, *pequyce*, *velhice*) e *-ura* (*dereitura*, *dulçura*, *fremosura*, *letradura*, *queentura*, *soltura*), não se registando, num texto desta natureza, nomes abstractos em *-eira*, e muito menos marcados por semas jocosos ou expressivos.

manifesta, faz muito x". O adjectivo *certeiro* já se encontra documentado na 1ª fase do Português Arcaico), e

(ii) "acto típico de x", como *ladroeira*. Este encontra-se documentado na *Arte de Furtar* (1645), como denotador de estado ("Meta cada hum a mão em sua consciencia, e achará a prova do que digo, que este mundo he huma *ladroeira*, ou feira da ladra, em que todos chatinaõ interesses, credits, honras, vaidades") e com leitura de estado ou atitudinal ("e se alguns ha, que não se mantenhaõ de outros viventes, tomaõ seu pasto dos frutos alheos, que não cultivaraõ; com que vem a ser huma pura *ladroeira*; tanto que até nas arvores ha ladroens")⁸.

No século XVIII documentam-se alguns gentílicos, denotando "que provém de x", como *mineiro*, *brasileiro*, *campineiro* (Viaro 2007). Segundo Cunha (1997), o gentílico *brasileiro* apenas se documenta no século XIX e *mineiro* já no século XVII.

Importa mencionar que desde sempre em português o sufixo *-eir-* concorreu com muitos outros sufixos também portadores de valor gentílico, pátrio ou étnico, razão pela qual ao longo dos séculos viria a revelar-se menos aproveitado para tais funções. Recorde-se que, de acordo com Viaro (2002), nas 420 *Cantigas de Santa Maria*, compostas no último quartel do séc. XIII (1270-1282), o sufixo predominante para esta função é *-ão*, *-ãa* (v.g. *perssião*, *romão*, *africão*, *aleimãa*, *aldeão*, *vilão*). Os demais étnicos são portadores de *-es* (*burgues*, *genoes*, *frances*, *marques*, *leones*), não ocorrendo formas com o sufixo *-ense*. O sufixo *-en-* está representado apenas no adjectivo *damasceno*.

A grande novidade do século XIX são os nomes de "estado em que se V intensamente", tais como *baboseira*, *barulheira*, *discurseira*, *maluqueira*, *tonteira*, todos usados em registos e/ou com objectivos

⁸ No século XVI João de Barros, nas *Décadas da Asia* (Década Segunda, Livros I-X), regista *ladroeira* com sentido colectivo, como se prova no excerto "e limpa gente como eles traziam, ao menos queimaria as naus que leixara no estaleiro. As quais ele desejava tanto queimar, como tomar a mesma cidade, porque não estava em razão leixar aquela *ladroeira* com os mouros mui escandalizados, e ir ao Mar Roixo e a Ormuz, pera, partido ele, saírem eles dali e fazerem-se senhores de toda aquela costa". (cf. www.corpusdoportugues.org).

expressivos. Os nomes *bebedeira*⁹ e *pasmaceira* já estão atestados no século XVIII. De acordo com o /www.corpusdoportugues/, *barulheira*, *maluqueira* e *tonteira* apenas se registam no século XX, e é igualmente neste século que se abonam *bigodeira* (muito usado por Eça de Queirós, em *A Ilustre Casa de Ramires*), *jeiteira*, *inverneira*, *preguiceira*.

Sumariando: pela sua natureza predicativa, os adjectivos denominais em *-eir-* prestam-se a serem marcados por traços convencionais de modalização e de intensidade, traduzidas por “que gosta de x, que pratica muito x” (*aventureiro*) e/ou por avaliação apreciativa (*certo*, *verdadeiro*) ou depreciativa (*grosseiro*, *interesseiro*), consoante a semântica da base e/ou os valores a esta associados. Os traços de quantidade e de intensidade estão igualmente presentes em muitos locativos (*barreiro*, *pedreira*), em nomes de fenómenos meteorológicos, como *nevoeiro*, *chuveiro* (sec. XV) significando ‘forte pancada de chuva’ (Viaro 2007: 63)¹⁰; e a nomes como *esterqueira* ou *lamaceiro* também estão associados semas avaliativos negativos.

As marcas avaliativa e intensiva/excessiva são, portanto, concomitantes de diversos valores denotacionais veiculados pelos derivados em *-eir-*, nada impedindo, *a priori*, que com eles continuem a coexistir.

3. Capitalização renovada de valores antigos: motivações possíveis

À medida que a língua se aproxima da actualidade, há maior abertura do sufixo à produção de nomes deadjectivais, de estado, de processo e/ou atitudinais em *-eir-*, e simultaneamente uma presença acrescida de marcas de intensidade e de expressividade, de depreciatividade (que pode ser moderada, empática, complacente), patente em *barulheira*, *discurseira*, *maluqueira*, *nervoseira*, *tonteira*, *vergonheira*.

⁹ No século XIV o equivalente era *bebedeçe* (Cunha 1997), por *bebedice*.

¹⁰ Também em documentos do século XVII (cf. /www.corpusdoportugues.org/) *chuveiro* acusa este valor intensivo: Frei Luís de Sousa, *Anais de D. João III*, 1631: “Cessou aquele espanto e terror; mas logo sobrevêo outro, que foi um *chuveiro* de água tao grossa e tao extraordinária, que zo passou polo encarecimento que dizemos de ‘chover a cantaros’”; e *Diário do Conde de Sarzedas* “deu hum pé de vento com hum *chuveiro* tam rijo que dezemparrão os artilheiros o cabrastante ...”.

Não sendo estes valores de intensidade/excessividade/avaliação (e os denotacionais a que estes se acoplam) exclusivos desta fase recente da história da língua, verifica-se todavia uma activação acrescida dos mesmos, exponenciando uma trilha antiga de subjectivização no uso do sufixo.

Estamos, pois, perante um processo de expansão e de enriquecimento semântico, a um tempo de tipo denotacional (reforço na formação de nomes de estado e/ou de atitude), e simultaneamente de tipo subjectivo, expressivo, interpessoal. Este mecanismo de subjectivização, de mais-valia expressiva associada a estes nomes, inscreve-se numa tendência há muito assinalada, e que as palavras de (Traugott 1986: 549) ilustram: “meanings tend to come to refer less to objective situations and more to subjective ones (including speaker point of view)”.

Que razões explicam esta inflexão na semântica de *-eir-*, nos seus primórdios tão centrada na formação de nomes concretos (agentes, objectos, locativos), quando a língua já dispõe de tantos e de tão expressivos sufixos de intensificação e de codificação da (inter)subjectividade?

É sabido que, na evolução das mudanças semânticas, os fenómenos de subjectivização e de pragmatização são tendencialmente posteriores aos demais¹¹: “Pragmatic meanings are grammaticalized latter than non-pragmatic (or prepositional) ones. [...] Meanings typically shift from what is said to what is meant” (Traugott & König 1991: 192-193).

É igualmente sabido que a subjectivização é um fenómeno pervasivo, transversal a vários paradigmas e sectores da língua. Os derivados em *-eir-*, mormente os adjectivos, e sobretudo quando predicam ser humano, desde cedo acumulam valores avaliativos favoráveis (*certeiro*, *verdadeiro*) ou desfavoráveis (*grosseiro*, *mentireiro*). Mas a percepção dos motivos que fazem com que o espectro de actuação

¹¹ Traugott considera como princípios de mudança semântica mais relevantes os que envolvem evolução de uma situação externamente representada para uma situação internamente (avaliativamente, cognitivamente) representada e evolução de situação textual para um estado de coisas assente nas crenças subjectivas do falante. Nas suas palavras, “If there occurs a meaning-shift which, in the process of grammaticalization, entails shifts from one functional-semantic component to another, then such a shift is more likely to be from propositional through textual to expressive then in reverse direction” (Traugott 1982: 256).

de *-eir-* se tenha alargado de modo tão sensível a nomes de atitude (*maroteira*), de qualidade e/ou de estado (*maluqueira*, *piroseira*, *inverneira*, *caloreira*), carreando consigo marcas de subjectividade e de expressividade tão impressionantes, muito provavelmente encontra-se também nas relações de força que os diferentes valores do sufixo foram acusando ao longo do seu transcurso histórico.

A correlação entre o funcionamento de *-eir-* e de outros sufixos cofuncionais, ao longo da história do português, ajuda certamente a explicar as tendências assinaladas. A grande polivalência do sufixo, associada à cristalização e/ou ao desgaste de alguns dos seus valores mais matriciais, propiciam a emergência de valores renovados, um dos quais envolve a valorização nos padrões derivacionais de marcas de subjectivização e de pragmatização que, *a priori*, já estavam neles subliminarmente inscritos.

O valor mais representado de *-eir-*, desde os primórdios da língua portuguesa, é o agentivo, como se comprova nos dados esquematizados nos Quadros III e IV.

Um estudo sistemático dos paradigmas derivacionais nas *Cantigas de Santa Maria*¹² permite confirmar que os derivados em *-eir-* são essencialmente denotadores de profissões. Segundo Viaro (2002), os nomes de profissões mais abundantes terminam em *-eiro*, *-ario*, *-airo* (*leitoario*, *marinheiro*, *tesoureiro*, *cavaleiro*, *pedreiro*). Os deverbais em *-dor* oscilam entre um valor denominativo (*leedor*, *fazedor*, *malfeitor*, *servidor*, *trobador*) e um valor predicativo (*amador*, *peccador*, *sabedor*, *vingador*).

O valor agentivo do sufixo desde cedo concorre, portanto, com o de um outro sufixo isofuncional: *-dor*. Mais tarde, *-eir-* e *-dor* viriam a sofrer a concorrência de *-ista*, que se associa tendencialmente às denominações de profissionais de actividades marcadas por especialização técnica (*anestésista*, *cientista*, *internista*, *pianista*), reservando-se as denominações em *-eir-* essencialmente para actividades artesanais e menos prestigiadas socioculturalmente (*calceteiro*, *guarda-soleiro*, *pedreiro*, *sapateiro*, *taberneiro*).

¹² A análise dos dados não pode abstrair da tipologia de textos em que estes ocorrem e em consonância com a qual foram processados. Não obstante as particularidades temáticas e poético-textuais das *Cantigas de Santa Maria*, a verdade é que dispomos de um estudo global, quantificado e criterioso dos derivados nelas presentes (Viaro 2002), o que nos fez não prescindir do seu concurso.

As denominações de artefactos ‘contentores’ (*açucareiro, bombo-neira, esmolneira, goteyra, louceiro, saladeira, tinteiro*) e de locativos (*pedreira*) incluem no seu semantismo traços de quantidade e/ou de intensidade (*esterqueira, lamaceiro*), que mais tarde se estendem a outras classes de nomes.

Por seu turno, os nomes de propriedade e/ou de estado (*cegueira, gagueira*), os atitudinais (*asneira, desgraceira, ladroeira, maroteira, piroseira*) e as denominações ou os predicadores de ser humano (*aventureiro, casamenteiro, fofoqueiro, usureiro*) também se prestam a carrear uma marca avaliativa, de sinal positivo ou negativo, consoante os valores associados aos denotados.

Ou seja, em nomes de ‘local onde existe grande quantidade de X, ou X em quantidade/intensidade elevada’ (*esterqueira, lamaceira, pedreira*), em nomes colectivos (*buraqueira, ladroeira*) e até de estado (*bebedeira, gagueira*), já estão presentes traços de quantidade e/ou intensidade e/ou depreciatividade que se generalizaram a outros nomes mais recentes, sejam deadjectivais (*tonteira*), atitudinais (*baboseira, griteira, parvoeira, parvalheira*), de processo (*tosseira*), ou de estado (*caloreira, pasmaceira*), e até mesmo a denominais em *-eir-* (*barulheira, bigodeira, inverneira, jeiteira, preguiceira*). Nestes casos de excessividade não se aplica a metáfora “more is up; less is down” (Heine *et al.* 1991: 51), tão invocada pela linguística cognitiva, se os denotados pela base forem marcados depreciativamente.

A par com o processo de marcação subjectiva e expressiva que atinge algumas classes de derivados portadores de *-eir-*, o sufixo passa a acoplar-se a base nominais para formar nomes de sentido avaliativo-intensivo (*bigodeira, inverneira, jeiteira*), alargando a sua amplitude denotacional, combinatorial e categorial.

O transcurso evolutivo dos valores deste sufixo, tal como o de outros, revela uma subjectivização/uma pragmatização de sinal crescente, marcada por acentuada transversalidade, que não anula os valores mais matriciais do sufixo. Trata-se certamente de marcas semânticas convencionais associadas ao uso deste sufixo¹³, distinguindo-os

¹³ Essas significações podem ser encaradas como implicaturas convencionais de sentido, uma vez que são parte do significado culturalmente convencionalizado, codificado e partilhado dos itens lexicais, e não interferem com as condições de verdade do seu uso.

assim dos mais neutros, como *-idade* (*singularidade, racionalidade*) ou *-ia* (*alegria, melhoria, ousadia*).

O traço de expressividade presente em *nervoseira* está ausente de *nervosismo*, marcado pela tecnicidade que caracteriza os derivados em *-ismo*. O mesmo se aplica a *teimosia* e *teimoseira* ‘teimosia intensa, excessiva e verbalizada como expressiva’, por oposição a *teimosia*, não marcado quanto a estes traços. O contraste entre *piroseira* e *pirosice* põe em relevo a marca pejorativa¹⁴, mas não necessariamente intensiva, de *-ice*, ao mesmo tempo que a depreciacção, a intensidade e a expressividade associadas ao derivado em *-eir-*.

A grande polivalência deste sufixo (no sentido de seleccionar várias classes de palavras de base e de gerar produtos de várias classes categoriais), na qual já se inscreve uma natural associação entre quantidade-excessividade-intensidade-depreciatividade-expressividade, e o facto de estes dois últimos traços não marcarem os sufixos cofuncionais *-dade* ou *-ia* mais neutros, na formação de nomes de estado e/ou de actividade, favorecem a associação de tais valores aos derivados em *-eira*.

A activação de valores expressivos, avaliativos, de manifestação da atitude do falante relativamente ao denotado e/ou de suporte de interactividade modalizada para com o interlocutor é certamente um expediente a que a língua recorre para inscrever a subjectividade no seu sistema derivacional, codificando-a de forma muito mais impressiva e reprodutiva, otimizando assim um recurso afixal tão representativo quanto *-eir-* para os fins intercomunicativos que, em última instância, definem e modelam o uso das línguas¹⁵.

Por fim, não existe correspondência entre a explicação de base cognitivista aventada em 2.1. e a realidade empírica observada, pois a

¹⁴ Esta marca depreciativa não é sistemática, como se observa em *meninice* ou *ledice*.

¹⁵ As duas seguintes tendências de mudança semântico-pragmática identificadas por Traugott & König (1991: 208-209) espelham de forma paradigmática a situação de mudança protagonizada por *-eir-*:

“Semantic-pragmatic Tendency I: Meanings based in the external described situation > meanings based in the internal (evaluative/perceptual/cognitive) situation [...]

Semantic-pragmatic Tendency III: Meanings tend to become increasingly situated in the speaker’s subjective belief-state/attitude toward the situation.”

deriva semântica proposta (cf. Quadro III) colide com a cronologia e com os percursos das co-ocorrências efectivamente registadas nas diferentes classes semântico-derivacionais em que *-eir-* está presente. Desde sempre os locativos (*lamaceiro, esterqueira*) acusaram marcas de grande quantidade, de intensidade, de excessividade e de avaliação (tendencialmente negativa), do mesmo modo que aos atributivos (*aventureiro, fofoqueiro, grosseiro, mentireiro, mexeriqueiro*) e aos atitudinais (*asneira, baboseira, maroteira*) também se associam traços de intensidade, de excessividade e de avaliação. Este filão alargou-se a outros nomes, deadjectivais e denominais, que explicitam propriedades/attitudes (*baboseira, parvoeira*), processos (*barulheira, tosseira*), estados (*nervoseira, preguiceira*), mesmo que metereológicos (*caloreira, inverneira*).

Uma concepção sequencialista da mudança dificilmente consegue representar os esquemas de simultaneidade dos vários valores do sufixo, desde os mais remotamente atestados, aos mais recentes. Por isso a necessidade de optar por uma análise que incorpore a multipolaridade – intrassistémica e extrassistémica – de rumos de mudança envolvidos, que consagre as diferentes modalidades de enriquecimento semântico e/ou pragmático activadas, que compagine a coexistência de subjectivização com semantização enriquecida, porque denotacionalmente diferenciada, explicando a prevalência de uns valores sobre os outros.

Referências

- Almeida, M. L. L.; Gonçalves, C. A. 2004. Polissemia sufixal: o caso das formas X-eiro – propostas e problemas. In: *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL, 237-246.
- Athanasiadou, A.; Canakis, B. C. C. (Eds., 2006). *Subjectification*. Berlin: Walter de Gruyter.
- Booij, G. 2005. *The grammar of words*. Oxford: Oxford University Press.
- Caetano Mocho, M. C. 2003. *Formação de palavras em Gramáticas Históricas do Português. Análise de algumas correlações sufixais*. Dissertação de Doutoramento em Linguística, especialidade Morfologia. Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas.
- Castilho, A. T. 2004. Reflexões sobre a teoria da gramaticalização. Contribuição ao debate sobre gramaticalização no contexto do PHPB. In: W. Dietrich; V. Noll (Eds.). *O Português do Brasil. Perspectivas da pesquisa atual*. Madrid/Frankfurt: Iberoamericana/Vervuert, 203-230.

- Castilho, A. T. 2006. Proposta funcionalista de mudança lingüística: os processos de lexicalização, semanticização, discursivização e gramaticalização na constituição das línguas. In: T. Lobo; I. Ribeiro et al. (Orgs.). *Para a história do português brasileiro: novos dados, novas análises*. Salvador: EDUFBA, tomo 1, 223-296.
- Coelho, J. S. B. 2004. *Semântica morfolexical. Contribuições para a descrição do paradigma sufixal do português arcaico*. Dissertação de Doutoramento em Letras, área de Linguística Histórica. Universidade Federal da Bahia. 2 tomos.
- Cunha, A. G. 1997. *Dicionário etimológico Nova Fronteira da Língua Portuguesa*. Rio de Janeiro: Nova Fronteira.
- Givón, T. 1979. *On understanding grammar*. New York/San Francisco/London: Academic Press.
- Ferreira, M.; David, M. www.corpusdoportugues.org.
- Heine, B.; Claudi, U.; Hünnemeyer, F. 1991. From cognition to grammar. Evidence from African languages. In: *Grammaticalization: a conceptual framework*. Chicago/London: The University of Chicago Press, 150-187.
- Hopper, P.; Traugott, E. 2003. *Grammaticalization*. Cambridge: Cambridge University Press. [second edition].
- Kurylowicz, J. 1975. The evolution of grammatical categories. In: *Essais Linguistiques*. Munich: Fink Verlag, II, 38-54.
- Lehmann, C. 2002. Thoughts on grammaticalization second, revised edition. In: ASSiDUE (Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt) 9, Erfurt: <http://www.uni-erfurt.de/sprachwissenschaft/ASSidUE/ASSidUE09.pdf>.
- Mattos e Silva, R. V. 1999. *Estruturas trecentistas: elementos para uma gramática do português arcaico*. Lisboa: Imprensa Nacional – Casa da Moeda.
- Traugott, E. 1982. From propositional to textual and expressive meanings: some semantic-pragmatic aspects of grammaticalization. In: W. P. Lehmann; Y. Malkiel (Eds.). *Perspectives in historical linguistics*. Cambridge: Cambridge University Press, 245-271.
- Traugott, E. 1986. From polysemy to internal semantic reconstruction. In: *Berkeley Linguistics Society*. 12, 530-550.
- Traugott, E.; König, E. 1991. The semantics-pragmatics of grammaticalization revisited. In: E. Traugott; B. Hein (Eds.). *Approaches to grammaticalization. Vol. 1 (Focus in theoretical and methodological issues)*. Amsterdam: John Benjamins, 189-218.
- Viaro, M. E. 2002. A sufixação nas cantigas de Santa Maria. Comunicação ao IX Congresso Brasileiro de língua portuguesa – IP São Paulo: PUCSP. Disponível em: <http://www.fllch.usp.br/dlcv/lport/maeviaro.html>.
- Viaro, M. E. 2007. Estudo diacrônico da formação e da mudança semântica dos sufixos derivacionais eiro/eira na Língua Portuguesa. In: G. Massini-Cagliari et al. (Eds.). *Trilhas de Mattoso Câmara e outras trilhas: Fonologia, Morfologia, Sintaxe*. São Paulo: Editora Cultura Acadêmica, 45-84.

Evolution des prépositions et emplois locatifs en français depuis le XVI^e siècle – syntaxe et conceptualisation de l'espace?

Jaroslav Štichauer
jstichauer@quick.cz
Université Charles, Prague (République tchèque)

RÉSUMÉ. L'article introduit d'abord le concept de locativité forte/faible, définie à l'aide des traits à la fois syntaxiques et conceptuels et exemplifié sur différents syntagmes prépositionnels à valeur locative (*en poche* vs *dans la poche*). Certains types de comportements syntaxiques (pronominalisation par les clitiques *le* et *y*, emplois coordonnés de deux prépositions, possibilité ou impossibilité d'anaphorisation discursive, etc.) sont mis en parallèle avec les propriétés conceptuelles de la configuration cible-site. Toute atteinte à la prototypicité de cette configuration entraîne des irrégularités syntaxiques, comme c'est le cas par exemple de la différence entre *Le vin est livré en tonneaux* vs *Le vin est conservé dans un tonneau*. Dans la première phrase, qui est une instance de la locativité faible, le complément prépositionnel *tonneau* ne peut pas être anaphorisé (**Il/ils est/sont en bois*) alors qu'une telle anaphorisation est tout-à-fait possible dans la deuxième phrase citée. L'analyse se poursuit par la suite dans la perspective diachronique afin d'identifier les différentes étapes qui jalonnent l'évolution de la locativité depuis l'introduction de la préposition *dans* dans la langue du XVI^e siècle. Les témoignages de trois grammaires (Meigret, Oudin, Condillac) des XVI^e, XVII^e et XVIII^e siècles attestent des changements notables intervenus dans les emplois locatifs concernés. On peut ainsi cerner trois phases différentes dans l'opposition entre locativité forte et locativité faible. Selon notre hypothèse, la phase intermédiaire, correspondant grosso modo aux XVI^e et XVII^e siècles, est caractérisée par un suremploi de *dans* par rapport à l'usage contemporain. Les syntagmes prépositionnels dans lesquels *dans* sélectionne des noms désignant des portions d'espace (*dans une plaine/île*) sont considérés comme une espèce de résidu provenant de cette étape. Cette évolution est enfin rapprochée de la phase de surgénéralisation de *dans* lors de l'apprentissage du français par l'enfant.

MOTS-CLÉ. Locativité, syntaxe, conceptualisation de l'espace, diachronie

ABSTRACT. The paper first expounds the concept of strong vs weak locativity defined by means of both syntactic and conceptual features and exemplified by various locative prepositional groups (*en poche* vs *dans la poche*). Certain kinds of syntactic behaviour (pronominalization through *le* and *y*

clitics, coordinated use of two prepositions, possibility or impossibility of anaphorization, etc.) are paralleled to conceptual properties of trajector-landmark (cible-site) configuration. Any violation of the prototypicality of this configuration is shown to bring about syntactic irregularities, as observed e.g. in the difference between *Le vin est livré en tonneaux* vs *Le vin est conservé dans un tonneau*. In the former sentence which is an instantiation of weak locativity, the prepositional complement *tonneau* cannot be anaphorized (**Il/ils est/sont en bois*), while similar anaphorization is quite possible in the latter one. The analysis goes on in the diachronic perspective in order to identify different stages in the development of strong/weak locativity, starting with the introduction of the locative preposition *dans* in the 16th century French. Three grammars (by Meigret, Oudin and Condillac) dating from the 16th to the 18th century provide evidence of notable changes occurred in the afore-mentioned locative constructions. Three different stages can thus be singled out in the development of the strong vs weak locativity opposition. According to our hypothesis, the intermediate stage corresponding roughly to the 16th and 17th centuries can be characterized by an overuse of *dans* when compared to current-day usage. The prepositional locative groups selecting space portion nouns (*dans une plaine/île*) are likely to be residual constructions from that period. Finally, parallels are drawn between the development and overuse of *dans* on the one hand and over-generalization of the same locative preposition during early stages of language acquisition process.

KEY-WORDS. Locativity, syntax, conceptualization of space, diachrony.

Commençons par une observation somme toute banale. Certaines suites prédicatives du type:

(1) *Leurs enfants sont en classe*

sont pronominalisables (cf., entre autres, Noailly (1999: 110) tout aussi bien par *le* que par *y*, comme le montrent les exemples (1a) et (1b), jugés par la plupart des personnes interrogées comme tout-à-fait acceptables:

(1a) *Les nôtres le sont aussi*

(1b) *Les nôtres y sont aussi*

A titre provisoire, on peut constater que cet exemple montre que certaines structures au format: être + *en* + DétØ + N (*en classe*, *en prison*, *en cave*, *en cage*, etc.) ont un double comportement

syntactique: elles ont un caractère à la fois nominal (elles sont donc pronominalisables par le clitique *le*) et locatif, comme le prouve la possibilité de leur pronominalisation par le clitique *y*. Ces séquences attributives (*être en* N) sont à rapprocher de celles du type *être à* + Dét + N qui affichent un comportement syntaxique analogue:

(2) *Nos enfants sont à la neige. Les leurs (y + le) sont aussi*

alors que que les vraies phrases à valeur locative ne sont jamais pronominalisables par *le*:

(3) *Les enfants sont dans une classe. Ils (*le + y) seront toute la matinée.*

A partir de cela, nous formulons l'hypothèse selon laquelle il existe en français deux degrés de locativité, à savoir un degré de locativité faible, représenté prototypiquement par la structure: X *être en* N (*Il est en cave*), et un degré de locativité forte, représenté prototypiquement par la structure: X *être dans* + Dét + N (*Il est dans (une + la + cette) cave*). Une expression à locativité forte réfère à une configuration spatiale plus ou moins précise qui correspond, dans la terminologie d'inspiration cognitiviste, au positionnement d'une „cible“ par rapport à „un site“ (Vandeloise 1986), d'une „figure“ par rapport à un „ground“ (Talmy 2000) ou d'un „trajector“ par rapport à un „landmark“ (Langacker 2008). Une expression à locativité faible (*Il est déjà en classe*) réfère certes à un site et à la position de la cible par rapport à ce site, mais la localisation topologique est en quelque sorte secondaire: la phrase est plus ou moins synonyme de: *Il suit un/des cours*. Comme le montre les tests (1)-(1b), la locativité de cette phrase s'avère pour ainsi dire „résiduelle“. On peut visualiser les différents types d'emplois locatifs et/ou attributifs par le tableau suivant: les traits [+Nom], [+Loc] traduisent en effet la capacité de ces phrases à être pronominalisées par *le* et/ou par *y*. Si elles ont un caractère nominal, elles se voient dotées du trait [+Nom], si elles sont locatives, elles sont dotées du trait [+Loc]. Les deux premières séquences correspondent aux deux degrés de locativité: la première, dotée des traits [-Nom], [+Loc], traduit la locativité forte, la deuxième, dotée des traits [+Nom], [+Loc], la locativité faible. La troisième phrase (*Il est à la mode*) accuse un caractère nominal (le syntagme à

la mode étant un adjectif composé prédicatif), alors que la quatrième phrase (*Il est à cent mètres de l'arrivée*) ni correspond ni aux critères de nominalité, ni à ceux de locativité et on peut donc considérer ce type de syntagme prépositionnel comme un adverbial à degré zéro de locativité et de nominalité:

[±Nom], [±Loc]	Suite	Pronominalisation
[-Nom], [+Loc]	<i>Il est dans un café</i>	<i>Elle (y + *I') est aussi</i>
[+Nom], [+Loc]	<i>Il est en classe</i>	<i>Elle (y + I') est aussi</i>
[+Nom], [-Loc]	<i>Il est à la mode</i>	<i>Celui-ci (*y + I') est aussi</i>
[-Nom], [-Loc]	<i>Il est à cent mètres de l'arrivée</i>	<i>Elle (*y + *I') est aussi</i>

Si l'on observe maintenant le comportement syntaxique des quatre prépositions locatives simples du français (*à, en, sur, dans*) dans divers types d'emplois à valeur locative, on constate tout de suite un certain nombre de divergences que nous essaierons de ramener, dans un premier temps, à ce même cadre explicatif qu'est l'hypothèse de locativité forte vs faible. La première divergence tient au format du syntagme prépositionnel: ce n'est que *dans* pour laquelle la présence d'un déterminant est, comme on sait, obligatoire (*dans* + Dét + N):

(4) *Il se trouve dans (une + la + cette + *E) forêt*

alors que ce n'est pas le cas pour les trois autres: *en forêt, sur mer, à mi-chemin*, etc.

Dans les reprises anaphoriques, on relève également certaines anomalies syntaxiques. Observons en effet les trois phrases dans (5)-(5b):

(5) *Jean est (en prison)_i*

(5a) *Jean est (dans une prison sur la côte atlantique)_i*

(5b) *Jean est (à la prison de Fresnes)_i*

Qui peuvent toutes être anaphorisées par:

(5c) *Il y_i apprend un métier*

L'anaphorisation de (5a) et de (5b) ne pose aucun problème, les deux structures correspondent en effet à des instances de locativité

forte. Celle de (5), qui est une instance de locativité faible, s'explique, croyons-nous, par la présence du trait [+Loc].

Les prépositions *dans* et *sur* admettent, à la différence de *à* et *en*, les emplois coordonnés du type: Prép1 + *et/ou* +Prép2 + Dét + N:

- (6) *cigognes dans et sur une volière* (Toile)
- (6a) *petits animaux (...) vivant (...) dans et sur le sédiment* (Toile)
- (6b) **oiseaux en et sur la cage*
- (6c) **Les flics grouillaient à et devant la Sorbonne* (exemple emprunté à Ruwet (1982: 319))

Ce type d'emploi est à rapprocher de la possibilité de commuter deux syntagmes prépositionnels quand il s'agit d'une localisation complexe, mais homogène du point de vue topologique. Là aussi, on relève des différences entre, d'un côté, *dans* et *sur*, et, de l'autre côté, *à* et *en*:

- (7) *Les bordeaux sont dans la cave sur deux étagères*
- (7a) *Les bordeaux sont sur deux étagères dans la cave*
- (7b) *?Les bordeaux sont en cave sur deux étagères*
- (7c) *??Les bordeaux sont sur deux étagères en cave*

Les phrases (7) et (7a) sont syntaxiquement identiques, leur seule différence tient au phénomène pragmatique de saillance qui n'est pas pertinent pour notre propos. Par contre, les exemples (7b) et surtout (7c) sont jugés comme plus ou moins douteux par la plupart des personnes interrogées. Ces trois types d'emplois sont visualisés dans le tableau (A) suivant:

Construction	dans	sur	en	à
Prép+Det+N/Prép + DétØ + N	+/-	+/+	+/+	+/+
Prép1+et+Prép2+N	+	+	-	-
Prép1+et+Prép2+N / Prép2+et+Prép1+N	+/+	+/+	-/-	-/+

Si l'on raisonne maintenant en termes fonctionnels *et/ou* cognitifs mettant en jeu les critères maintes fois décrits dans la littérature (cf., entre autres, Herskovits 1986, Vandeloise 1986, Vandeloise 1993, Talmy 2000) tels que inclusion partielle ou totale de la cible

par rapport au site, la relation contenant/contenu et la position de la cible par rapport à l'axe vertical/horizontal du site, on s'aperçoit qu'il existe un type de correspondances analogues par rapport au comportement syntaxique que l'on vient de décrire, d'où la différence d'interprétation entre (8) et (8a):

- (8) Le vin est livré en tonneau(x)
 (8a) Le vin est conservé dans un tonneau métallique

(8a) décrit en effet la relation contenant/contenu, alors que dans (8), ce type de relation est secondaire, la phrase met en avant les modalités de transport (par opposition à *livré en bouteilles*, etc.)

Il est facile de relever des emplois analogues illustrant la relation d'inclusion de la cible par rapport au site comme dans (9) et (9a): (les exemples proviennent de la Toile):

- (9) *le risque d'apparition de cette maladie invalidante est majoré de 30% par rapport aux enfants vivant en milieu sain*
 (9a) *permettre à l'élève et aux adultes d'évoluer dans un milieu sain*

Dans (9), le syntagme *en milieu sain* désignent beaucoup plus le contexte sanitaire, sa valeur locative n'est que résiduelle, alors qu'elle est beaucoup plus présente dans (9a) où le syntagme *dans un milieu sain* renvoie à une certaine réalité topologique (*maison, école, etc.*). On comprend qu'on peut retrouver ce même type de dichotomie dans les emplois métaphoriques opposés à ceux utilisés au sens locatif propre:

- (10) *Il a la victoire en poche*
 (10a) *Il a une lettre dans sa poche*

La cible peut également être localisée par rapport à l'axe vertical et/ou horizontal que représente prototypiquement le site. On peut ainsi opposer:

- (11) *les bateaux en mer*
 (11a) *les enfants jouent sur la plage*
 (11b) *il veut s'installer (dans la plaine + dans une île)*

Nous allons analyser un peu plus loin l'exemple (11b) qui présente une certaine anomalie. Le tableau (B) ci-dessous résume la situation:

Relation cible/site	dans	sur	en	à
Inclusion partielle/totale de la cible par rapport au site	+	-	+	+
Relation contenant/contenu	+	-	+	-
Situation de la cible par rapport à l'axe vertical/horizontal	+/+	+/+	-/+	+/+

On voit apparaître clairement un certain nombre de correspondances entre la syntaxe et la cognition: comme l'ont montré, à titre d'exemple, les phrases (8) et (8a). Les différences d'ordre syntaxique entre *en tonneau* et *dans un tonneau*, par exemple la possibilité ou l'impossibilité d'une reprise anaphorique du type *il est fait en bois de chêne*, sont liées, en l'occurrence, à la qualité de la relation fonctionnelle contenu/contenant. Il n'est donc pas surprenant de constater que toute métaphorisation a des répercussions sur la syntaxe. Une locution métaphorique comme (12):

(12) *I est dans le pétrin*

interdit évidemment toute reprise anaphorique du type:

(12a) **Elle y est aussi*/**Elle est dedans aussi*

Etant donné que la métaphorisation ou le figement sont cependant des valeurs scalaires, on observe certains cas où les reprises anaphoriques s'avèrent plus ou moins possibles. On a ainsi d'un côté (13):

(13) *Les enfants sont à la neige* (= ils participent à un stage de ski)

qui semble admettre des enchaînements anaphoriques comme:

(13a) *Ils sont à la (neige)_i. Elle_i est très poudreuse cette année.*

et, de l'autre côté, (14):

(14) *Les enfants sont à (l'université)_i.*

(14a) ??*Elle_i est très ancienne.*

(14b) ??*Elle_i date du XIV^e siècle/ ??Sa_i fondation date du XIV^e siècle*

Il en est de même pour les emplois métonymiques du type:

(15) *Paul travaille dans la chaussure* (= secteur, industrie)

(15a) ?*Il est dedans depuis dix ans*

où la reprise par *dedans* semble plus acceptable pour les professionnels du métier, pour qui la métonymie est apparemment beaucoup moins pertinente.

Formulons maintenant, à partir de ce qu'on vient de voir, trois hypothèses de base:

- a) il existe une corrélation entre l'aspect syntaxique et l'aspect cognitif, ce qui veut dire que toute irrégularité syntaxique entraîne une anomalie sur le plan cognitif. La régularité syntaxique est prototypiquement corrélée à ce qu'on vient de définir comme locativité forte, par contre la locativité dite faible correspond à des emplois moins prototypiques d'une préposition. Il nous semble que cette hypothèse tient même si l'on accepte une vision plutôt géométrique de la cognition spatiale et que l'on raisonne en fonction d'un sens idéal d'une préposition locative comme le fait par exemple Herskovits (1986: 39ff): "The ideal meaning of a preposition is a geometrical idea, from which all uses of that preposition derive by means of various adaptations and shifts.", ou au contraire si l'on adopte une approche plus fonctionnelle (cf. Vandeloise 1986, Vandeloise 1993).
- b) la locativité est une valeur scalaire, il existe un continuum entre les emplois prototypiques à locativité forte et ceux à locativité faible. Si le syntaxique et le cognitif sont corrélés, la prototypicité peut être testée à l'aide d'outils syntaxiques. Si l'on constate par exemple que l'anaphorisation de la phrase (11a), reprise ici comme (16), pose quelques problèmes d'acceptabilité:

(16) *Les enfants jouent sur la plage. ??Ils jouent dessus.*

on peut en déduire qu'il s'agit d'un niveau "affaibli" de locativité. En effet, la suite *jouer sur la plage* est fortement concurrencée par *jouer à la plage* (sur la Toile, on relève quelque 5200 occurrences pour *jouer sur la plage* contre environ 3500 pour

jouer à la plage) – la préposition à représentant, sur l'échelle de locativité, des valeurs beaucoup moins prototypiques.

- c) Les quatre prépositions locatives primaires *dans*, *sur*, *à* et *en* constituent, tant sur le plan syntaxique que sur le plan cognitif/conceptuel, non pas des paires d'opposition binaires, mais un ensemble structuré selon le modèle wittgensteinien de ressemblance de famille. Là, nous rejoignons Jackendoff 1996: 5: "Conceptual structure is also not entirely digital, in that some conceptual features and some interactions among features have continuous (i.e., analog) characteristics that permit stereotype and family resemblance effects to be formulated.", cf. aussi Jackendoff (2002: 354). La structuration des emplois locatifs selon ce principe permet en effet de mieux expliquer la concurrence de certains syntagmes prépositionnels (*dans une île* vs *sur une île*, *en ville* vs *dans la ville*, etc.).

Passons maintenant à la perspective diachronique pour voir l'évolution des emplois locatifs concernés depuis le français préclassique, donc depuis le XVI^e siècle. On admet depuis Gougenheim 1970 que la préposition *dans* commence à s'imposer massivement en français dès les années cinquante du XVI^e siècle et notamment grâce à Ronsard. Comme le fait remarquer Gougenheim (1970: 54): "La naissance de *dans* à côté de *en* au XVI^e siècle a permis à la langue moderne de différencier ces deux outils grammaticaux. *Dans* a pris la valeur spatiale et matérielle de l'ancien *en*. *En* a exprimé l'intégration, la prise de possession par le dedans (...)" . Tout au long de la seconde moitié du XVI^e siècle et encore abondamment au siècle suivant, on trouve de nombreux exemples d'emplois qu'on peut considérer comme déviants par rapport à l'usage contemporain:

- (16) *Quand les vignes gellent en mon village (...)*, Montaigne (Frantext)
- (17) (...) *aucuns Espagnols prindrent terre en une contrée fertile et plaisante (...)* – (ibid.)
- (18) (...) *le Pape, representant Dieu en terre* (ibid.)
- (19) (...) *il monta en la chambre de la princesse où, s'asseyant en une chaise prés du lict, il se mit fort tendrement sur les enquestes de son indisposition.* Marie de Gournay, 1626 (Frantext).

Si l'on se limite, sur le plan purement descriptif, à la seule opposition entre *dans* et *en*, on observe que leur évolution suit la trajectoire habituelle du changement linguistique maintes fois décrite dans la littérature (cf., à titre d'exemple, Brinton-Traugott (2005: 26): $A > A \sim B > B$ où la forme initiale *A* n'est définitivement remplacée par la nouvelle forme *B* qu'après une période plus ou moins longue de coexistence des deux formes. Citons, parmi d'innombrables exemples, ces deux vers de Corneille:

(20) *Mon sceptre dans ma main par la tienne affermi* (Le Cid, 1636, Frantext)

(21) *La bague de la Reyne est encore en ma main* (Don Sanche d'Arragon, 1650, Frantext)

Il est aussi intéressant de voir comment cette problématique a été abordée dans les grammaires de l'époque. Parmi les nombreux ouvrages du genre parus entre la moitié du XVI^e siècle et la fin du XVIII^e siècle, nous en retiendrons trois qui nous paraissent particulièrement pertinents pour notre propos. Le premier est le *Tretté de la grammeze françoëze* (1550) de Louis Meigret¹ dont nous citerons le passage suivant (p. 168) que nous transcrivons en orthographe moderne pour une plus grande commodité de lecture:

Nous disons bien *on le mène en prison*, sans article: nous ne dirons pas toutefois *on descend ce vin en cave*, pour *en la cave*, Il faut aussi entendre que la locution sans article est plus générale, que celle qui est par l'article (...) de sorte que quand on dit *il est en prison*, ou *en cave*, nous comprenons qu'il est en quelque prison, ou en cave que ce soit.

Ce témoignage est particulièrement éclairant, nous semble-t-il, car il représente une première ébauche d'une théorisation des emplois locatifs vers le milieu du XVI^e siècle. Meigret n'ignore pas, il est vrai, la préposition *dans*, mais elle n'apparaît, au côté de *dedans*, que comme une simple variante, comme le montrent ses exemples d'emplois partitifs (ibid. p. 164):

(22) (...) *j'ai mis du brochet, de la carpe dedans cet étang* (...)

(23) (...) *j'ai mis en mes étables du mouton, du boeuf, de la vache* (...)

¹ Nous citons d'après l'édition de W. Foerster. 1888. Heilbronn: Verlag von Gebr. Henninger

La principale opposition est pour lui celle du syntagme à déterminant zéro (*en cave*) et du syntagme déterminé (*en la cave*) et elle correspond à l'opposition entre un emploi locatif référentiel, autrement dit un emploi à locativité forte, et un emploi déréférentialisé, proche d'une valeur attributive à locativité faible. Une telle configuration d'emplois locatifs correspond d'ailleurs à celle que l'on connaît aujourd'hui dans d'autres langues romanes, cf. pour l'italien (24) et pour l'espagnol (25):

(24) *Pietro a ormai la vittoria in tasca vs Paolo a le chiave nella tasca*

(25) *valores cotizados en bolsa vs valores cotizados en la bolsa de Londres*

La *Grammaire françoise rapportee au langage du temps*, (1656) d'Antoine Oudin² dresse fidèlement un état des lieux concernant l'évolution des prépositions et des emplois locatifs vers le milieu du XVII^e siècle (p. 312):

En se met aussi lors qu'on parle de chose non materielle contenant: par exemple, il sçait en son ame, en son esprit, en sa conscience, &c. & si on attribuë à ces choses spirituelles une qualité des materielles, alors on se peut servir de dans: v.g. Cela est bien avant dans son ame: il est fort avant dans mon esprit. Ainsi ces choses qui contiennent materiellement, il est dans sa chambre, il est dans son logis, ce qui se dit plus proprement qu'en sa chambre, & en son logis. Et tout cecy se doit entendre du verbe estre: car autrement on peut dire: il est logé en ce lieu là: il luy a dressé cette lettre là en son logis, &c.

Même si son enseignement ne cadre pas toujours avec l'usage de l'époque, il s'avère, dans les grands traits du moins, conforme à la tendance générale. En effet, la préposition *dans* semble occuper désormais une position dominante quant aux emplois à locativité forte (cf. *il est dans sa chambre, il est dans son logis*), en revanche, sa principale rivale *en* ne sélectionne désormais que des compléments abstraits et – ce qui fait encore la différence par rapport à l'usage contemporain – elle peut être aussi utilisée (cf. *il est logé en ce lieu là*) dans des constructions où c'est le verbe qui est doté d'un sens intrinsèquement locatif. Une petite statistique obtenue grâce à Frantext

² Nous citons d'après l'édition de Rouen, chez Jean et David Berthelin.

confirme cette tendance. Si l'on ne considère, à titre d'exemple, que des occurrences de *habiter en* et ceux de *habiter dans*, on relève, pour la période (i) antérieure à 1700 et (ii) pour la période entre 1700 et 1800:

	< 1700	1700 – 1800
<i>habiter en</i>	30	1
<i>habiter dans</i>	5	17

Avec Condillac et ses *Principes généraux de grammaire* (1798)³, on découvre déjà l'essentiel de ce qu'est l'usage contemporain:

A désigne seulement le lieu où est une chose: *dans* le désigne avec un rapport du contenu au contenant. (...) *En* diffère de *dans*, parce que le terme qu'il indique se prend toujours d'une manière indéterminée. *J'étois en ville* signifie *je n'étois pas chez moi*; et je n'ajoute pas au mot *ville* l'adjectif *la*, parce qu'en pareil cas il n'est pas nécessaire de le déterminer. Il me suffit de faire entendre que j'étois quelque part dans la ville. Si, au contraire, je veux dire que je n'étois pas sorti hors des portes, je détermine ce mot, et je dis: *j'étois dans la ville*.

Ces définitions, étonnamment modernes dans leur terminologie („rapport du contenu au contenant“) délimitent clairement les sphères d'emploi locatif des trois prépositions *dans*, *en* et *à* et ne diffèrent guère de la position des grammaires modernes. On peut en conclure que le processus qui mène de l'opposition *en ville* vs *en la ville* (Meigret) à celle de *en ville* vs *dans la ville* semble avoir été achevé vers cette fin du XVIII^e siècle.

Revenons maintenant au type d'emplois locatifs représentés par (11b), c'est-à-dire les syntagmes prépositionnels dans lesquels *dans* sélectionne des portions d'espace „bidimensionnel“ tels que *plaine*, *désert*, *île*. Le test d'anaphorisation par *dedans* révèle une anomalie par rapport au statut de *dans* comme préposition prototypique à locativité forte:

(26) *Il vit depuis des années dans une île déserte. ??Il se plaît beaucoup dedans.*

³ Nous citons d'après l'édition disponible sur le site: <http://gallica.bnf.fr>

Sur le plan diachronique, on trouve d'abord le syntagme *en une isle*, comme par exemple chez Rabelais (1542, Frantext):

(27) (...) *les envoya vivre en une isle là auprès* (...)

Le syntagme *dans une isle* n'apparaît dans Frantext qu'en 1602 (Fauchet), alors que *sur une isle* est nettement plus tardif: 1681 (Bosuet) ou même 1794 (Mme de Staël) pour la graphie moderne (*sur une île*). Pour le XX^e siècle, on relève 134 occurrences pour *dans une île* contre 70 pour *sur une île*. Le tableau est tout-à-fait analogue pour *dans une plaine*, attesté pour la première fois en 1636 (Mairet). Ainsi, des emplois comme (28):

(28) *Dans une plaine extrêmement verte nous crûmes voir une eminence couverte de neige* (Guez de Balzac, 1657, Frantext)

sont nettement antérieurs à ceux de *sur une plaine* comme (29) qui n'est attesté pour la première fois qu'en 1770:

(29) *Vois de longs bataillons rangés sur une plaine* (...) (J. Delille, Frantext)

Voici encore un tableau documentant le nombre d'occurrences des deux syntagmes *en un desert* et *dans un desert* dans trois périodes de cinquante ans. Cette petite statistique rejoint complètement celles que l'on vient de relever pour les compléments prépositionnels *plaine* et *île*.

Période	1550-1600	1600-1650	1650-1700
<i>En un desert</i>	5	7	2
<i>Dans un desert</i>	1	5	15

Ajoutons que les données que l'on peut obtenir grâce à Internet permettent de voir une évolution en cours dans le sens d'une certaine „standardisation“ d'emplois locatifs: la préposition *dans* étant confinées de plus en plus dans des emplois plus ou moins prototypiques (relation contenant/contenu, entre autres) et *sur* tend à la remplacer dans des emplois dans lesquels la relation cible/site est sensible à l'axe horizontal du site. Ainsi, on relève quelque 147 000 occurrences de *dans une île* contre 975 000 pour *sur une île*, 56 000 occurrences

pour *dans une plaine* contre 18 000 pour *sur une plaine*, 106 000 pour *dans un désert* contre 535 pour *sur un désert*. Les trois types de syntagmes évoluent donc à des vitesses différentes, le phénomène de fréquence lexicale y est certainement pour quelque chose.

Ces témoignages et ces données statistiques à l'appui, nous formulons, l'hypothèse suivante: l'opposition entre ce que nous appelons la locativité forte et la locativité faible a connu, depuis le XVI^e siècle une évolution en trois étapes. La première correspond grosso modo à la dichotomie relevée par Meigret et que l'on a vue un peu plus haut. L'étape moderne est caractérisée par une séparation de plus en plus nette entre les emplois à locativité faible (prototypiquement *en* + DétØ + N) et ceux à locativité forte, tout cela avec un parallélisme dans la conceptualisation de l'espace par la langue. L'étape intermédiaire reste par définition à mi-chemin entre ces deux pôles, mais il nous semble que l'irruption de *dans* a entraîné en français préclassique une restructuration générale des emplois prépositionnels locatifs: la préposition *dans* tendait en effet à être employée dans des syntagmes exprimant la locativité forte, quelle que soit la relation conceptuelle entre la cible et le site, autrement dit la préposition *dans* a d'abord monopolisé l'expression de la locativité forte, référentielle. A preuve, on peut citer d'innombrables emplois de *dans* qui nous paraissent aujourd'hui déviants par rapport au standard contemporain. Un seul exemple à titre d'illustration (Baro, 1628, Frantext):

(30) *j'ay nulle blessure dans la main*

Schématiquement, on peut visualiser cela dans un petit tableau récapitulatif:

Locativité forte – faible	Locativité forte – faible	Locativité forte – faible
Etape 1	Etape 2	Etape 3
<i>en la cave – en cave</i>	<i>dans</i> + tout N locatif – <i>en la cave/en cave</i>	<i>dans (la+une) cave – en cave</i>

Le rapprochement entre la diachronie et la problématique de l'apprentissage de la langue par l'enfant a été souvent fait dans la littérature, cf., entre autres DeGraff (1999: 474ff). Dans ce cas précis,

nous sommes également tentés de dresser ce type de parallélisme en nous référant à Martinot 1999: 115:

Le fonctionnement des premières prépositions présente trois particularités: la surreprésentation de l'une d'elles, *dans*, comme si elle jouait la fonction prototypique de toutes les prépositions, c'est ce phénomène que l'on désigne par *surgénéralisation*; la disparité statistique qui en résulte et ne correspond pas à celle du français standard; mais surtout un certain décalage, qui peut passer inaperçu, entre la préposition employée par l'enfant et celle qui serait plus naturelle dans la langue cible.

Il serait certainement prématuré de vouloir tirer des conclusions hâtives de cette coïncidence fort intéressante entre la surreprésentation de *dans* comme prototypiquement locative sur le plan diachronique et la surreprésentation de la même préposition à un certain niveau du processus de l'acquisition de la langue, mais ceci constitue sans doute une matière fort stimulante pour de nouvelles recherches.

Références

- Berthonneau, A. M. 1999. A propos de *dedans* et de ses relations avec *dans*. *Revue de Sémantique et Pragmatique*. **6**: 13-41.
- Brinton L. J.; Traugott, Closs E. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- Croft, W.; Cruse, D. A. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- DeGraff, M. 1999. Creolization, Language Change, and Language Acquisition: An Epilogue. In: M. DeGraff (Ed.). *Language Creation and Language Change*. Cambridge MA: The MIT Press.
- Fortis, J.-M. 2004. L'espace en linguistique cognitive: problèmes en suspens. *Histoire-Épistémologie-Langage*. **26(1)**: 43-88.
- Gougenheim, G. 1970. *Etudes de grammaire et de vocabulaire français*. Paris: Picard.
- Herskovits, A. 1986. *Language and Spatial Cognition*. Cambridge: Cambridge University Press.
- Homma, Y. 2006. Réflexions sur l'analyse de l'identité d'une préposition: le cas de «dans». *Modèles linguistiques*. **XXVII(2/54)**: 25-36.
- Jackendoff, R. 1996. The architecture of the Linguistic-Spatial Interface. In: Paul Bloom *et al.* (Eds.). *Language and Space*. Cambridge MA: The MIT Press, 1-30.

- Jackendoff, R. 2002. *Foundations of Language*. Oxford: Oxford University Press.
- Kleiber, G. 1990. *La sémantique du prototype*. Paris: PUF.
- Langacker, R. 2008. *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press.
- Martinot, C. 1999. Premières prépositions: *Dans* ou à ?. *Revue de Sémantique et Pragmatique*. **6**: 115-132.
- Noailly, M. 1999. *L'adjectif en français*. Paris-Gap: Ophrys.
- Ruwet, N. 1982. *Grammaire des insultes et autres études*. Paris: Seuil.
- Spillebout, G. 1985. *Grammaire de la langue française du XVIIIe siècle*. Paris: Picard.
- Talmy, L. 2000. *Toward a Cognitive semantics, vol I. Concept Structuring systems*. Cambridge MA: The MIT Press.
- Tremblay, M.; Dupuis, F.; Dufresne, M. 2003. Les prépositions dans l'histoire du français: Transitivité, grammaticalisation et lexicalisation. *Verbum*. **XXV(4)**: 549-562.
- Vandeloise, C. 1986. *L'Espace en français*. Paris: Le Seuil.
- Vandeloise, C. 1993. Les analyses de la préposition *dans*: faits linguistiques et effets méthodologiques. *Lexique* **11**: 15-40.

Recordando o Professor Jaromír Tláškal

Jaroslava Jindrová
Faculdade de Filosofia, Universidade Karlova, Praga
(República Checa)

A 20 de Maio de 2008 faleceu em Praga Jaromír Tláškal, professor associado do Instituto de Línguas Românicas da Faculdade de Letras da Universidade Carolina de Praga, com apenas 59 anos de idade. O Professor Jaromír Tláškal dedicou grande parte da sua vida aos estudos de linguística românica e geral, foi membro do Conselho Redactorial da Revista de Filologia Moderna editada em Praga e docente da Secção Portuguesa e Francesa. Apesar de nos últimos dois anos ter lutado com uma doença grave, todos os seus colegas e estudantes acreditaram que ia voltar a trabalhar. A morte do colega amado e respeitado representa um duro golpe para todos os que o conheceram.

Jaromír Tláškal nasceu a 14 de Fevereiro de 1949 em Praga. Formou-se na Faculdade de Letras da Universidade Carolina, escolheu filologia românica como especialização, mais concretamente portuguesa e francesa. Em 1972 defendeu a tese de licenciatura dedicada à problemática do conjuntivo nas frases subordinadas no português moderno e logo depois começou a trabalhar na Escola Estatal de Línguas como professor de língua portuguesa e francesa. Ao mesmo tempo continuou nos estudos linguísticos e poucos anos mais tarde apresentou a dissertação de doutoramento. Devido à sua convicção pessoal e política não pôde seguir a carreira científica no ambiente académico da Faculdade de Letras; porém, nunca deixou de acompanhar as novas correntes e tendências na linguística. Ainda como professor de língua prática começou a escrever artigos científicos publicados tanto no seu país de origem como no estrangeiro. Na metade dos anos oitenta do século passado concentrou as suas forças numa monografia dedicada à fonética e fonologia do português europeu. Pouco tempo depois da Revolução de Veludo em 1989, foi convidado para trabalhar no Instituto de Línguas Românicas da Faculdade de Letras em Praga, nos primeiros anos como assistente e mais tarde como professor associado. Embora a sua especialização fosse linguística francesa (morfologia e sintaxe), Jaromír Tláškal dava

igualmente aulas de fonética, fonologia e lexicologia aos alunos de Estudos Portugueses. Como sucessor do professor Zdeněk Hampl representou a filologia portuguesa também no estrangeiro e nessa qualidade foi convidado para trabalhar na equipa dos autores da enciclopédia monumental de línguas românicas *Lexikon der romanistischen Linguistik (I-VIII)*, publicada em Tübingen sob a orientação do professor Günter Holtus¹, para a qual escreveu a parte de morfologia flexional portuguesa. Continuou também os estudos de fonética e fonologia do português, publicou vários artigos dedicados a este tema e em 1990 publicou a primeira versão do manual de Fonética e Fonologia do Português Europeu Contemporâneo destinado, como material de apoio, aos estudantes universitários de português. Este manual, alargado e actualizado, foi publicado em 2006 como monografia².

Porém, a obra mais importante do professor Jaromír Tláškal é, sem dúvida, a monografia escrita em francês *La transposition en français contemporain – contribution à l'étude du problème*³, trabalho complexo apresentado como dissertação de habilitação para o grau de professor associado. A monografia é prova da erudição do autor, sucessor das melhores tradições da escola linguística estrutural praguense. Jaromír Tláškal foi igualmente tradutor e editor da obra de um dos representantes mais destacados da linha do Círculo Linguístico de Praga, professor Josef Vachek⁴.

Todavia, o professor Tláškal não foi só homem de ciência. Como membro do corpo docente da Faculdade de Letras orientou teses de licenciatura e teses de doutoramento, foi membro de várias comissões e participou na elaboração de materiais para o ensino secundário e superior de línguas românicas. Como professor teve grande prestígio entre os colegas e entre os estudantes.

¹ Portugiesisch: Flexionslehre. In Holtus, G. – Metzeltin, M. – Schmitt, Ch. (Eds.). *Lexikon der Romanistischen Linguistik*. Band VI, 2. Galegisch, Portugiesisch. Tübingen: Niemeyer, 1994, p. 160-172.

² *Fonetika a fonologie současné evropské portugalštiny*. Praha: Karolinum, 2006.

³ *La transposition en français contemporain – contribution à l'étude du problème*. Praga: AUC, 2000.

⁴ Vachek, J. *Lingvistický slovník pražské školy*. Praga: Karolinum, 2005 (tradução e edição em colaboração com V. Petkevič).

O Instituto de Línguas Românicas perdeu um dos seus melhores professores, cujo nome sempre será honrado na sua memória. No entanto, em primeiro lugar, perdemos um bom amigo – compreensivo, honesto, com grande prestígio moral e ético, sempre disponível para ajudar todos. O seu desaparecimento deixou um grande vazio que dificilmente será preenchido. Nos nossos corações Jaromír continuará presente para sempre.

Bibliografia das obras mais importantes do Doc. PhDr. Jaromír Tláškal

- L'emploi du subjonctif dans les propositions subordonnées de la langue portugaise contemporaine (contribution au problème). *Philologica Pragensia* 19/2, 1976, p. 53-72.
- Contribuição à semântica do subjuntivo português nas proposições subordinadas, *Ibero-Americana Pragensia* XI, 1977, p. 236-238. (1978a)
- Corrélation entre les formes du futur et la catégorie de la locution en portugais. *Linguistics: special issue*, 1978, p. 173-184.
- Remarques sur la catégorie de la personne en portugais. *Folia Linguistica* XII, 1978, p. 367-383.
- Remarques sur les voyelles nasales en portugais. *ZPSK*, 33, 1980, p. 562-570.
- Dynamique du système phonologique des voyelles orales en portugais du Brésil. *Phonetica Pragensia* VI, 1980, p. 69-79.
- Remarques sur l'accent tonique en portugais du Brésil. *ZPSK (Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationslehre)* 34, 1981, p. 341-351.
- Dynamique du système du portugais parlé au Brésil (contribution au problème). *Boletim de Filologia* XXVI, 1981, p. 29-49.
- Bemerkungen zu Tempora und Modi im Portugiesischen. In Schmidt-Radefeldt, J. (ed.). *Portugiesische Sprachwissenschaft*. Gunter Narr Verlag Tübingen, 1983, p. 105-118.
- Observações sobre tempos e modos em português. Coleção linguística. Volume I, *Estudos de Linguística Portuguesa*, Coimbra Editora 1984, p. 237-255.
- Portugiesisch: Flexionslehre“. In: Holtus, G. – Metzeltin, M. – Schmitt, Ch., ed.: *Lexikon der Romanistischen Linguistik*. Band VI, 2. Galegisch, Portugiesisch. Tübingen, 1994, p. 160-172.
- Remarques sur l'accord entre la forme et la fonction. *Linguistica Pragensia* 1/96, 1996, p. 35-45.
- Contribuição a uma característica funcional dos relativos no português falado. *Revista Internacional de Língua Portuguesa* 14, 1996, p. 148-153.
- Notas sobre o relativo *QUE* e a valência verbal no português falado. *Ibero-Americana Pragensia* XXXI, 1997, p. 29-39.

- La transposition en français contemporain – contribution à l'étude du problème.*
Praga: Karolinum, 2000.
- Remarques sur la transposition en français. In Radimská, J. (ed.). *Approche du texte, aspects méthodologiques en linguistique et en littérature. Opera romanica 2.* České Budějovice: Editio Universitatis Bohemiae Meridionalis, 2001, p. 127-153.
- Alguns aspectos do infinitivo flexionado português: a sua capacidade (potencialidade) transposicional. In Čermák, P. – Tláškal, J. (eds.). *Las lenguas románicas: su unidad y diversidad. Homenaje al Profesor Bohumil Zavadil con ocasión de su 65º cumpleaños.* Praga, 2005, p. 227-239.
- Fonetika a fonologie současné evropské portugalštiny* (Fonética e Fonologia do Português Europeu Contemporâneo). Praga: Karolinum, 2006.