

# Unidades monoverbais e pluriverbais, diacronia e tratamento informático no *corpus* metalinguístico do português quinhentista

Maria Helena Paiva

mhelenapaiva@netcabo.pt

Faculdade de Letras da Universidade do Porto

Centro de Linguística da Universidade do Porto (Portugal)

RESUMO. Quando se aborda o estudo linguístico de épocas em que a variação atinge a *palavra*, é indispensável conceituar essa unidade. Considerando que a questão ganha em ser reexaminada no âmbito de uma língua determinada, neste caso o português, na primeira parte do artigo analisam-se os critérios habitualmente retidos ou considerados relevantes na definição de ‘palavra’: menciona-se o nível prosódico como não descurável mas também não susceptível de fornecer critérios decisivos (1.1), discutem-se as definições de Bloomfield (1.2) e de Meillet (1.3), avalia-se a eficácia do critério de impossibilidade de inserção (1.4), tendo em conta o futuro e o condicional (1.4.1) e os advérbios em *–mente* (1.4.2), depois do que passa a interrogar-se a diacronia (1.5). Considera-se que, na relação entre o português e o latim a nível do vocábulo, são discerníveis três derivas: a permanência do vocábulo lexical (1.5.1) que, no plano fonológico, se caracteriza pela impermeabilidade ao contexto, o que atesta a mobilidade da palavra (1.5.1.1); a consolidação da alomorfa contextual, que afecta o, artigo e pronome e outros morfemas dependentes (1.5.2) e a caducidade do vocábulo funcional (1.5.3), em parte substituído por unidades pluriverbais, em que frequentemente se dão fenómenos de gramaticalização e em que se distinguem gradações de diversos tipos (1.5.3.1). Na segunda parte indaga-se sobre o modo como se projectam os conteúdos definidos anteriormente nos textos dos gramáticos e ortógrafos quinhentistas: depois da caracterização do *corpus* e da síntese do tratamento informático a que este foi submetido (2.1), aborda-se o delineamento do percurso que vai da palavra gráfica à palavra propriamente dita, explicitando os princípios metodológicos aplicados (2.2). São passados em revista diversos tipos de delimitações (2.2.1): rupturas (2.2.1.1), junções simples (2.2.1.2), junções com elisão ou crase (2.2.1.3), a que se segue a descrição da introdução e do uso do apóstrofo (2.2.1.4) e do traço de união (2.2.1.5). Na terceira parte são definidas identidades e diferenças entre palavra e unidades pluriverbais; começa-se por uma retrospectiva crítica (3.1) sobre a adequação dos diferentes conceitos de palavra às especificidades da sua representação gráfica e às soluções tentadas pelos gramáticos ao longo do período e/ou fixadas na nor-

ma gráfica. As unidades pluriverbais (3.2), são tratadas numa primeira vertente, em que se estuda a trajectória que consiste na transformação de unidades pluriverbais em palavras, e se distinguem diferentes graus de coesão interna (3.2.1); e numa segunda, em que se descreve a trajectória inversa, da palavra à unidade pluriverbal, mediante a selecção dos aspectos mais directamente relacionados com o tratamento informático do *corpus* (3.2.2).

PALAVRAS-CHAVE. Palavra - Unidades pluriverbais - Tratamento informático - Grafia quinhentista - História do Português - História da Gramática.

ABSTRACT. When focusing on the linguistic study of the periods when variation affects the word, it is essential to conceptualize this unit. Considering that the subject gains in being re-examined in the scope of a given language (in this case, Portuguese), in the first part of the article I have analysed the criteria that are usually retained or that are considered relevant to define 'word': the prosodic level is mentioned as not being discarded but also not being able to provide decisive criteria (1.1); the definitions of Bloomfield (1.2) and Meillet (1.3) are discussed; the effectiveness of the criteria of impossibility of insertion is evaluated (1.4), the future and the conditional are taken into account (1.4.1) and also the *-mente* adverbs (1.4.2). Then, the diachrony is questioned (1.5). It is considered that in the relationship between Portuguese and Latin regarding the vocable, three drifts can be observed: the permanence of the lexical vocable (1.5.1) which, at the phonological level, is impervious to context, which certifies the mobility of the word (1.5.1.1); the consolidation of the contextual allomorphy of *o*, article and pronoun, and of other dependent morphemes (1.5.2), and the lapse of the functional words, partly replaced by multi-word expressions in which grammaticalization frequently occurs and in which different kinds of gradations are noticed. In the second part, it is inquired into the projection of the contents previously defined on the texts of the grammarians and the orthographers of the 16<sup>th</sup> century: after the description of the *corpus* and the synthesis of the computer processing (2.1), the outline of the path from the graphic word to the word itself is drawn and the methodological principles which have been applied were exposed (2.2). Different kinds of delimitations are reviewed (2.2.1): breaks (2.2.1.1); simple unions (2.2.1.2); unions with elision and contraction (2.2.1.3), followed by the description of the introduction and the use of the apostrophe (2.2.1.4) and the hyphen (2.2.1.5). In the third part, the identities and the differences between the word and the multi-word expressions are defined; it includes a critical retrospection (3.1) about the adequacy of the different concepts of 'word' to the specifications of its graphic representation and the solutions that were tried by grammarians along this period and / or that were fixed in the writing norm. The multi-word expressions (3.2) are addressed in a first outline in which I study the trajectory of transformation of the multi-word expressions into words and different degrees of cohesion are established (3.2.1); in a second outline, the inverse trajectory, from the word to the multi-word expressions is described, by means of selection of the aspects more closely connected with the computer processing (3.2.2).

KEY-WORDS. Word - Multi-word expressions - Computer processing - XVIth century writing - History of Portuguese - History of Grammar.

## O – Introdução

A palavra gráfica, definida como um grafema ou uma sequência de grafemas delimitados por espaços, constitui habitualmente a unidade mínima do texto e do tratamento informático nas línguas actuais, fortemente normalizadas, bem como o ponto de partida para a conceituação de “unidades pluriverbais”; mas é frequente a falta de indagação sobre o que de propriamente linguístico e/ou de convencional existe no que continua a designar-se por *palavra*, realidade tanto mais difícil de definir quanto mais fortemente condicionada pelo propósito de encontrar traços constantes na pluralidade das línguas, apesar das diferentes tipologias, do variável peso da história e dos equilíbrios que em cada língua se definem.

Sendo assim, restringiremos a problemática inerente a unidades mono- e pluriverbais a um quadro concreto preciso, directamente observável, de que emergem, através da própria variação e do delineamento de uma deriva específica, coordenadas gerais não descuráveis na explicitação dos conceitos em estudo; a área de pesquisa fica circunscrita a uma língua (o português), a um determinado tipo de discurso (os textos metalinguísticos) e a um determinado período da história (compreendido entre 1536 e 1606)<sup>1</sup>.

De facto, a identificação de “palavra” mediante critérios linguísticos é indispensável para períodos, como é ainda o século XVI, em que a variação atinge a delimitação desse segmento da cadeia fónica<sup>2</sup> e em que, relativamente a “unidades pluriverbais”, as práticas oscilam entre a união, que as transforma numa palavra gráfica<sup>3</sup>, e a conservação da separação que mantém como unidades pluriverbais seqüências hoje uniformemente grafadas como uma palavra<sup>4</sup>, havendo,

---

<sup>1</sup> Como ponto de partida para este artigo, cf. Paiva (2002).

<sup>2</sup> Assim, reportando-me ao *corpus* descrito no *Apêndice* e usando, nesta nota e nas duas seguintes, [ / ] para separar e [ - ] para reunir as formas ocorrentes, de acordo com a grafia actual: *a/língua*, *de/dentro*, *D/alentejo*, *antr/elles*, *desd/então*, *des/que*, *que/trazem*, *se/faz*, *principal/parte* ; *a-proveitar*, *a-té*, *a-quelle*, *na-quella*, *emtam*, *em-comendando*, *de-termina*. Ter-se-á em conta que nas transcrições dos textos do *corpus* o itálico é sempre da minha responsabilidade.

<sup>3</sup> Como *antre/dour/a/minho*, *o/qual*, *cada/hum*, *assi/como*.

<sup>4</sup> Por exemplo: *com-tudo*, *em-quanto*, *por-tanto*.

na opção pela junção ou pela disjunção, tendências suficientemente definidas para suscitarem reflexão.

Como admitia A. Martinet, “Un terme scientifique naît avec sa définition”, condição com que é incompatível a explicitação precisa do conteúdo de “mot”, “terme emprunté au vocabulaire courant.” (Martinet (dir.) 1969: s.v. “Mot”). Análogas reservas foram expressas por B. Pottier: “On peut se demander [...] si le terme de *mot* est bien utile dans le domaine linguistique, s’il ne correspond pas uniquement à une réalité *orthographique*, phénomène secondaire par rapport au caractère oral du langage (Pottier (dir.) 1973: s.v. “Mot”)⁵.

Não admira por isso que se dê a rarefacção da produção linguística sobre a conceituação de “palavra”, ao mesmo tempo que a preocupação de rigor se transfere para termos que delimitam unidades segundo ângulos de visão sistematizantes e pré-definidos, específicos de diferentes autores ou escolas, mas que só parcialmente recobrem o conteúdo do termo ou o ultrapassam (como *monema*, *lexema*, *lexia simples*, *gramema* e *sintema*, *lexia composta*, *lexia complexa*).

## 1 – Definição de palavra

### 1.1 – A palavra prosódica

É sabido que não há coincidência entre a palavra propriamente dita e a palavra prosódica, caracterizada por “um único acento principal”: “existem palavras morfológicas que apresentam mais de um acento e, inversamente, existem palavras morfológicas desprovidas de acento” (Mateus, Brito, Duarte, Faria, Frota, Matos, Oliveira, Vigário & Villalva 2003: 1061-1062).

Apesar disso, e apesar de a descrição anterior se referir ao português contemporâneo, a linguística histórica não pode alhear-se do nível prosódico, quer na conceituação de palavra (ver 1.4.1 e 1.4.2), quer na interpretação de particularidades gráficas como

---

<sup>5</sup> Na mesma linha, uma das posições mais peremptórias é a de Mounin (1974), que inicia a definição de “mot” como “Unité significative empirique de la grammaire traditionnelle” e termina considerando que “Le mot n’est pas une réalité de linguistique générale”.

indícios de atonicidade (ver 2.2.1.2) e na avaliação da amplitude de fenómenos de elisão e crase (ver 2.2.1.3).

### 1.2 – A definição de Bloomfield

De todas as definições, a mais citada continua a ser a de Bloomfield (1933)<sup>6</sup>, apesar da artificialidade do teste de produção de frases monoverbais que, além de contextos metalinguísticos, se ajusta apenas às interjeições, a situações específicas a que correspondem usos codificados como vocativos ou imperativos e apesar de ficarem excluídas palavras funcionais como preposições e conjunções e, em geral, todas as palavras átonas. Por isso, Câmara (1985: 37) define “vocábulo fonológico” como correspondendo “no plano mórfico à ‘forma livre’ de Bloomfield”.

### 1.3 – A definição de Meillet

Antes de Bloomfield, Meillet, ao expor os critérios a que deve obedecer, no âmbito da gramática comparativa, a determinação de etimologias, considerava que deve ter-se em conta que “Un mot est défini par l’association d’un sens donné à un ensemble donné de sons susceptible d’un emploi grammatical donné.” (Meillet 1921: 30).

Embora o propósito de Meillet não fosse a definição de “palavra”, geralmente é como tal que a frase é citada, dada a relevância dos traços de identificação que reúne: a palavra é uma unidade codificada e como tal é susceptível de ser transmitida através de gerações e de espaço a espaço (é o que torna possível a existência da etimologia como disciplina científica); essa unidade, além de ser, em termos saussureanos, um signo linguístico composto de significante e de significado, pauta-se por regras gramaticais determinadas.

Mas estes traços são comuns a afixos, quer lexicais, quer gramaticais, o que exige que sejam definidos critérios complementares que procuraremos analisar na sua adequação ao português.

---

<sup>6</sup> “A free form which is not a phrase, is a *word*. A word, then, is a free form which does not consist entirely of (two or more) lesser free forms; in brief, a word is a *minimum free form*.” (Bloomfield 1933: 178).

#### 1.4 – O critério de inseparabilidade ou impossibilidade de inserção

Flexionada ou não, primitiva ou derivada, a palavra caracteriza-se pelo grau máximo de coesão interna: não é possível incluir novos elementos no seu interior, motivo por que o critério de impossibilidade de inserção possui uma eficácia muito geral. A sua aplicabilidade deve ser discutida, no português, relativamente à tmesa no futuro e no condicional e aos advérbios em *-mente*.

##### 1.4.1 – O futuro e o condicional

*Chamarás*, a par de *chamar-me-ás* e de *chamá-lo-ás*, *chamarias* a par de *chamar-me-ias* e de *chamá-lo-ias*, entroncam num dos tipos de futuro românico, do presente e do passado, respectivamente, a partir de perífrases que têm originariamente carácter modal : *clamare as* (<*habes*), *clamare ias* (<*habias*). A posição dos pronomes átonos depois do infinitivo e antes do auxiliar foi corrente no castelhano até ao século XVI e no provençal arcaico (Lausberg 1963: 408) e é quando se dá a soldagem plena dos dois elementos constitutivos, ou seja, quando deixa de haver duas palavras e passa a haver uma só, que os pronomes átonos se tornam externos a essa unidade. No português, a designação de *mesóclise*, sugerindo que os pronomes átonos se posicionam no meio da palavra, pode fazer perder de vista que, na realidade, eles se situam no ponto em que se encontram duas palavras que não chegaram a soldar-se completamente, como o comprova o facto de, nas formas com tmesa como *chamar-me-ás*, a vogal temática se manter aberta, não sofrendo a redução inerente à atonicidade, como acontece em *chamarás* e em todas as formas arizotónicas<sup>7</sup>.

##### 1.4.2 – Os advérbios em *-mente*

Quanto aos advérbios em *-mente*, sequências como “*rápida e prontamente*” remontam à construção latina do mesmo tipo (cf. “*Bona et grata mente*”) que, além do português, só sobrevive em castelhano,

---

<sup>7</sup> Sobre a existência, no português contemporâneo, de dois acentos nas formas com mesóclise, veja-se Mateus *et al.* (2003: 1062), bem como Pereira (1999: 185-189), que refere a restrição do uso deste tipo de formas (Pereira 1999: 187).

mas foi corrente nas fases antigas do francês, do provençal, do italiano e do aragonês (Pottier 1976: 228-231). O facto de o adjectivo apresentar a forma feminina em concordância com o substantivo *mente* e de ser susceptível de superlativação (*rapidissimamente*) comprova que se trata de duas palavras que mantêm identidade mórfica; a transposição de *mente* para o termo da sequência só é possível porque essa unidade não é um sufixo: pode haver elipse de palavras, não de partes de palavras. Como diz Câmara (1985: 121-122), há “a coordenação de dois ou mais adjectivos, subordinados a um único vocábulo *mente* no fim da sequência”.

No plano prosódico, cada um dos elementos constitutivos mantém o acento (cf. Mateus *et al.* 2003: 1062), o que confere aos advérbios em *-mente* um “estatuto particular”, “reflexo sincrónico da sua origem compósita” (Pereira 1999: 217); por outras palavras, não se deu a perda de autonomia que acompanha os fenómenos de aglutinação (Câmara 1985: 38-39), facto que pode projectar-se na grafia (ver, adiante, 2.2.1.5).

#### 1.4.3 – Outras restrições

São de outra natureza as restrições à validade do critério de não inserção: por um lado, entre duas palavras (p. ex.: *tão contente, muito bem, para ti, com todos, entre vários*) pode não ser possível inserir uma terceira sem que elas deixem de ser duas palavras; mas principalmente o critério adequa-se também à identificação de unidades pluriverbais (Zuluaga 1980: 98; Gross 1996: 18-19) (como *céo da boca, haver mester, Vossa Mercê, cada um, por causa de, posto que*), não sendo por isso possível através dele estabelecer a fronteira entre esses dois tipos de unidades. Outros critérios, que são afinal particularizações ou consequências da fixidez da unidade, são por isso bivalentes, como a não alteração da ordem dos elementos (Zuluaga 1980: 105).

#### 1.5 – Interrogando a diacronia

Raramente se procuram na diacronia indícios para a conceituação de palavra (cf. Juilland & Roceric 1972: 71). Contudo, uma perspectiva histórica relativa ao vocábulo patenteia fundamentalmente três derivas: a permanência do vocábulo lexical (1.5.1), que contrasta

com a caducidade do vocábulo funcional, que atinge, por ordem decrescente, conjunções, advérbios e preposições, substituídos por unidades pluriverbais, aglutinadas ou não (1.5.3.); um tipo diferente de deriva, que engloba o pronome e artigo *o*, e outros morfemas dependentes<sup>8</sup>, deverá ser tida em conta, à luz das relações entre palavra e contexto (1.5.2).

### 1.5.1 – Permanência do vocábulo lexical

Quanto à primeira, mesmo tendo em conta que as diferenças de carácter diastrático e diafásico afectam acentuadamente esse plano do latim vulgar, uma ampla zona de identidade caracteriza o léxico românico na sua relação com o latim clássico.<sup>9</sup>

#### 1.5.1.1 – A impermeabilidade em relação ao contexto

Dentro desta continuidade destaca-se, no plano fonológico, a muito geral diferença de evolução de numerosos fonemas consonânticos (para restringimos a observação aos casos mais gerais e simples que, contudo, configuram contrastes fortemente simétricos), conforme a sua posição é inicial ou interna e, neste caso, intervocálica. É neste contexto que se dão fenómenos de sonorização, fricativação ou queda<sup>10</sup>, casos particulares do fenómeno de assimilação dupla que

---

<sup>8</sup> Deve-se a Câmara (1959: 104) a conceituação de “forma dependente” – “Ampliando a classificação de Bloomfield, podemos dizer que, além da forma livre, que aparece não raro isolada, e da forma presa, que só aparece ligada a outras e por elas condicionada, há a forma dependente, que é autónoma embora nunca apareça isolada.” – posteriormente retomada no *Dicionário de Linguística e de Gramática* (Câmara 1984: s.v. *forma*). Dado que, no português, as “formas dependentes” são sempre morfemas gramaticais, e que se impunha distingui-los dos morfemas ligados ou afixos, a designação de *morfemas dependentes* (cf. Pena 1999: 4324-4325) é, de entre as designações conhecidas, aquela que melhor se ajusta às concepções expressas neste artigo.

<sup>9</sup> Cf. Bourciez (1967): “Les mots employés par les Latins en parlant sont en principe ceux qui se trouvent chez les écrivains classiques: subst. *homo, filius, manus, aqua, panis, ferrum, rota, asinus, cervus, caelum*[...]; adj. *Rotundus, plenus, calidus, siccus, niger, novus, bonus*; verbes *habere, facere, dormire, bibere, currere, videre, credere*, etc. Ces mots, et des centaines d’autres se retrouvent encore sans exception dans toutes les langues romanes.” (Bourciez 1967: 53-54).

<sup>10</sup> Exemplos de sonorização: NĀPU- > *nabo*, RŌTA- > *roda*, IŌCU- > *jogo*, PRŌFĒCTU- > *proveito*, RŌSA- > *rosa*, FACĒRE > *fazer*; de fricativação: CABĀLLU- > *cavalo*; SCŌPA- > *escoba* > *escova*; de queda: PEDE- > *pé*; LŪNA- > *lua*; FĪLU- > *fiu*.

decorre da permanência do contexto fonológico inerente à fixidez do vocábulo. O facto de, em posição inicial, esses fenómenos não se darem, é uma consequência da variabilidade do contexto fonológico, decorrente da mobilidade da palavra e da extensa gama combinatória de palavras na sequência das quais cada palavra se situa.

É certo que existem em todas as línguas românicas casos de sonorização das oclusivas surdas iniciais, e que o facto tem sido relacionado com “uma variação fonético-sintáctica” que se mantém em vários dialectos sardos<sup>11</sup> e “que originalmente constituía uma tendência geral” (Meier 1990: 14).

Na mesma linha, mas situada no cerne da evolução do consonantismo ocidental inicial, Lloyd (1987: 235 e ss., 244, 368) defende a existência de um fenómeno geral de variação das oclusivas iniciais que apresentariam as mesmas evoluções que no interior de palavra: depois de pausa ou de palavra terminada em consoante, a oclusiva inicial mantém-se forte; depois de palavra terminada em vogal, desenvolver-se-ia um alofone “débil ou frouxo”, igual ao que aparecia entre vogais no interior da palavra (Lloyd 1987: 235). Mas a total ausência de comprovação documental do fenómeno, a insuficiência da explicação por “fonologização da variação” (Lloyd 1987: 235, 243), não são supridas pelo papel atribuído ao substrato celta<sup>12</sup>, que não elimina a dificuldade em compreender-se como se passa da assumida sonorização generalizada das oclusivas surdas iniciais após palavra iniciada por vogal aos resultados prevalentes nas línguas românicas ocidentais. Pelo improvabilíssimo ensurdecimento das oclusivas iniciais? Ou pela conservação dos fonemas latinos originários? As duas últimas objecções englobam a tese de H. Meier.

---

<sup>11</sup> Assim, “em posição inicial absoluta [...] ou pos-consonântica: (il)las terras”, “em posição intervocálica: (il) la derra” (Meier 1990).

<sup>12</sup> “La influencia céltica estaría detrás de la adopción de la sonorización como realización fonética de la variación en occidente y su posterior fonologización [...]. El sustrato celta habría sido un factor concomitante en el cambio, un factor que habría coadyuvado a determinar la dirección que ese cambio tomó en el romance occidental.” (Lloyd 1993: 245).

Em suma, em comparação com a conservação, o número de fenómenos de sonorização de oclusivas iniciais latinas<sup>13</sup> é demasiado reduzido para pôr em causa as consequências do carácter variável ou fixo do contexto. Quanto a outros fenómenos de sândi, a sua frequência na escrita medieval, decorrente da maior proximidade desta em relação à língua oral, contrasta com a raridade com que a palavra foi estavelmente afectada pelo contexto, motivo por que o fenómeno se revela com consequências históricas escassas.<sup>14</sup>

A fixidez interna, a par da variabilidade dos contextos efectivos, torna-se assim um critério para o reconhecimento de ‘palavra’; essa variabilidade decorre de opções facultativas do locutor, dentro da gama de restrições gramaticais que afectam cada signo, o que origina outro princípio de identificação: o da possibilidade exclusiva de comutação com unidade da mesma classe. O princípio, explicitado negativamente por Zuluaga<sup>15</sup> para a detecção de unidades pluriverbais, pode ser transposto para o âmbito da ‘palavra’, para cuja existência e delimitação dá um contributo particularmente eficaz (cf. 3.1 e 3.2.1).

---

<sup>13</sup> São comuns ao português e ao castelhano vários casos de sonorização de p-, como *pastinaca-* > cast. *biznaga*, port. *bisnaga* (cf. García de Diego 1951: 85); port. *bostela* < \**pustella* por *pustula* (Nunes 1956: 86); mais frequente, embora mantendo carácter esporádico, é a sonorização da velar inicial, fenómeno por vezes já documentado em latim e comum a várias línguas românicas, e que no português é patente, entre outros casos, em *gato* < *cattu-*, *gamela* < *camella-*, *gaiola* < \**caveola-*, *gávea* < *cavea*, *gorgulho* < *curculio-* (Menéndez Pidal 1904: 200; García de Diego 1951: 85; Nunes 1956: 87; Williams 1975: 71)

<sup>14</sup> Além de *irmão* < GERMĀNU- (Nunes 1956: 161; Williams 1975: 72), e de *Eanes*, patronímico proveniente de *Joane* (Nunes 1956: 161), os casos apontados pelas gramáticas históricas citadas são na origem unidades pluriverbais submetidas a dinâmicas próprias no plano do conteúdo e da forma que é afectada pelas tendências coevas ou posteriores ao momento em que surgiram como, por exemplo, as formações baseadas em *hora*, quer remontando ao latim vulgar, no caso de *agora* < HĀC HŌRA, quer constituídas já em português, como aconteceu com *essora*, *outrora*, *embora*.

<sup>15</sup> Zuluaga (1980: 98): “Las unidades fraseológicas rechazan [...] su conmutación por equivalentes, p. ej.: [...] *a brazo partido* - \* *a brazo quebrado* [;] *corriente y moliente* - \* *común y moliente*.”

### 1.5.2 – A consolidação da alomorfa contextual: *o*, artigo e pronome e outros morfemas dependentes

A história do artigo e pronome *lo/no/o*, em várias das suas fases, só é compreensível se se perspectivar a integração das palavras morfológicas em unidades fonológicas no interior das quais se dão fenómenos que restringem a amplitude do princípio de impermeabilidade da palavra à acção modificadora do contexto.

Tendo como ponto de partida o acusativo do demonstrativo latino *ILLŪ*, a redução da forma, decorrente do enfraquecimento de sentido, conjugado com a frequência e a atonicidade, é um facto românico (Lausberg 1963: 350-355); os contornos próprios ao português, a partir da forma *lo*, são originados por fenómenos fonéticos contextuais de assimilação, que englobam a palavra de cujo acento a forma átona depende, e originam os segmentos terminais *lo* (*dáslo* > *dállo* > *dálo* = *dá-lo*; *perlo* > *pello* > *pele*) ou *no* (*põelo* > *põeno* = *põe-no*; *enlo* > *enno* > *eno*, posteriormente *no*) e *o* que, resultando da queda de *-l-* (que se mantém quando é proveniente de uma geminada e quando é inicial), surge ainda em consequência da acção do contexto intervocálico, designadamente após preposições: *a lo rio* > *ao rio*.

A alomorfa foi comum ao pronome e ao artigo no português medieval,<sup>16</sup> por exemplo “*Viste las* (= *Vistes as*) *jurás*” (*Cancioneiro da Vaticana*) e ainda renascentista “*Vede los* (= *Vedes os*) *Alemães*” (*Lusíadas*) (Williams 1975: 146) e está documentada em Fernão d’Oliveira, ao nível do uso (2 oc.): “despoys das pessoas que acabão em *.s. mudamolo* (= *mudamo’lo*) *.s.* em *.l.*” (GR 31 7-9); “[o] costume

---

<sup>16</sup> Destaca-se a particular frequência depois de morfemas dependentes como, em documentos de Entre Douro e Minho: *per la condiction* (1262), *por lo dito moesteyro* (1344) *tras la granja* (1424 (Maia 1986: 644-648); nos *Diálogos de São Gregório*, a tendência à “constituição de um único vocábulo fonético”, o que se traduz numa só unidade gráfica, ocorre com “após, depós, despós, ambos, todos, per, por”, sendo os totais das formas aglutinadas como *apolos* e não aglutinadas, como *após os*, respectivamente de 258 e 11 ocorrências (incluindo *sobrelo*, interpretado como *sobre+elo* (Silva 1989: 143-144); refira-se finalmente a menor frequência do alomorfe *no*, como em “nem *no* abbade nem *no* conuento” (1299), uma vez que depois de nasal pode ocorrer também o alomorfe *lo*: (“cõ *la* dita herdade (1333) (Maia 1986: 648), embora nos *Diálogos de São Gregório* a forma corrente depois de nasal seja *no* (Silva 1989: 149-150)

da nossa língua que he amiga *dabрила* (= d'abri'la) boca" (GR 47 12-13). No plano descritivo, comprovando a capacidade de observação e de relação do autor da primeira gramática, o facto é referido como consequência do "parentesco ou vizinhança" entre "letras" (entenda-se 'fonemas') e facilidade "em se mudar hũas em outras": "*amarano* seu deos por *amarão* o seu deos" e considerado afim do que ocorre nas contracções "*no*" por "em o" e "*pollo*" em lugar de "por o" (OGR 25 27-31). Duas outras ilustrações de factos gerais – "*amaylo* vosso deos" (31 9) e "*queremno* bem" (35 11) – podem indiciar que a alomorfa do artigo é então, na linguagem oral, menos rara do que a sua representação na escrita faz supor.

Quanto à questão em aberto, da autonomia formal da palavra em relação ao contexto, importa ter em conta, tomando como referência o português padrão contemporâneo, que a alomorfa contextual foi eliminada no artigo e tende a reduzir-se no pronome, restringindo-se à ênclise e mesóclise verbais<sup>17</sup>, à ênclise nos pronomes átonos *nos* e *vos*, e mais raramente, ao advérbio *eis*.

A densidade de morfemas dependentes afectados pelo contexto não é irrelevante: enquanto os morfemas ligados sofreram as mudanças fonéticas regulares internas à palavra (o morfema verbal da segunda pessoa do plural –TIS > -des, o morfema modo-temporal do imperfeito –BA > -va), os morfemas dependentes são, de todas as palavras, aquelas que participam em menor grau da impermeabilidade formal ao contexto, ou seja, constituem o sector em que os fenómenos de sândi são mais frequentes: quando as duas palavras se aglutinam e são ambas modificadas por assimilação, como em *pele*, *polo*, *ambolos*, *todolos* e *no*; ou quando a primeira das palavras, terminada em vogal, a perde por elisão, como nos casos, ocorrentes no *corpus*, em que assinalámos o facto por duplo apóstrofo: *d"oliveyra*, *d"aver*, *desd"então*, *contr"elles*, *com"estas* (cf. 2.2.1.3). Uma vez mais é a maior proximidade do oral que a prática acusa; ela será afectada pela

<sup>17</sup> Só "na linguagem popular e na literária popularizante de Portugal" é hoje possível o uso do alomorfe nasal depois dos advérbios *não* e *bem* e depois de "outras palavras terminadas em ditongo nasal", como "E assim pedia, num dó tamanho, / Não no tirassem lá donde estava." (António Nobre) (Cunha & Cintra 1984: 280).

posterior normalização que restringirá as contracções, na escrita, a pouco mais do que aos pronomes átonos como *mo* e congêneres e às preposições *a*, *de*, *em*, e *por*, quando antes de morfemas dependentes.

Em todos os casos analisados até este momento, a permanência da palavra enquanto signo tem como consequência que a relação genética se estabeleça mediante o delineamento da evolução nos planos fonológico, semântico e gramatical. Não assim relativamente ao terceiro tipo de deriva.

### 1.5.3 – Caducidade do vocábulo funcional e seus sucedâneos

Da comparação entre o elenco de advérbios, preposições e conjunções em latim e nas línguas românicas ressalta a frequente não sobrevivência dos vocábulos latinos<sup>18</sup>.

#### 1.5.3.1 – Unidades pluriverbais, gramaticalização e gradação

Entre os novos meios de expressão avultam sequências de palavras que constituíram uma unidade no plano do significado, que se tornou inanalísável, e no plano da forma, que apresentava diversos graus de coesão interna; o grau máximo consistiu na aglutinação, que se traduzia na existência de um único acento tónico e na constituição de uma unidade que apresentava as características do que se designa por palavra.

As palavras assim nascidas resultam de vários tipos de combinações, como preposição e advérbio: DE PŌST > *depois*, ĪN TŪNC > *então*; substantivo precedido de preposição: AD NŌCTEM > *ontem*,

---

<sup>18</sup> “Dos advérbios latinos, originados, na maior parte, de nomes ou pronomes, poucos passaram às línguas românicas” (Ali 1966: 183); “Grande parte das preposições latinas se perderam” (Câmara 1985: 177); “Da respeitável série de conjunções que faziam parte do idioma latino poucas passaram às línguas românicas” (Ali 1966: 220); “en dehors de quelques conjonctions simples, les langues romanes disposent surtout de conjonctions composées qui contiennent, d’une part, un élément purement conjonctionnel et, de l’autre, un ou plusieurs éléments (le plus souvent adverbiaux ou prépositionnels qui constituent l’expression explicite de la fonction et du sens de la conjonction” (Hermann 1963: 21).

AD MANEĀNĀM > *amanhã*, substantivo precedido de demonstrativo: HĀC HŌRA > *agora*; e apresentam como traços predominantes: a expressão de conceitos básicos, ao mesmo tempo que constituem o único meio monoverbal de expressão desses conceitos, a perda de densidade semântica, a inserção num paradigma (advérbio, preposição ou conjunção) e correspondente partilha dos atributos e das regras que definem a classe.

Nesta enumeração, não exaustiva, são notórias as afinidades com características (algumas das quais não incontroversas) dos processos de gramaticalização: “perda de substância semântica e [...] fonológica” (De Mulder (2001), citando Lehman) diminuição da capacidade de opção do locutor – “plus un signe est grammaticalisé, moins le locuteur peut décider lui-même de l’exprimer ou pas”) o carácter unidireccional do fenómeno, do “termo lexical ao elemento gramaticalizado” (De Mulder 2001: 13, 14; Peyraube 2002: 51), e a existência de graus, quer porque há “etapas transitórias” (demonstrativo > artigo definido > marcador de género) (Peyraube (2002), citando Greenberg) quer porque a distância entre o ponto de partida e o ponto de chegada é variável.

Se é certo que existem unidades pluriverbais em todas as classes de vocábulos, e se os processos de gramaticalização se dão também em unidades monoverbais<sup>19</sup>, elas adquirem uma importância específica no âmbito dos vocábulos gramaticais, constituindo uma deriva própria, pela homogeneidade dos traços evolutivos, designadamente a relação entre génese e funcionamento, marcada pela direcção ao mais geral e ao mais abstracto e o carácter gradual; no plano do uso, este traduz-se pela consolidação e expansão da sequência (o que é depreensível pela quantificação), nos planos da forma, do conteúdo e da relação entre ambos, pela perda progressiva da autonomia dos elementos de formação e, relativamente ao elenco de unidades assim constituídas, pela existência de uma hierarquia de níveis de coesão.

---

<sup>19</sup> Como: HŌRA- > *hora* substantivo > *ora* advérbio > *ora* conjunção, *ora...ora* conjunção; MĀGIS advérbio > *mais* advérbio > *mas* conjunção; *segundo*, *salvo*, etc. que funcionam também como preposições, etc.

Sem abordarmos, ao nível teórico, a questão do carácter gradual da mudança linguística, afirmada pelos neogramáticos e regressada à actualidade em tempos recentes, os traços referidos devem projectar-se numa diferença de tratamento das unidades pluriverbais quer o sejam no ponto de partida, em determinada fase do passado ou na fase da língua em estudo.

## 2 – A palavra no *corpus* gramatical quinhentista

As dificuldades na conceituação de palavra e as relações entre unidades mono- e pluriverbais conferem um interesse particular aos textos dos gramáticos e ortógrafos quinhentistas, porque por um lado através deles é possível avaliar a amplitude da variação, as tendências dominantes e as tendências prevaescentes e, por outro lado, ponderar a hipótese de que a dimensão metalinguística e as concepções de escrita se projectem na delimitação das sequências gráficas mínimas.

### 2.1 – Caracterização do *corpus* e súmula do tratamento informático

O *corpus* (descrito no *Apêndice*) reúne as obras metalinguísticas sobre o português publicadas entre 1536 e 1606 e distribuídas por duas sincronias, 1536-1540 e 1574-1606, separadas por um hiato editorial de 34 anos. Constituído para a avaliação do papel dos gramáticos (*lato sensu*) na constituição da norma do português moderno (cf. Paiva 2002), ele apresenta uma extensão compatível com a determinação de tendências gerais e de fenómenos de amplitude grande ou média e, relativamente à primeira sincronia, faculta a distinção dos contributos de Fernão d’Oliveira e de João de Barros, cujas obras, de extensão aproximada, são tratadas exaustivamente. A segunda sincronia é menos informativa, por ser menos extensa (26,8% do total), por ser constituída por amostra aleatória ampla (1/4 de cada obra) e porque, dentro das contingências da História, Duarte Nunes de Leão contribui com 23% para a informação total, enquanto a parte de Pêro Magalhães de Gândavo é apenas de 3,1%. Sendo assim, a segunda sincronia deve ser vista como um todo (em que se divisam contudo matizes diferenciais) que permite conhecer a distância entre a génese da dinâmica desenhada na primeira sincronia e os resultados que se projectam no discurso metalinguístico, algumas décadas depois.

Visto que a definição de tendências só é possível pela delimitação de um fundo de constância/identidade sobre o qual se destaca a variação, e também porque se concebeu a mudança como síntese de diferenças que começam por ser captadas pontualmente e vão coincidindo na formação de coordenadas, o *corpus* foi submetido a tratamento informático que conduziu à produção de um *Índice Geral de Vocábulos*. São os dados resultantes desse processo que se utilizam aqui.

## 2.2 – Da palavra gráfica à palavra propriamente dita

A primeira condição para que os objectivos anteriormente definidos (cf. 2 e 2.1.) sejam alcançados é que, nas pré-edições dos textos, as intervenções do editor sejam sempre assinaladas, mediante a aplicação de um código de convenções; a segunda é que se parta da palavra gráfica, tal como ela existe na imprensa quinhentista<sup>20</sup>.

Única realidade objectiva imediatamente depreensível, a palavra gráfica é um grafema ou uma sequência de grafemas cuja delimitação mais geral é, simultaneamente, o espaço anterior e o espaço seguinte, que incluem os sinais de pontuação, neles contidos.

No plano da delimitação, as diferenças, relativamente a usos actuais, são a frequente omissão do traço de translineação, os alógrafos posicionais que assinalam o início do vocábulo ou o seu termo, a maiúscula quando funciona como fronteira de palavra e a nota tironiana que, representando a conjunção copulativa “e”, quando minúscula, é um ideograma que se auto-delimita. O apóstrofo e o traço de união são descritos na *Ortografia* de Leão (1576), mas o seu uso restringe-se a essa obra e mesmo aí é raro e esporádico (cf. 2.2.1.4 e 2.2.1.5).

### 2.2.1 – Segmentações gráficas

#### 2.2.1.1 – Rupturas

Escritas como “*a verbios*” (OGR 19 29), “*da hy*” (BDL 55R 1, 58V 14) cortam a unidade inseparável que é a palavra, destacando segmentos, sempre monossilábicos, que são homónimos de morfemas

---

<sup>20</sup> Cf. Paiva (2002: IV, pp. 2-21).

dependentes como *a*, artigo (“*a quella* – BGR 6R 20), preposição (“*A quy*” – BGR 40R 11, 50R 3), pronome (“*a vendo*” = *havendo* – BDL 59R 3), *da*, contracção da preposição e artigo (“*da hi*” – OGR 10 16; “*da qui*” – Barros, 5 ocorrências, a par de 1 ocorrência de *daqui*), *mos*, contracção dos pronomes *me* e *os* (“*have mos*” - OGR 53 23).

Estas rupturas decorrem principalmente: primeiro, da projecção, sobre a escrita, do facto de, em latim, as preposições funcionarem também como prefixos (“*de termina*” – BGR 20v 19, “*de sejar*” (BGR 25V 19), “*a proveitar*” (BGR 16R 8, 40R 20), “*a proveitão*” (OGR 59 11), o que deve também ser relacionado com a existência, na língua coeva, de formas como *terminar* ou *proveito*; segundo, da insuficiente maturidade do reconhecimento metalinguístico de morfemas dependentes, designadamente, em Barros, a não identificação das contracções das preposições com o artigo; terceiro, de uma “comutação” ingénua praticada entre palavras com um segmento comum (em Barros: “*da quelle*” (5 oc. num total de 8) como “*na quelle*” (4 oc. num total de 6), ainda que este recorte anule as relações de conteúdo e de emprego gramatical (cf. Meillet 1921: 1.3) que são inerentes aos signos que são palavras; quarto, também em Barros, do predomínio do pendor analítico e da tendência a remontar a uma suposta origem, ignorando a monossemeia (“*quál quer*” – 55V 8; “*em tam*”, 4 oc., a par de 8 de “*entam*”).

Finalmente, das 58 rupturas deste tipo documentadas no *corpus*, 54 afectam o início da palavra e 4 o fim. É significativo que em Oliveira se encontrem apenas 4, e em Barros todas as restantes, e ainda que não estejam documentadas na segunda sincronia, o que reflecte o aumento da consciência metalinguística. Neste aspecto, devem ser relacionadas com um dos tipos de junção (a de partículas), tratado a seguir.

### 2.2.1.2 – Junções simples

Relativamente aos pronomes átonos usados na contiguidade do verbo, a norma, muito estável, apesar de excepções pontuais, e sem alteração ao longo do tempo, é a separação gráfica quando a forma ocorre antes do verbo, e a junção quando ocorre depois; assim: “*a põe*” (OGR 21 17), “*se segue*” (GRE 1 6), “*vos mostrei*” (LRI 62V 23); “*Lembrate*” (BGR 11v 22), “*Dizeime*” (GDD 56 5), “*atrevamonos*” (LRT 62R 20); “*temolo*” (OGR 28 11), “*conhecemno*” (BGR 9v 14).

Em mesóclise as práticas variam entre a opção por um único segmento gráfico (“*diloemos*” – OGR 39 15; “*farteey*” – BGR 43V 23), ou por dois, sendo então o pronome agregado à parte da palavra coincidente com o infinitivo (“*ficárnos ã*” – BDL 56R 10; “*veerte ei*” – LRT 6 18).

As contracções com modificação simultânea dos elementos constituintes apresentam todas as possibilidades: separação: “*no lo*” (nos + lo) – OGR 27 14; junção: “*nolo*” (nós + lo) – OGR 17 29, “*nolo*” (nos + lo) – LRT 7- 35, “*eillo*” (eis + lo) – BGR 28V 21; junção e aglutinação gráfica ao verbo: “*tenhovo*” (BGR 39V 3).

O princípio de não intervenção relativamente a práticas constantes que foi adoptado nas pré-edições, teve como consequência que estas sequências gráficas fossem reproduzidas tal como ocorrem. No *Índice de Vocábulos*, elas aparecem integradas sob os lemas e sub-lemas que nelas estão reunidos: assim, *temolo* figura sob o lema **ter** (verbo) e sob o lema **o** (pronome), que inclui os sub-lemas necessários à estruturação do quadro da alomorfa, em conexão com a separação ou a junção.

A junção gráfica dos clíticos que seguem o verbo, em contraste com a separação quando o antecedem, não pode deixar de suscitar reflexão.

A junção relaciona-se com a alomorfa que afecta os dois simultaneamente: em “*vedallo*” (GRE 37 4) o locutor reconhece o infinitivo de *vedar* e o mesmo pronome que em “*seguiolos*” (OGR 6 22), mas o fonema único que os reúne pertence tanto ao termo do verbo como ao início do pronome; a regra que se generaliza é assim a junção gráfica, mesmo quando não há alomorfa.

Outra razão não menos relevante é o facto de entre a forma verbal e o clítico não poder inserir-se nenhuma palavra, o que não acontece quando o clítico se situa antes do verbo. Mesmo restringindo a observação às ocorrências de *nos*, constata-se que é mais livre do que hoje a ordem de palavras: entre a forma átona e o verbo pode incluir-se a negação – “*porque nos não podemos salvar*” (OGR 13 10)<sup>21</sup>

<sup>21</sup> Neste parágrafo, a referência restringe-se à localização de *nos*.

– um quantificador – “de que *nos muito* servimos” (BGR 37R 11), “à conquista das quais *nos mais* demos que às treladações latinas” (BGR 56R 19) – mas também o sujeito – “necessidade que *nos o costume* já pôs” (OGR 30 25), “de que *nos Julio César* avisava nos guardássemos” (LRT 115 9).

Finalmente, não pode deixar de pensar-se em razões de ordem prosódica que são contudo dificilmente detectáveis neste caso.

Há outro tipo de junções que têm carácter esporádico e que foram assinaladas nas pré-edições. Das 334 junções desse tipo, que atingem 1% das ocorrências, 92,7% consistem na aglutinação gráfica de partículas (67,9%) e de unidades pluriverbais (24,8%), sendo as restantes atípicas e imputáveis aos compositores.

As junções do primeiro tipo afectam, por ordem decrescente de frequência: *a*, preposição (65 oc.: *a*)[*mi* – OGR 13 23, *a*][*ler* – BGR 1v 2), *a*, artigo (49 oc.: *a*)[*boca* – OGR 12 2, *a*][*cartinha* – BGR 1V 1), *de*, preposição (23 oc.: *de*)[*greçia* – OGR 58 9, *de*][*cujo* – BGR 1V 11), *e*, conjunção (16 oc.: *E*)[*assi* – OGR 6,15, *E*][*ajuntando* – BGR 20V 23), *o*, artigo (13 oc.: *o*)[*til* – OGR 22 18, *o*][*livro* – BGR 6V 21).

Apresentam como traços constantes: o facto de serem esporádicas, de se darem com elementos antepostos e não pospostos, de serem simultaneamente palavras de frequência elevada e de forma muito curta ou maximamente curta. Como as rupturas anteriormente tratadas (*a quella* ou *da hi*: 2.2.1.1), a sua frequência decresce acentuadamente na segunda sincronia, o que manifesta, neste aspecto, a diminuição da distância entre palavra gráfica e palavra real.

São de teor totalmente diverso junções como *amesma*, *paraque*, ou *outrotanto*, que resultam da tendência a fazer coincidir unidade pluriverbal e unidade gráfica. Dos dois primeiros gramáticos, é Oliveira quem mais pratica este tipo de aglutinação gráfica (36,6%), de acordo com a aptidão a captar e a exprimir desse modo a monossemia; em Barros a percentagem decresce (22,6%), e a disparidade acentua-se se se tiver em conta que *oqual* e flexão representam 73,3% desse total (a par de 30,7% em Oliveira).

Pelo contrário, na segunda sincronia, as junções quase se restringem a unidades pluriverbais: Gândavo escreve “*assicom*, *convemasaber*” (*convem a saber*), Leão “*perque*, *desque*”.

### 2.2.1.3 – Junções com elisão ou crase

São fenómenos de sândi que estão na origem de aglutinações gráficas (que assinalamos por *''*), como *su''alteza* (BDL 50V 16), *ate''gora* (LRT 62V 9) e ainda *s''espantem* (OGR 33 16), *d''homens* (OGR 6 13), *antr''ellas* (BCA 5 11), *desd''o* (LRI 7-13).

Não admira por isso que apresentem uma frequência dependente do grau de atenção que o gramático prestou ao nível prosódico e da importância que lhe atribui na escrita. Daí, a sua frequência máxima em Fernão d'Oliveira, com a percentagem de 0,80%, quando a percentagem geral no *corpus* é de 0,37%.

As ocorrências distribuem-se por duas constelações: a primeira é formada por sequências frequentes, algumas das quais reúnem atributos habitualmente seleccionados como identificadores de unidades pluriverbais, designadamente, monossímia, forma fixa que não admite inserção, alteração da ordem, comutação com sinónimos ou transformações sintácticas correntes; as palavras que as constituem ficam internamente sujeitas a um contexto permanente, o que constitui, na linguagem oral, uma condição favorável à produção dos mesmos fenómenos; a memorização da unidade, também na sua dimensão fonológica e enquanto todo susceptível de ser reproduzido, inclui o fenómeno que de forma mais ou menos frequente a afecta. Daí que se constate um aumento das percentagens de elisão e crase nas unidades pluriverbais do *corpus*, embora este não seja suficientemente extenso e diversificado para que o facto se possa reduzir a percentagens com carácter probante. Contudo, das 3 oc. do nome da região *Entre Douro e Minho* em Oliveira, 2 apresentam a elisão da vogal final de *Douro*: *antre][dour''a][Minho* (68 20, 43 24); Barros, cuja escrita é pouco permeável a este tipo de fenómenos, escreve *su''alteza* (BDL 50V 16).

*Ainda agora, até aqui, até agora* não são expressões monossémicas (embora possa haver matizes diferenciais) e outros testes usados para a avaliação da fixidez das expressões ou não são aplicáveis (impossibilidade de inserção, alteração da ordem) ou são de aplicação melindrosa e de eficácia duvidosa: a comutação de uma das palavras por outra de conteúdo aproximado, oposto ou da mesma classe assenta na competência do locutor e estamos separados por mais de quatro séculos da língua dos textos; o exame das concordâncias atesta que hipóteses aventadas não estão documentadas, embora um *corpus*

mais extenso pudesse facultar provas que transformassem hipóteses em certezas. O primeiro argumento para a retenção destas sequências é a frequência que, apesar de baixa em consequência das dimensões do *corpus*, não é explicável pelo cruzamento de probabilidades. O segundo argumento é a proporção de fenómenos de sândi: das 7 oc. de *ainda agora* em Oliveira, 5 apresentam a forma *aindagora*; das 9 oc. de *até aqui* no *corpus*, as 6 que são usadas por Barros apresentam de forma constante a elisão da vogal inicial de *aqui*: *a te qui, té qui, té quy*. Embora *até agora* só ocorra duas vezes no *corpus*, apresenta a forma constante *até gora*, quer em Oliveira, quer na *Ortografia* de Leão, facto particularmente significativo porque entre as duas obras medeiam quarenta anos e porque Leão, fortemente relatinizante, sistematizante e normalizante (principalmente nesta obra) e consequentemente pouco interessado pela oralidade, não usaria esta forma se ela não fosse corrente.

A segunda constelação é constituída por morfemas dependentes cuja vogal final se elide antes de palavra iniciada por vogal: o fenómeno afecta pronomes (*s"entendem* – OGR 53 4; *est"ano* – OGR 43 25), conjunções (*com"este* – OGR 42 15), mas principalmente preposições: *de* perde a vogal e constitui uma unidade gráfica com qualquer lexema iniciado por vogal (*d"Afonso* – OGR 58 25; *d"agua* – OGR 43 13; *d"aprender* – BDL 50V 6, 50V 12); as preposições dissilábicas integram-se na mesma tendência (*desd"as* – (OGR 11 32; *contr"elles* – OGR 6 26; *antr"ellas* – OGR 29 18, BCA 5 11), *sobr"elle* (OGR 21 3).

A evolução acusa-se já em Barros, em cuja obra, das 12 aglutinações gráficas deste tipo, 8 afectam a preposição *de* (*d"ano* – 47V 19; *d"alma* – 59R 2). Na segunda sincronia (em que estão documentadas apenas duas elisões deste tipo: *desd"o* – LRI 7- 3 e *d"espigas* – LRI 78 9) está claramente patente o rumo que se projectará na normalização da grafia, que se processará pelo afastamento da linguagem oral: à variabilidade e à frequência dos fenómenos de sândi sucederá a fixação das combinações e a sua restrição a determinados morfemas dependentes; ficam excluídas as conjunções, as preposições dissilábicas, enquanto *de* passa a contrair-se unicamente com vocábulos gramaticais.

Em síntese, a fixação da palavra gráfica, neste aspecto, conduz à sua identificação com a palavra propriamente dita. Por outro lado,

a consolidação das contracções impõe a dissociação entre palavra e monossemia, uma vez que, no plano do conteúdo, se mantém claramente distinto o contributo das duas palavras originárias.

#### 2.2.1.4 – O apóstrofo

A condenação de grafias em que se pratica a aglutinação da preposição *de* “com as dições [...] que começam em vogal” como “anel *douro*, homem *darmas*, *delle*, *della*” constitui a nona das vinte “Regras Geeraes” que, na *Ortografia* de Leão correspondem aos principais conteúdos programáticos da obra, preconizando o Autor que, se se quiser “na sriptura tirar o .e. como se tira na pronunçiação”, se use o apóstrofo no ponto de “detracção” da vogal (55R 18 – 55V 8). A questão é retomada no capítulo “Dos apostrophos” (67R 13 – 68R 10), onde se defende o uso do sinal em três casos: primeiro, nas “dições acabadas em vogal, em que mais cõmummente comemos & tiramos a dicta ultima vogal, [que] são estas, *de*, *me*, *te*, *se*, *que*, *ante*, *no*, *na*, *esse*, *este*, *aquelle*, *outro*” (67V 3-6); segundo, em “nomes compostos”, como *Montagraço* ou *Portalegre*, que deverão escrever-se *Mont’agraço* e *Port’Alegre* (67V 21 - 68R 3), terceiro, em “nomes próprios & cognomes” como *Dalmeida* ou *Dantas* que devem ser separados pelo apóstrofo: *d’Almeida*, *d’Antas* (68R 4-10).

A distância entre a doutrina e a sua aplicação é grande, quantitativa e qualitativamente.

Quantitativamente, porque o uso do apóstrofo é raro: na amostra da *Ortografia* incluída no *corpus* e que, como foi dito, constitui um quarto da totalidade da obra, o apóstrofo é usado apenas 21 vezes, das quais 3 em ocorrências metalinguísticas em que o grau de atenção e a imposição de coerência são maiores (*ap’lo*, abreviatura de *apóstolo* – 69R 1; *Mont’agraço*, que retoma o exemplo anteriormente dado, agora ilustrando o uso do hífen – 77V 24; *m’argüirá* – 77R 4). Factores análogos explicam que, das 18 oc. restantes, 8 se encontrem na *Dedicatória*, o que nos situa já no plano qualitativo.

De facto, é no plano qualitativo que a distância é maior: a observação de todas as ocorrências com apóstrofo comprova que da lista, acima transcrita, de palavras seleccionadas como aquelas em que mais frequentemente a vogal final é elidida, a única com que é usado o apóstrofo é *de*, não naqueles casos que haviam sido

descritos como próprios da linguagem oral como *douro* ou *darmas* (55R 18 e sgs), mas em contracções perfeitamente consolidadas, algumas maximamente correntes (que reproduzimos no masculino do singular): *d’algum*, *d’aquela*, *d’elle*, *d’este*, *d’entre*, *d’o*, *d’outro*.

O uso do apóstrofo em LRT representa como que um regresso às origens praticado unicamente em casos de génese inequívoca, sem qualquer incidência sobre os fenómenos de sândi observáveis na língua coeva e sem consequências no conceito de palavra, uma vez que Leão separa palavras preexistentes. Não admira por isso que o próprio Autor abandone o uso do apóstrofo, totalmente ausente da amostra da *Origem* incluída no *corpus*.

#### 2.2.1.5 – O traço de união

Descrito no mesmo capítulo que o apóstrofo, o hífen ou traço de união relaciona-se directamente com o conceito de palavra, visto que é usado “quando se ajuntão em hũ corpo duas dições diferentes ficando feitas hũa soo como *passa-tempo*, *guarda-porta*, *val-verde*, *Mont’-agraço*.” (LRT 77V 22-24).

Pela primeira vez na tradição gramatical portuguesa, um sinal gráfico assinala o facto linguístico que João de Barros tinha descrito no âmbito do conceito de “figura”, herdado da gramática clássica e na esteira da *Gramática* de Nebrija, e aplicado a cada uma das partes do discurso: enquanto “Nome simples é aquelle, as partes do qual estremadas hũa da outra nam sinificam cousa algũa” (BGR 9R 10-12), “Nome composto tem o contrario deste, por que partido em duas partes, sempre per hũa dellas entendemos cousa algũa, como. *Guarda porta*, que é composto deste verbo, *guardar*, e deste nome *porta*.” (9R 14 – 17).

A utilização do mesmo exemplo – *guarda-porta* – confirma dois factos: primeiro, que os exemplos tendem a ser transmitidos; segundo, que Leão conhece a *Gramática* de Barros, como outras coincidências não fortuitas atestam. Mas, sendo diferentes o teor das obras e as partes em que se inserem, o que condiciona o ponto de vista, Barros retém, como traço identificador, a conservação pelo menos de parte do conteúdo – “sempre per hũa dellas entendemos cousa algũa” – o que se ajusta aos outros exemplos que selecciona: *rede fole*, *arquibanco*, *torcicolo*, *morfeuge*, *puxavante*, *traspé* (9R 24 – 9V 5), independente-

mente do facto de constituírem ou não uma só unidade gráfica. Para Leão, o que está em causa é a explicitação da norma relativa ao uso de um sinal, justificado porque duas “dições” formam um só “corpo”, deixando de haver referência à combinação de conteúdos, que contudo é indescernível em *val-verde* e *Mont'-agraço* (exemplo já dado a propósito do apóstrofo (ver 2.2.1.4.), mas agora incluindo também o hífen) e puramente residual em *passa-tempo* e *guarda-porta*.

Na amostra da *Ortografia*, além destes casos, o hífen ocorre só mais uma vez, no início absoluto do texto: “Orthographia he sciencia de bem screver [...]. E diz se de orthos, que quer dizer directo, & grapho, screvo, como se dicessemos sciencia de *directa-mente* screver.” (1R 8 14).

O primeiro dos factores que se projectam na forma *directa-mente* é a influência de Nebrija e a prática de separação de *mente* que o gramático andaluz adopta em todas as suas obras em castelhano (Quilis 1981: 87): “Orthographía é palavra de origem grega que quer dizer “sciencia de bien & *derecha mente* escribir” (Nebrija 1492 [1981]: 16-17)<sup>22</sup>.

Dois outros aspectos devem ser retidos: a forma situa-se no início absoluto do texto, talvez revisto à luz de doutrina expressa na parte final, mas de qualquer modo objecto de maior auto-domínio e grau de consciência no uso da língua. Mais relevante é o reconhecimento por Leão do estatuto particular dos advérbios em *mente* e da não completa soldagem das palavras originárias, facto que traduz pelo traço de união. As consequências desse reconhecimento não ultrapassam este caso, única ocorrência em contexto não estritamente metalinguístico do traço de união em todo o *corpus* (cf. 1.4.2).

### 3 – Unidades monoverbais e pluriverbais

#### 3.1 – Retrospectiva

A definição de ‘palavra’ de Meillet (1.3) é a que mais amplamente se ajusta ao tratamento informático do *corpus*, e à necessidade, que lhe é inerente, de classificação exaustiva de segmentos gráficos

---

<sup>22</sup> Compare-se a definição de Barros: “Esta palavra, Orthografia, é grega: quer dizer ciência de escrever *dereitamente*.” (BGR 40R 3-4).

cuja participação nos atributos que delimitam esse conceito tem vindo a ser discutida: enquanto signo memorizado identificável pela associação entre uma forma, um conteúdo e um conjunto de regras gramaticais em que se insere a sua actualização, a definição reúne os traços simultâneos imprescindíveis para que uma ocorrência seja considerada uma palavra.

A primeira forma de testar a validade da definição consiste em avaliar a sua aptidão para excluir unidades que não são palavras.

As não-palavras são de vários tipos: há-as na área metalinguística, como quando Leão condena a forma *bernio*, na expressão *manto bernio*, por *hibernio*, “o que não he menos grosseria, que se dixeremos *Taliano* por Italiano & *Lemão* por Alemão.” (LRT 61V 19 – 62R 1). A inexistência de *taliano* ou *lemão* resulta de que, enquanto signo memorizado, reproduzível e reconhecível, a forma tem carácter fixo, ou seja não pode ser truncada.

No âmbito geral do uso da língua escrita, há sequências gráficas que não existem como palavras: porque são menos que uma palavra como a *[[quella* (cf. 2.2.1.1) e esse todo que é a palavra é inseparável (cf. 1.4); ou porque são mais que uma palavra como *a][boca* (cf. 2.2.1.2): entre *a* e *boca* pode inserir-se *sua* (*a sua boca* – BGR 36V 12) (cf. 1.4) e com certeza outras palavras, ainda que não documentadas nas concordâncias; além disso, *a* não é a única forma possível antes de *boca*, uma vez que comuta com *hũa* (*hũa boca* – OGR 16 13) e com a sua própria elipse (“este órgão *da* língua & *boca*” – OGR 4 24); *boca* não é também a única unidade possível depois de *a* (visto que ocorre também em *a língua* (OGR 7 31), *a terra* (OGR 8 25), *a necessidade* (OGR 9 23), etc: consequência da variabilidade dos contextos, a comutação com unidade da mesma classe (cf. 1.5.1.1) ganha em ser verificada pelas concordâncias, o que reduz os inconvenientes da aplicação de testes assentes na competência do locutor.

Constituem também mais que uma palavra, mas agora com carácter regular, as unidades gráficas que integram os clíticos que se seguem ao verbo (cf. 2.2.1.2), o que pode ser testado, quando não há alomorfa, pela alteração da ordem dos elementos, decorrente de restrições sintácticas: “*Damoslhe*” (OGR 23 21), mas “*não lhe damos*” (OGR 19 18); “*chamamoslhe*” (OGR 60 6), mas “*verdade é que de costume lhe chamamos [...]* til” (OGR 22 27-28). Quando

existe alomorfia, a alteração da ordem acompanha-se da alteração da forma, e é necessário ter em conta a dimensão gramatical da palavra e a existência de um morfema virtual aglutinador da variação contextual.

A variação contextual, limitada a morfemas dependentes e, em certos casos, à palavra de que dependem, individualiza um outro tipo de palavra (cf. 1.5.2) em que os significantes, afectados por fenómenos de fonética sintáctica em determinados contextos, perdem a uniformidade a favor de um determinado número de alomorfes que, adentro desse condicionalismo, tendem a tornar-se constantes e obrigatórios; além de *lo* e *no*, pronomes que, na esfera do verbo, foram referidos no parágrafo anterior, incluem-se neste quadro as combinações do pronome *o* com o pronome pessoal tónico *nós* (*nó lo*), com os átonos (*mo*, *to*, *lho*, *no lo*, *vo lo*) e com o advérbio *eis* (cf. 2.2.1.2), bem como outros morfemas dependentes em que se dão fenómenos de elisão e crase (cf. 2.2.1.3) e cujo maior contingente é o das preposições.

As combinações em que há dupla alomorfia – *nó lo*, *no lo*, *vo lo* – funcionam como um bloco inseparável com ordem fixa que se desloca enquanto todo; é a rede de associações virtuais própria da alomorfia que estabelece nexos entre estas formas e as formas não modificadas pelo contexto. Posteriormente ao período em estudo, o traço de união assinalará o estatuto próprio destas combinações.

O rumo da evolução que cristalizará na norma da linguagem escrita está definido no período em estudo (cf. 2.2.1.3): a maior proximidade da linguagem oral que, em Fernão d'Oliveira, é ao mesmo tempo um aspecto que o aproxima da Idade Média e um ponto importante no núcleo de ideias que configuram a sua coerência, cederá face à restrição da representação de fenómenos de sândi na escrita, simultânea à consolidação de formas gráficas que traduzem a projecção do nível prosódico sobre o nível morfológico (cf. 1.1). Assim, deixam de ocorrer formas tipicamente oliveirianas como *d"homens*, *contr"elles*, ou *com"este*; as combinações do pronome *o* com pronomes pessoais restringem-se às formas átonas; tornam-se cada vez mais raras sequências de formas não contractas; ao mesmo tempo dá-se a especialização da tendência, cuja expansão é contudo lenta e assimétrica: *de+o*, artigo definido, deixa de ser possível, embora *de+um*, artigo indefinido, subsista até aos nossos dias e as contracções de *em*

apresentem um permanente recuo sobre as contracções de *de*: *em+o*, artigo, está documentado desde Oliveira (*em as línguas* – OGR 10 10) a Leão (*em os vocábulos* – LRT 8V 6), mas da contracção de *em* com o indefinido há apenas uma ocorrência em todo o *corpus*: “*nũa certa parte*” (GDD 68 1).

O apóstrofo (cf. 2.2.1.4) poderia facultar a representação de fenómenos de oralidade, sem dúvida permanentes em certas condições porque, registados por Fernão d’Oliveira (*s’entendem, est’ano*), continuam a ouvir-se hoje, mas Leão, que o introduz, faz dele um uso não só esporádico, mas regressivo (*d’o, d’este*) na medida em que separa o que, aparecendo de modo constante fundido numa só forma gráfica, reflecte a identificação entre palavra gráfica e palavra propriamente dita.

### 3.2 – Unidades pluriverbais

#### 3.2.1 – Da unidade pluriverbal à unidade monoverbal

Entre os meios de renovação da língua, contam-se unidades formadas originariamente de duas palavras (e raramente mais de duas) que, frequentemente seguidas e associadas a um conteúdo que tende a tornar-se específico, passam a constituir um bloco prévio ao acto de discurso.

Na medida em que são fixas e participam de traços identificadores de ‘palavra’, como a não alteração da ordem dos elementos (Zuluaga 1980: 105), uma parte destas unidades pluriverbais tende a transformar-se em unidades monoverbais, mas nem sempre o trânsito é completo e subsiste a possibilidade de inserção de elementos no interior do conjunto, o que, bem como a conservação de traços prosódicos originários, caracteriza em português o futuro e o condicional (1.4.1) e os advérbios em *–mente* (1.4.2).

O traço de união (cf. 2.2.1.5), usado pela primeira vez por Duarte Nunes de Leão, designadamente para assinalar o estatuto particular destes advérbios, terá por função marcar determinadas unidades pluriverbais reconhecíveis cujos elementos constitutivos mantêm identidade prosódica e semântica.

Mas frequentemente, de um ponto de vista sincrónico, nada revela a génese pluriverbal de *agora, então* ou *ontem* (cf. 1.5.3.1). O locutor pode eventualmente estabelecer entre *também* e as palavras

originárias *tam* (ainda sem ditongo à data da constituição da unidade) e *bem* uma relação de tipo genético, mas essa relação não oculta a identidade de funcionamento: *tam*, *bem* e *também* possuem a mobilidade que as identifica como três palavras, cada uma das quais comuta com unidade da mesma classe (*tão* /  *muito* /  *menos* + *bem*; *tão* + *bem* /  *mal* /  *devagar*; *também* /  *igualmente* (cf. 1.5.1.1); mas quando *tam* e *bem* passam a integrar *também*, a variabilidade de contextos é eliminada e é *também* que passa a ter exclusivamente esse atributo, ao mesmo tempo que a distância entre o conteúdo da palavra originada e os conteúdos das palavras originárias dilui os nexos de relação. Não é por isso ocasional que das 199 ocorrências de *também* no *corpus*, das quais 181 na primeira sincronia, não haja um único caso de separação dos elementos de formação.

A não estabilização da palavra gráfica exigiu que se definisse o conceito de palavra, o que permitiu observar quais os aspectos de que os gramáticos têm mais dificuldade consciência e as consequências da reflexão metalinguística na delimitação da unidade de base (cf. 2.2). Essa não estabilização tem contudo a vantagem de, na dependência do tempo e da margem de intervenção decorrente de concepções individuais, poder reflectir a relação entre unidades pluriverbais e unidades monoverbais.

*Por que* / *porque*, causal ou final, é já, no período em estudo, uma conjunção caracterizada pela opacidade; é a percepção da monossemita que a junção gráfica reflecte, enquanto uma perspectiva analítica, genética e conservadora propicia a grafia com separação dos elementos de formação; a primeira atitude é a de Fernão d'Oliveira que, nas 218 oc. da conjunção, usa em 92,2% um só segmento gráfico, e em 7,7%, dois; a segunda atitude é a de João de Barros que, nas 143 oc. da conjunção, adopta a separação em 95,8% dos casos, em contraste com 4,19 % das ocorrências em que são reunidos os dois elementos. É a opção de Fernão d'Oliveira que reflecte a percepção do rumo da deriva: das 107 oc. da conjunção na segunda sincronia, todas consistem numa só palavra<sup>23</sup>.

---

<sup>23</sup> Não estão documentadas, relativamente a *porque*, ocorrências metalinguísticas que permitam relacionar as formas gráficas com a tipologia dessas ocorrências, mas é

Embora nem todos os casos apresentem uma tão acentuada clareza, o exame amplo da variação da delimitação gráfica comprova que a consciência da especialização funcional de uma unidade originariamente pluriverbal se traduz, dentro de determinados condicionalismos, pela junção gráfica, ou seja, pela sua transformação numa unidade monoverbal.

### 3.2.2 – Da unidade monoverbal à unidade pluriverbal

Restringindo-me aos aspectos em que o percurso sintetizado neste título se relaciona mais de perto com o tratamento informático, começarei por referir que o *corpus* (cf. 2.1), integrando-se no discurso escrito culto de carácter didáctico, pode considerar-se representativo das unidades pluriverbais propriamente linguísticas correntemente usadas nessa modalidade de expressão, compreendidas por um lado entre as colocações, “restrições de combinações estabelecidas pelo uso” (Corpas Pastor 2003: 74) e as paremias cujo estudo exigiria outro tipo de abordagem e a selecção de outras fontes.

Na dimensão metalinguística a amplitude dos usos aumenta, mas tem carácter esporádico, motivo por que unidades pluriverbais características da linguagem familiar e da linguagem popular estão escassamente representadas.

O tratamento informático constituiu o meio de exploração e de organização do material linguístico; a identificação, classificação e incorporação de unidades pluriverbais foi prevista desde o início, mas foi concretizada numa fase avançada do percurso, de tal modo que o *Índice Geral de Vocábulos* inclui as unidades pluriverbais, integradas segundo princípios gerais e sempre detectáveis através do sistema de remissões.

A primeira vantagem do tratamento informático decorre da necessidade de classificação exhaustiva das unidades da língua actualizada nos textos, o que, ultrapassando caracterizações muitas

---

significativo que, como observa Gómez Asencio (2004: 10), Nebrija escreva *por que* “na generalidade da obra”, mas *porque* “como uma só palavra” quando dá exemplo da conjunção composta (Nebrija 1492 [1981]: 199).

vezes decorrentes da selecção de casos claramente tipificados, constitui um desafio à capacidade de observação, reflexão, generalização e sistematização. Não é por isso ocasional que, embora o interesse e a teorização sobre unidades pluriverbais remonte à tradição gramatical clássica, as pesquisas conducentes à elaboração do núcleo fundamental das línguas, designadamente do francês e do português, tenham dado importantes contribuições sistematizantes na articulação entre concepções teóricas e funcionamento directamente observável<sup>24</sup>.

O primeiro dado útil fornecido pelo computador é a frequência com que uma sequência se repete, o que fornece a base que sustenta a hipótese de que essa sequência seja uma unidade pluriverbal: das 49 oc. de *exemplo*, no singular, em Barros, 23, ou seja 46,9% integram-se na expressão *per semelhante exemplo*. A expressão não é usada por nenhum outro autor, é uma peculiaridade de Barros que se integra no seu esquema expositivo e não apresenta nenhum outro traço que a identifique como uma unidade pluriverbal, ou seja, o indício não é confirmado.

A aplicação dos testes habituais que permitem avaliar a fixidez de uma sequência é imprescindível, mas a segunda das vantagens da exaustividade da análise que o tratamento informático faculta consiste em testar a eficácia dos testes.

Limitando a ilustração da afirmação ao teste de impossibilidade de inserção, constata-se que, das 61 oc. de *ainda que* no *corpus*, e das 35 oc. em Oliveira, 8 são interrompidas por *porém; ainda (porém) que* foi considerada uma variante atestada da locução, da qual constitui um sub-lemma. Internamente a *aver de* parece hoje impossível incluir uma palavra; mas as concordâncias comprovam a possibilidade contrária na época em estudo: “onde se *há sempre de seguir M.*” (GRE 37 10-11), “Outros muitos vocábulos se *haviam necessariamente de pegar aos Romanos*” (LRI 24 18-19); assim, ao sub-lemma *haver de*, incluído em *haver*, segue-se um sub-lemma de segundo grau, *haver(...)*

---

<sup>24</sup> Vejam-se Gougenheim *et al.* (1964), Gougenheim (1970) e Cruz (1987: 311-421).

*de*, o que deixa aberto o caminho a investigações posteriores que especificuem o tipo de palavras que podem interromper a sequência e em que condições a interrupção se dá.

A integração no contexto é indispensável para resolver homografias: “*a sua louvada velhice*” (OGR 3 8) inclui uma unidade pluriverbal, mas não “*Trouxe a sua casa*” (OGR 3 3). Muito mais do que isso, as concordâncias facultam o estabelecimento de relações de semelhança, de diferença e de gradação que se projectam em questões teóricas referentes a unidades pluriverbais. Assim, no exemplo apresentado por Barros (de verbos “neutros” usados com preposições), “*folgo no Inverno por amor do estudo*”, pode discutir-se se *por amor de* é uma unidade pluriverbal, caracterizada pela “não actualização dos elementos” (Gross 1996: 14) ou se, relativamente aos dois pólos do dilema explicitado por A. Martinet – “*Figements ou modèles?*” (Martinet 1979: 252), ela se situa no segundo, enquanto decorrente de um molde virtual disponível para um número ilimitado de preenchimentos; uma posição restritiva e mais de âmbito lexicológico que sintáctico optará pela negativa, tanto mais que o substantivo núcleo é usado no sentido próprio. Mas quando Oliveira usa a mesma sequência – “*o qual vocabolo [bombarda] chamarão assi por amor do som que ella lança*” (OGR 51 14-15) – não restam dúvidas de que se trata de uma unidade pluriverbal: na frase que serve de exemplo a Barros, *amor* é ainda *gosto por alguma coisa*, mas o contexto contém implícita uma relação entre *o estudo* e o *acto de folgar*, que torna o segundo a razão de ser do primeiro. Dá-se assim, simultaneamente, o deslizar do conteúdo para o âmbito da causa e a fixação da sequência.

As concordâncias facultam uma leitura vertical que selecciona a informação relativamente a uma unidade gráfica e, no âmbito do *corpus*, a esgota ao nível dos paradigmas, ao mesmo tempo que, horizontalmente, cada unidade é integrada nos contextos em que foi actualizada. Na primeira dimensão, esboça-se o quadro da produtividade das formas para a constituição de unidades pluriverbais e das relações entre polissemia e polilexicalidade; na segunda, tornam-se transparentes mudanças de conteúdo a partir das virtualidades semânticas das palavras seleccionadas, que passam a ser associadas, em bloco, a um efeito decorrente da sua actualização: *com tudo* só é

usado por Fernão d'Oliveira, embora com um número de ocorrências (33) que, comparado com o de *porem* (30, incluindo as 7 oc. de *ainda (porém) que*) não deixa dúvidas sobre a estabilidade da sequência; meio de referência ao contexto anterior e de ligação ao contexto seguinte, a expressão passa a incluir o matiz de indiferença perante o que se disse, face ao que vai dizer-se a seguir, ou a irrelevância do que acaba de dizer-se relativamente ao que se segue, ou seja, passa a veicular o nexos da relação entre dois factos ou duas séries de factos; por outras palavras, *com tudo* tornou-se uma conjunção poliverbal adversativa, o que pode ser ilustrado pelos dois passos seguintes:

"& assi como os gregos tem isto [acento tónico antes da antepenúltima sílaba, no caso das enclíticas] pode ser que também outras gentes o tem com"elles & *com tudo* se pronúnciao ambos aquelles acentos ou qual delles elles o saibão: eu não dou conta mais que escassamente da minha língua" (35 28-31); [sobre a variação na primeira pessoa do indicativo de *ser*: *som* ~ *sou* ~ *so* ~ *são*]: "no parecer da premeira pronúnciação com .o. & .m. que diz *som*. he o mui nobre johã de barros & a rezão que dá por si é esta: que de *som* mais perto vem a formaçã do seu plural o qual diz .*somos*. *com tudo* sendo eu moço pequeno fui criado em são domingos D"évora onde fazião zombaria de mim os da terra porque o eu assi pronúnciao segundo que o aprendera na beira." (71 29 – 72-2)

Quando se passa das concordâncias para o *Índice de Vocábulos*, onde a informação relativa a unidades pluriverbais é reunida, é possível delimitar questões que se situam na linha do tempo, por exemplo, em que grau a fixidez das unidades pluriverbais resiste às tendências linguísticas inovadoras na época, o que caracteriza os processos de extinção, que relações se estabelecem entre os processos de criação, fixação, expansão e o sistema de que emanam.

Mas todas as questões a que o tempo dá forma ou expressão ancoram no intemporal e a ele conduzem: há unidades pluriverbais que devem ser interrogadas à luz de processos cognitivos e que justificam que seja testada a hipótese MGMF ("more general more frequent") de Hagège (Peyraube 2002: 52); há dinâmicas de direcção recíproca entre a língua e o discurso que merecem ser reexaminadas, bem como a relação entre o individual e o colectivo: se certas expressões

se expandem e se tornam correntes é porque usá-las é uma forma de integração no grupo restrito ou numa comunidade mais ampla, e há o juízo subjectivo de que, parafraseando Fernão d'Oliveira a propósito de “dições usadas”, “quem não usa dellas é desentoado, fora do tom & musica dos nossos homens dagora” (52 9).

Se vier a provar-se, como penso, que ao longo do século XVI se dá a amplificação e a ramificação do inventário de unidades pluriverbais aptas à expressão da racionalidade e à articulação do pensamento abstracto, isso quererá dizer que, como escreveu ainda Fernão d'Oliveira, “as falas [...] sempre se conformão com os conceitos ou entenderes, juyzos e tratos dos homens.” (50 4-5).

## REFERÊNCIAS

- Ali, S. 1966. *Gramática Histórica da Língua Portuguesa*. São Paulo: Melhoramentos [6ª ed.].
- Bloomfield, L. 1933. *Language*. London: George Allen [reimp. 1961].
- Bourciez, E. 1967. *Éléments de Linguistique Romane*. Paris: Klincksieck [5.ème éd.].
- Câmara, J. Mattoso. 1959. *Princípios de Lingüística Geral*. Rio de Janeiro: Livraria Acadêmica [3ª ed.].
- Câmara, J. Mattoso. 1984. *Dicionário de Linguística e Gramática*. Petrópolis: Vozes [11ª ed.].
- Câmara, J. Mattoso. 1985. *História e Estrutura da Língua Portuguesa*. Rio de Janeiro: Padrão [4ª ed.].
- Corpas Pastor, G. 2003. *Diez Años de Investigación en Fraseología: Análisis sintáctico-semánticos, contrastivos y traductológicos*. Madrid: Iberoamericana.
- Cruz, M. L. Segura da. 1987. A norma lexicológica no tratamento do “corpus” de frequência. In: M. F. Bacelar do Nascimento et al. *Português Fundamental*. Vol II, T. I: *Inquérito de Frequência*. Lisboa: INIC/CLUL, 311-421.
- Cunha, C.; Lindley Cintra, L. F. 1984. *Nova Gramática do Português Contemporâneo*. Lisboa: Sá da Costa.
- De Mulder, W. 2001. La linguistique diachronique, les études sur la grammaticalisation et la sémantique du prototype: présentation. *Langue Française*. **130**: 8-32.
- García de Diego, V. 1951. *Gramática Histórica Española*. Madrid: Gredos [reimp. 1981].

- Gómez Asencio, J. J. 2004. Conjunciones, conjunciones compuestas y locuciones conjuntivas en antiguas gramáticas del Español. *Revista Española de Lingüística*. **34**: 1-37.
- Gougenheim, G. 1970. *Études de Grammaire et de Vocabulaire Français*. Paris: Picard.
- Gougenheim, G. et alii. 1964. *L'Élaboration du Français Fondamental*. Paris: Didier.
- Gross, G. 1996. *Les Expressions Figées en Français. Noms composés et autres locutions*. Paris: Ophrys.
- Hermann, J. 1963. *La Formation du Système Roman des Conjonctions de Subordination*. Berlin: Akademie-Verlag.
- Juilland, A.; Roceris, A. 1972. *The Linguistic Concept of Word. Analytic Bibliography*. The Hague/Paris: Mouton.
- Lausberg, H. 1963. *Linguística Românica*. Lisboa: Fundação Calouste Gulbenkian, 1974.
- Lloyd, P. 1987. *Del Latín al Español. Fonología y Morfología Históricas de la Lengua Española*. Madrid: Gredos, 1993.
- Maia, C. Azevedo. 1986. *História do Galego-Português, Estado linguístico da Galiza e do Noroeste de Portugal desde o século XIII ao século XVI*. Coimbra: INIC.
- Martinet, A. 1979. *Grammaire Fonctionnelle du Français*. Paris: Crédiff/Didier.
- Martinet, A. (Dir.). 1969. *La Linguistique, Guide Alphabétique*. Paris: Denoël.
- Mateus, M. H.; Brito, A. M.; Duarte, I.; Faria, I. H. ; Frota, S.; Matos, G.; Oliveira, F.; Vigário, M.; Villalva, A. 2003. *Gramática da Língua Portuguesa*. Lisboa: Caminho [5ª ed.].
- Meier, H. 1990. La etimología iberorromance en el siglo XX. *Boletim de Filologia*. **XXXI**: 5-20.
- Meillet, A. 1921. *Linguistique Historique et Linguistique Générale*. Paris: Champion [reimp. 1982].
- Menéndez Pidal, R. 1904. *Manual de Gramática Histórica Española*. Madrid: Espasa-Calpe [16ª ed., 1980].
- Mounin, G. 1974. *Dictionnaire de la Linguistique*. Paris: P.U.F.
- Nebrija, A de. 1492. *Gramática de la Lengua Castellana*. Estudio y edición de Antonio Quilis. Madrid: Editora Nacional [1981].
- Nunes, J. J. 1956. *Compêndio de Gramática Histórica Portuguesa*. Lisboa: Clássica [5ª ed.].
- 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.

- Pena, J. 1999. Partes de la morfología. Las unidades del análisis morfológico. In: I. Bosque; V. Demonte. *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa-Calpe, III, 4312-4328.
- Pereira, M. I. P. 1999. *O Acento de Palavra em Português. Uma análise Métrica*. Dissertação de doutoramento. Coimbra.
- Peyraube, A. 2002. L'évolution des structures grammaticales. *Langages*. **146** : 46-58.
- Pottier, B. 1976 [trad. esp.]. *Lingüística Moderna y Filología Hispánica*. Madrid: Gredos [1ª ed. fr.: 1968].
- Pottier, B. (Dir.). 1973. *Le Langage*. Paris: Retz.
- Quilis, A. 1981. Ver Nebrija, A. de. 1492.
- Silva, R. V. Matos 1989. *Estruturas Trecentistas, Elementos para uma gramática do português arcaico*. Lisboa: Imprensa Nacional.
- Williams, E. 1975 [trad. port.]. *Do Latim ao Português*. Rio de Janeiro: Tempo Brasileiro [1ª ed. ing.: 1961].
- Zuluaga, A. 1980. *Introducción al estudio de las expresiones fijas*. Frankfurt-am-Main: Peter Lang.

## APÊNDICE

Primeira sincronia

## 1. Fernão d'Oliveira

*Grammatica da lingoagem portugueja* -----1536-----OGR ----- 23 538 oc.---- 35,7 %

## 2. João de Barros

Texto contínuo da "Cartinha"----- 1539-----BCA

*Grammatica da lingua portugueja*----- -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

## 3. Pêro de Magalhães de Gândavo

*Regras que ensinam a maneira de escrever e Orthographia da lingua Portugueja [...]*-

-----1574----GRE

"Dialogo em defeção da lingua Portugueja" --1574 --GDL

----- 2 097 oc. ----- 3,1%

## 4. Duarte Nunes de Leão

*Orthographia da Lingoa Portuguesa* ----- 1576 ---LRT 8 698 oc. ---- 13,5 %

## 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.

Recensões

