

COMPLEXIDADE MORFOLÓGICA E CUSTOS DE PROCESSAMENTO LEXICAL

Alina VILLALVA *
Carina PINTO**

- RESUMO: Embora muitas palavras sejam formadas por mais de um constituinte morfológico, nem todas são habitualmente consideradas palavras complexas. No quadro da análise morfológica do Português, o conceito de ‘palavra complexa’ divide as palavras formadas por um radical, um constituinte temático e, eventualmente, um ou dois sufixos de flexão, das palavras formadas por estes mesmos constituintes e ainda os que são trazidos pelos processos de derivação, modificação ou composição. Esta distinção é redutora porque todas as palavras contêm algum grau de complexidade, mas não há instrumentos de análise que permitam medi-la. Procuraremos contribuir para a discussão da avaliação da complexidade das palavras com base em dados do processamento da leitura. A literatura apresenta diversas descrições que mostram que a estrutura morfológica desempenha um papel importante no processamento visual. Neste trabalho procuraremos encontrar novas evidências, testando hipóteses relacionadas com a composicionalidade das palavras. Usamos os métodos de *priming* morfológico e decisão lexical sobre três conjuntos de derivados em *-oso*: o primeiro é formado por estruturas composicionais, o segundo é constituído por palavras onde ocorre um alomorfe do sufixo e o terceiro contém palavras onde ocorre um alomorfe da forma de base. Os resultados obtidos confirmam que o processamento das palavras derivadas é sensível à sua estrutura morfológica e mostram também que as estruturas composicionais envolvem menores custos de processamento. Estas evidências permitem-nos propor critérios a ter em consideração na avaliação da complexidade das palavras.
- PALAVRAS-CHAVE: Complexidade morfológica. Processamento visual. Acesso lexical. Derivação.

Introdução

As línguas são sistemas complexos, formados por módulos complexos que acolhem diversos domínios complexos.³ A complexidade de um desses módulos, o léxico, deve ser avaliada em três planos distintos: o primeiro diz respeito ao cálculo do

* Universidade de Lisboa (UL), Faculdade de Letras, Centro de Linguística, Lisboa - Portugal. Departamento de Linguística Geral e Românica. alinavillalva@campus.ul.pt

** Universidade de Lisboa (UL), Faculdade de Letras, Centro de Linguística, Lisboa - Portugal. pintocarinaal@gmail.com

³ Veja-se Gong e Coupé (2011) para uma descrição da discussão ainda não concluída sobre a relação entre a complexidade das línguas e de outros sistemas complexos. Veja-se também a discussão do conceito de complexidade linguística em Mufwene (2012).

índice de complexidade de cada unidade lexical;¹ o segundo diz respeito ao cálculo da complexidade das palavras; e o terceiro diz respeito à avaliação do grau de complexidade das relações entre palavras (no domínio das relações morfológicas podemos considerar a partilha de um mesmo afixo contra a partilha de afixos concorrentes, por exemplo). Estas três dimensões são de grande relevância para a compreensão do processamento de palavras e do acesso lexical.

O presente trabalho procura contribuir para a discussão da avaliação da complexidade lexical no segundo domínio, olhando para a computação da complexidade de um tipo particular de palavras derivadas (i.e. adjetivos em *-oso*). O nosso objetivo último é o de avaliar se um maior custo para o processamento na leitura destas palavras decorre de um maior índice de complexidade lexical, ou não, mas, para o alcançar, é necessário saber (i) como se podem atribuir índices de complexidade às palavras e aos seus constituintes e (ii) como se podem medir os custos de processamento.

A discussão sobre a complexidade das palavras aqui apresentada tomará dados do Português Europeu² para um estudo de caso. A análise morfológica baseia-se na caracterização das unidades lexicais apresentada em Villalva e Silvestre (2015) e na tipologia das estruturas de palavras complexas de Villalva (2000, 2008). No que diz respeito à avaliação experimental e ao estudo do processamento, as referências principais são Taft e Forster (1975), Taft (1979, 1994), Rastle et al. (2000) e Rastle et al. (2004).

O conceito de complexidade

O conceito de complexidade surge frequentemente na investigação em diversos domínios do conhecimento, mas nem sempre é usado de forma rigorosa, o que pode levar a interpretações variáveis e imprevisíveis. O trabalho de Simon (1962) representa, provavelmente, o início do debate teórico nesta área. O autor defende que a complexidade é própria dos sistemas que assumem uma forma hierárquica: os sistemas complexos são formados por subsistemas, que por sua vez apresentam os seus próprios subsistemas, não havendo limite estabelecido para esta cadeia. Pode, assim, extrair-se que, independentemente do seu conteúdo específico, os sistemas hierárquicos são sempre sistemas complexos.

Sendo as línguas sistemas hierárquicos, formados por módulos hierarquizados que acolhem diversos domínios hierarquizados, pode concluir-se que os sistemas linguísticos e todos os seus subsistemas são sistemas complexos. No entanto, a aplicação do conceito de complexidade a cada disciplina do conhecimento suscita questões particulares, e o que se sabe sobre a sua aplicação ao domínio particular da linguística é ainda escasso.

¹ Assume-se aqui, à semelhança de Villalva e Silvestre (2015), que o léxico contém diversos tipos de unidades lexicais, nomeadamente radicais e afixos, palavras e sequências de palavras lexicalizadas.

² Considerando que a formação de adjetivos em *-oso* não apresenta diferenças significativas nas diversas variedades do Português contemporâneo, admitimos que os resultados do estudo apresentado não espelhem apenas o quadro do Português Europeu.

Nos termos de Gong e Coupé (2011, p.370), as línguas podem ser vistas como sistemas complexos adaptativos (em inglês, *complex adaptive systems*), porque envolvem um número significativo de unidades e módulos geradores de complexidade estrutural a vários níveis. As línguas são sistemas complexos adaptativos porque os seus agentes – os falantes - têm a capacidade de mudar o próprio sistema (cf. STEELS, 1997, 2000; BECKNER et al., 2009).

Uma outra reflexão sobre complexidade linguística é enunciada por Bane (2008, p.69) da seguinte forma:

- I. A complexidade linguística mantém-se constante ao longo do tempo;
- II. A gramática de uma dada língua não é mais complexa do que a gramática de qualquer outra língua.

Esta hipótese é interessante porque permite assumir que a variação linguística não afeta o grau de complexidade das línguas, embora possa configurá-la de forma distinta. Ainda que a ‘riqueza’ ou ‘pobreza’ dos sistemas morfológicos flexionais seja frequentemente invocada (por exemplo, na correlação com processos sintáticos), o que se pode presumir é que uma menor complexidade de um subsistema será compensada por uma maior complexidade de outro subsistema. No entanto, não existe, tanto quanto é do nosso conhecimento, forma de medir a “riqueza” ou, por outras palavras, a complexidade de nenhum dos componentes de qualquer um dos sistemas linguísticos.

Igualmente interessante é o contributo de Mufwene (2012), que descreve o conceito de complexidade no domínio da linguística, considerando os seguintes aspetos:

- I. a complexidade das unidades (e.g. o tamanho do inventário fonético) e regras de cada subsistema linguístico, a que dá o nome de complexidade *bit*;
- II. a complexidade interativa, que diz respeito às relações entre unidades e regras, no seio de cada um dos módulos, e ainda às relações entre os diversos módulos.

Esta é uma reflexão importante porque sugere que a complexidade das unidades é distinta da complexidade do modo como estas unidades interagem e estabelece ainda que a complexidade que caracteriza cada subsistema também é distinta, diferenciando-se igualmente da complexidade que caracteriza o sistema globalmente.

A conjugação destas posições permite construir uma hipótese de análise da complexidade linguística assente no seguinte conjunto de pressupostos:

1. é necessário eleger um ponto de equilíbrio no sistema linguístico, o que pode ser controlado a partir da consistência da amostra selecionada;
2. dado que a complexidade das línguas é constante, mas a dos seus subsistemas é variável, importa selecionar para cada análise apenas um subsistema e analisá-lo independentemente;

3. considerando que o cálculo da complexidade do sistema assenta em dois vetores (o valor intrínseco de cada unidade e o valor da sua interação com outras unidades numa estrutura), deve dirigir-se a análise para um único destes domínios.

Assim, no presente trabalho, mostramos os resultados de um estudo experimental realizado numa amostra homogênea de falantes do Português Europeu, que considera palavras complexas derivadas por intervenção do sufixo *-oso*. Assumimos aqui que os dados do processamento na leitura de palavras podem trazer informações relevantes para a compreensão da complexidade das estruturas morfológicas.

Não há muito trabalho feito no domínio da complexidade morfológica e o que existe diz respeito à flexão. O trabalho sobre complexidade morfológica tem sobretudo sido desenvolvido pelo Surrey Morphology Group,³ no âmbito de um projeto realizado entre 2009 e 2015 (cf. Morphological complexity: Typology as a tool for delineating cognitive organization). Têm ainda sido avançadas algumas tentativas de quantificação da complexidade das palavras considerando, por exemplo, a descrição das estruturas linguísticas e a sua função, a previsibilidade da sequência das palavras no discurso, as regularidades estruturais das línguas ou o tamanho do inventário fonético, entre outras (cf. GONG; COUPÉ, 2011). No entanto, estas medidas são vagas, tornando a medição da complexidade demasiado subjetiva. Neste quadro, a pesquisa sobre complexidade das estruturas morfológicas derivadas tem ainda um caráter pioneiro.

Por outro lado, a discussão do conceito de complexidade permite compreender que a distinção habitualmente considerada no domínio da análise morfológica entre palavras simples e palavras complexas é tão importante quanto simplista e redutora. Esta distinção permite colocar em campos opostos as palavras constituídas por um radical e os seus especificadores temáticos e flexionais (se forem requeridos pelo radical) e as palavras que integram um ou mais afixos (derivacionais ou modificadores) ou dois ou mais radicais. E esta é uma distinção de grande relevância: a identificação do conjunto das palavras simples dá a conhecer o conjunto de radicais inalisáveis que estão presentes no léxico desta língua, bem como as suas propriedades gramaticais e os seus valores semânticos básicos; e o conjunto de palavras complexas, que na realidade é um conjunto de conjuntos de palavras complexas, dá a conhecer os processos de formação de palavras em uso (passado ou contemporâneo) na língua. O problema que a distinção entre palavras simples e palavras complexas não resolve é o de que nem todas as palavras simples são igualmente simples, como nem todas as palavras complexas o são em igual medida. Nenhuma fonte de informação sobre o léxico do Português permite, tanto quanto é do nosso conhecimento, obter o conjunto de palavras simples desta língua. Também não é possível aceder ao conjunto de palavras complexas globalmente, mas é possível gerar subconjuntos de palavras complexas formadas por um dado prefixo ou por um dado sufixo, embora seja necessário filtrar os resultados, dado que a busca

³ Cf. <<http://www.smg.surrey.ac.uk/>>

é feita a partir de um critério ortográfico (e.g. palavras que terminam em *-idade*) e não de um critério morfológico (e.g. palavras que contêm o sufixo *-idade*). No caso das palavras simples, nenhuma estratégia ortográfica permite pesquisá-las. Vejamos, então, que tipo de problemas a aferição da complexidade das palavras simples e das palavras complexas pode suscitar.

A complexidade das palavras simples

Interessa-nos determinar se todos os radicais que podem ocorrer em palavras simples têm o mesmo índice de complexidade morfológica ou se existem fatores que afetam esse valor. A hipótese que colocamos é a de que o peso dos diversos constituintes não é constante e que é necessário identificar os fatores que fazem oscilar esse peso.

Um destes fatores dirá talvez respeito à especificação categorial dos radicais. Villalva e Silvestre (2015) estabelecem uma distinção básica entre radicais que podem ocorrer em palavras simples (cf. [*tóxic*]o) e aqueles que não podem (cf. [*aqu*]ífero). Os primeiros são classificados como predicados inerentemente intransitivos, embora também possam ocorrer como predicadores transitivos (cf. *neuro[tóxic]o*), como complementos (cf. [*toxic*]idade) e como modificadores (cf. [*tóxic*]o-dependente). Os segundos, habitualmente chamados radicais neoclássicos por se tratar de empréstimos extraídos de radicais latinos e gregos, são classificados como predicados inerentemente transitivos porque não podem ocorrer em palavras simples: estes radicais ocorrem como complemento em alguns tipos de derivados (cf. [*aqu*]oso) e ocorrem em compostos morfológicos, podendo ser o seu núcleo (cf. *aquí[fer]o*) ou o seu complemento (cf. [*aqu*]ífero).

Nesta secção consideraremos apenas o primeiro tipo de radicais, ou seja, aqueles que podem ocorrer em palavras simples⁴. Ainda que não haja valores de grandeza disponíveis, admite-se que a quantidade de palavras simples que pertence a uma única classe de palavras é considerável. Palavras como *perna*, *grossola* ou *pedir* pertencem inequivocamente ao domínio dos nomes, dos adjetivos e dos verbos, respetivamente. Os radicais que integram estas palavras terão, portanto, uma única especificação categorial:

- (1) [*pern*]_{RN} a
 [*gross*]_{RADJ} o/a
 [*ped*]_{RV} ir

Há, no entanto, um conjunto, tão ou mais importante do que o anterior, que é o dos radicais que podem ocorrer em diferentes palavras simples, como se verifica nos seguintes exemplos:

⁴ Os radicais transitivos serão referidos na secção seguinte, a propósito da discussão dos compostos morfológicos.

- (2) a. $[murch]_{RADJ}$ *o/a*
 $[murch]_{RV}$ *ar*
- b. $[dan\zeta]_{RN}$ *a*
 $[dan\zeta]_{RV}$ *ar*
- c. $[velh]_{RADJ}$ *o*
 $[velh]_{RN}$ *o*
- d. $[sec]_{RN}$ *a*
 $[sec]_{RV}$ *o*
 $[sec]_{RADJ}$ *o/a*
 $[sec]_{RV}$ *ar*

Muito possivelmente, nem todos estes radicais terão o mesmo estatuto lexical: alguns serão unidades lexicais de pleno direito, possuidoras de uma especificação categorial inequívoca (e.g. $[murch]_{RADJ}$; $[dan\zeta]_{RN}$; $[sec]_{RADJ}$), outros poderão ser radicais subespecificados (cf. $[velh]_{R[+N]}$); outros ainda poderão ser o resultado da intervenção de processos de conversão (cf. $[murch]_{RV}$; $[dan\zeta]_{RV}$; $[sec]_{RN}$; $[sec]_{RV}$) (cf. VILLALVA, 2013). Não é essa a questão que aqui importa discutir – o que importa saber é se esse diferente estatuto tem consequências para o cálculo da complexidade das palavras em que os radicais ocorrem. Assume-se aqui, como premissa, que o menor grau de complexidade pertence aos radicais intransitivos não-ambíguos (cf. $[pern]_{RN}$) e que o maior grau de complexidade se encontra nos radicais que foram objeto de formação por conversão (cf. $[murch]_{RV}$; $[dan\zeta]_{RV}$; $[sec]_{RN}$; $[sec]_{RV}$). Mas esta é uma assunção meramente estipulativa e que requer validação. Por esta razão, na constituição das listas que serviram de base ao trabalho experimental, selecionámos duas séries (1ª e 2ª) de radicais inerentemente intransitivos e inequivocamente nominais (e.g. *veneno*, *mentira*, *luxo*, *conflito*). Procurámos, desse modo, controlar fatores exógenos à questão que nos propusemos analisar.

Por outro lado, alguns radicais apresentam uma única forma, qualquer que seja a estrutura em que ocorrem, mas outros não. A existência de formas alternantes pode estar relacionada com questões morfofonológicas (cf. 3a) ou podem decorrer de circunstâncias lexicais, como a introdução de empréstimos neoclássicos (cf. 3b):

- (3) a. *cão*
canil
- b. *veia*
venoso

As duas séries de dados antes referidas (1ª e 2ª) são formadas por radicais que não têm formas alternantes.

Em suma, no trabalho experimental aqui relatado procurámos controlar os fatores que podem afetar o nível de complexidade das palavras simples, tendo considerado aspectos que dizem respeito apenas aos radicais.

A complexidade das palavras complexas

A avaliação da complexidade das palavras complexas eleva o nível de dificuldade já antes referido a um patamar ainda desconhecido, dado que, como referimos anteriormente, não há trabalho feito neste domínio nem para o Português nem para outras línguas. Neste domínio, as linhas de investigação em aberto passam, pelo menos, pela aferição comparada do peso da afixação face à composição,⁵ da prefixação face à sufixação (ou outros tipos de afixação, quando disponíveis); pela comparação de configurações que envolvem diversos níveis de encaixe e diversas interações entre prefixação, sufixação e composição; e ainda pela avaliação da produtividade dos diversos processos. Relevantes para o presente trabalho são, no entanto, apenas as questões que dizem respeito ao estatuto dos radicais inerentemente transitivos e à alomorfia nos sufixos derivacionais.

A atestação de radicais inerentemente transitivos no Português aumenta exponencialmente a partir do século 18, a par do desenvolvimento de terminologias científicas e técnicas. As palavras que os transportam entram frequentemente como empréstimos, por exemplo do Francês (cf. *termómetro*), fazendo elas próprias uso de empréstimos extraídos do léxico do Latim e do Grego Antigo. Compare-se, por exemplo, a forma das palavras em (4) e a plausibilidade da sua compreensão pelos falantes. O radical [*pedr*], que ocorre em (4a) é semanticamente equivalente ao radical de origem latina [*petr*], que ocorre em (4b) e ao radical de origem grega [*lit*] que ocorre em (4c). Mas só o primeiro ocorre numa palavra simples (i.e. *pedra*) à qual está associado um valor semântico acessível aos falantes. Assim, presume-se que a semântica das palavras que contêm este radical seja mais transparente e que o processamento dos seus derivados seja composicional. Quanto aos outros dois radicais, nenhum deles pode ser conhecido fora do contexto das palavras complexas onde ocorrem e a interpretação dessas palavras complexas pode, aliás, não ser composicional.

- (4) a. *pedra*
 pedreira
- b. *pétreo*
 petrificar
- c. *litografia*
 megalítico

⁵ Inclui-se aqui apenas a composição morfológica (que envolve a sequencialização de radicais e não de palavras). A composição morfossintática e a composição sintática não são processos morfológicos de formação de palavras.

As palavras que usamos para construir a 3ª série de dados contêm variantes de radicais que só ocorrem em palavras complexas (e.g. *aquoso* / água; *medroso* / medo).

A existência de formas alternantes também pode afetar os afixos. No caso específico do sufixo *-oso*, verifica-se a existência de uma forma alomórfica (i.e. *-uoso*) cuja distribuição não é determinável no Português contemporâneo. Consequentemente, as palavras que contêm esta variante do sufixo estão lexicalizadas. A segunda série dos dados usados no nosso estudo contêm este alomorfe do sufixo (cf. *conflituoso*, *luxuoso*).

A hipótese de que partimos foi a de que as palavras que contêm formas alternantes do radical (3ª série) ou do sufixo (2ª série) revelam maiores custos de processamento e que esse seja um índice fiável para o cálculo da complexidade morfológica das palavras complexas.

O cálculo da complexidade morfológica

Provavelmente, o grau de complexidade de cada palavra não corresponde a uma simples soma aritmética dos índices de complexidade associados a cada um dos seus constituintes, mas sim a uma função algorítmica que estamos longe de poder construir. Há, portanto, dois campos de trabalho a explorar: por um lado, é necessário identificar o que pode determinar o índice de complexidade de cada um dos constituintes morfológicos e o algoritmo que pode permitir calcular o índice de complexidade de cada palavra; por outro é necessário encontrar instrumentos de validação das estipulações construídas pela análise linguística.

Nas secções seguintes exploraremos a possibilidade de encontrar uma forma de validação na análise dos dados sobre processamento morfológico e lexical obtidos experimentalmente. O trabalho que aqui relatamos contribui para esta discussão sobre o cálculo da complexidade morfológica ao olhar para o processamento de palavras complexas composicionais e palavras complexas lexicalizadas, subdividindo estas últimas em dois subconjuntos (alomorfia da base ou alomorfia do sufixo).

Complexidade morfológica e processamento

Vários estudos experimentais têm vindo a demonstrar que a natureza da estrutura morfológica das palavras tem um papel a desempenhar no reconhecimento visual das palavras e no acesso lexical. Alguns destes estudos, realizados no âmbito de uma vasta amplitude de paradigmas estão sumariados em McQueen e Cutler (1997), estando disponível uma descrição mais recente do estado da arte em Pinto (2017).

A experimentação desenvolvida no quadro do paradigma de *priming* tem ajudado na identificação de propriedades e na verificação do seu papel na ativação lexical. Os resultados de estudos como os de Frost et al. (2005) e Velan e Frost (2011) sugerem a existência de dois processos organizados hierarquicamente:

1. o estágio morfo-ortográfico caracteriza as fases mais precoces do reconhecimento da palavra – o processo é ativado quando o *input* visual (i.e. a palavra lida) é complexa e composicional – nesta fase é ativado o reconhecimento das formas (base e afixo);
2. o estágio morfo-semântico é ativado posteriormente – nesta fase, as formas anteriormente reconhecidas são semanticamente interpretadas.

Há, no entanto, outros trabalhos (cf. TAFT; FORSTER, 1975; TAFT, 1979, 1994) que sugerem que os constituintes morfológicos são armazenados no léxico mental como entradas lexicais distintas e que o acesso ao significado se produz através da compreensão do significado de cada unidade. Estudos de *priming* morfológico, como o de Laudanna e Burani (1995) ou o de Järvikivi et al. (2006) mostraram que as palavras que contêm afixos com diversos alomorfes apresentam uma latência de resposta mais longa. Provavelmente, o processamento visual das palavras morfológicamente complexas pode ser afetado por diversos fatores deste tipo, ou seja, por propriedades dos constituintes morfológicos, nomeadamente a existência de alomorfia e a sua produtividade. Características deste tipo tornarão os afixos mais salientes, o que poderá afetar a probabilidade de serem ativados autonomamente durante o processo de reconhecimento visual das palavras.

Muitas das evidências de que a estrutura morfológica desempenha um papel no processamento visual das palavras complexas é compatível tanto com o modelo de processamento global das palavras, como pelo modelo de processamento autónomo dos constituintes morfológicos das palavras. O que ainda não é claro é quando e como os processos de análise morfológica entram em ação no processo de compreensão.

Avaliação experimental

Como contributo para a discussão sobre complexidade morfológica, e de forma a testar algumas premissas sobre o processamento morfológico e a complexidade dos constituintes morfológicos, foram realizadas quatro experiências: uma de decisão lexical e três de *priming* morfológico. Com a realização destes dois tipos de procedimentos experimentais pretendemos verificar se a introdução de um *prime* que corresponde à forma de base de um alvo derivado facilita ou não o processamento visual destas palavras complexas. Uma resposta positiva poderá indiciar que a análise morfológica está disponível e que facilita a compreensão dos derivados.

Por outro lado, a realização do mesmo teste de *priming* com três diferentes intervalos de exposição ao *prime* permitirá verificar se o processamento morfológico é realizado numa fase mais inicial ou num momento mais tardio do processamento visual. Também se deve admitir a hipótese de que o processamento morfológico se verifica em diversos estágios do processamento das palavras, em função da complexidade dos constituintes das palavras complexas.

Por último, e considerando que os dados morfológicos testados são formados por três conjuntos de palavras derivadas pelo mesmo sufixo, sendo que um dos grupos integra estruturas composicionais, enquanto os dois outros integram alomorfes de um dos constituintes, coloca-se como hipótese de partida que palavras com uma estrutura composicional sejam acedidas mais rapidamente do que palavras com uma estrutura perturbada pela ocorrência de alomorfes.

Dados morfológicos

Esta experiência foi planeada a partir do processo de adjetivalização denominal em *-oso*. Para a seleção das palavras foi tido em consideração o número de sílabas, selecionando-se formas derivantes com duas ou três sílabas (cf. *veneno*, *luxo*, *medo*) e derivados com quatro ou cinco sílabas (cf. *venenoso*, *luxuoso*, *medroso*). Foi também realizado um levantamento da frequência de ocorrência das palavras no Português Europeu, através da base de dados do CRPC, que permitiu selecionar formas derivantes com frequência de ocorrência altas. O *Corpus de Referência do Português Contemporânea* (CRPC), é um corpus textual eletrônico, que integra mais de 1,6 milhões de palavras provenientes de transcrições de gravações e registros orais da variedade europeia e de outras variedades do Português. Foi produzido pelo CLUL e está disponível em: <www.clul.ul.pt/pt/recursos/183-reference-corpus-of-contemporary-portuguese-crpc>.

Os derivados têm sempre uma frequência de ocorrência baixa, porque essa é uma característica inerente à sua própria natureza de palavras derivadas. Por último, e tendo em conta a estrutura morfológica dos derivados, foram organizadas três séries de dez palavras, que correspondem às seguintes condições:

- 1^a derivados com uma estrutura composicional (e.g. *venenoso*) - a base é um radical de um nome (i.e. *veneno*), o sufixo tem a forma *-oso* e o derivado é um adjetivo (na forma do masculino e singular), que pode ser parafraseado por uma expressão do tipo $X]_{RN} \textit{oso} = \textit{'que tem X']}_{ADJ}$ (i.e. *venenoso* = ‘que tem veneno’);
- 2^a derivados com alomorfia do sufixo (e.g. *luxuoso*) – a base é o radical de um nome (i.e. *luxo*), o sufixo tem a forma *-uoso* e o derivado é um adjetivo (na forma do masculino e singular), que pode ser parafraseado por uma expressão do tipo $X]_{RN} \textit{oso} = \textit{'que tem X']}_{ADJ}$ (i.e. *luxuoso* = ‘que tem luxo’);
- 3^a derivados com alomorfia da base (e.g. *medroso*) – a base é um alomorfe de um nome (i.e. *medr-* vs. *med-*), o sufixo tem a forma *-oso* e o derivado é um adjetivo (na forma do masculino e singular), que pode ser parafraseado por uma expressão do tipo $X]_{RN} \textit{oso} = \textit{'que tem X']}_{ADJ}$ (i.e. *medroso* = ‘que tem medo’).

Metodologia

A avaliação da hipótese acima enunciada levou à concepção e aplicação de diversas experiências que envolvem tarefas de decisão lexical e *priming* morfológico. Nas seções seguintes apresenta-se a caracterização da amostra dos informantes, a descrição dos procedimentos experimentais adotados e a discussão dos resultados obtidos.

Amostra

Todas as experiências que integram este estudo foram realizadas na população portuguesa, tendo sido selecionada uma amostra de indivíduos saudáveis e estudantes do ensino universitário, na região de Lisboa e de Leiria. Este estudo obteve parecer favorável da Comissão Nacional de Proteção de Dados (Autorização nº 7788/2013).

Esta amostra foi sujeita a critérios de exclusão passíveis de provocar alterações cognitivo-linguísticas, como as seguintes:

- 1) Existência prévia de Acidente Vascular Cerebral (AVC);
- 2) Existência de Epilepsia;
- 3) Ocorrência de Traumatismo Crânio-Encefálico (TCE);
- 4) Depressão *major*/Esquizofrenia diagnosticadas por médico especialista;
- 5) Alterações visuais não corrigidas;
- 6) Alterações auditivas não corrigidas;
- 7) Alterações da linguagem escrita caracterizadas no DSM IV;⁶
- 8) Doença médica grave que potencie o aparecimento de alterações linguísticas;
- 9) Toxicod dependência/Alcoolismo;
- 10) Bilinguismo.

No total, foram recolhidos os dados de 116 sujeitos. Foram obtidos os respectivos consentimentos informados. Todos os sujeitos foram submetidos a um despiste de alterações da linguagem oral e escrita. Nenhum dos sujeitos apresentou qualquer comportamento desviante.

Procedimento

O presente trabalho relata quatro experiências: uma tarefa de decisão lexical e três tarefas de *priming* morfológico com decisão lexical, que utilizaram três diferentes tempos de exposição ao *prime*: 50 milissegundos (= ms), 100ms e 150ms.

⁶ DSM IV (*Diagnostic and Statistical Manual of Mental Disorders*) é o manual da Associação Americana de Psiquiatria, consensualmente utilizado para o diagnóstico de perturbações mentais.

As experiências foram construídas com recurso ao software E-Prime® 2.0. Os estímulos visuais foram apresentados num computador Compaq Presario®. A tabela seguinte mostra a distribuição de sujeitos,⁷ por experiência:

Tabela 1 – Distribuição de sujeitos pelas quatro experiências

Apresentação do <i>prime</i> durante 50ms	32
Apresentação do <i>prime</i> durante 100ms	30
Apresentação do <i>prime</i> durante 150ms	27
Decisão lexical	27

Fonte: Elaboração própria.

No caso da decisão lexical, o teste incluiu a lista experimental formada pelas trinta palavras apresentadas como alvo na prova de *priming* (exs. *venenoso*, *luxuoso*, *medroso*). No caso do *priming* morfológico, os testes incluíram três séries de dez pares de palavras, sendo o *prime* fornecido pela forma derivante do alvo (exs. *veneno-venenoso*; *luxo-luxuoso*; *medo-medroso*). Nestas experiências, foram usadas, como *fillers*, pseudo-palavras construídas especialmente para este fim. As pseudo-palavras que utilizámos foram construídas com base num procedimento sistemático de substituição das sílabas das palavras alvo (ex. *bexigoso* – *goxiboso*). Os *fillers* são utilizados, canonicamente, para motivar a decisão lexical e para escamotear o objetivo do teste, desempenhando assim a função de distratores.

Todas as experiências se iniciaram com a apresentação de seis itens de treino (i.e. palavras que não fazem parte do teste). Tanto estes itens de treino como os itens experimentais (i.e. as palavras a testar) surgiam no centro do écran, na fonte Times New Roman, tamanho 18, e em letra minúscula. Todos os itens surgiam a preto sobre um fundo branco e eram precedidos por uma máscara (+), que era também apresentada no centro do écran, durante 500ms, e que servia como ponto de fixação. No caso da experiência de *priming* morfológico, após a exibição da máscara surgia o *prime* (durante 50, 100 ou 150ms), imediatamente seguido pela palavra alvo, que ficava disponível no écran até à tomada da decisão lexical por parte dos sujeitos. A tomada da decisão lexical era concretizada recorrendo ao teclado do computador.

Resultados

Para a análise dos dados obtidos foi utilizado o software SPSS® versão 20, que serviu de base à análise estatística. De acordo com os procedimentos habituais, foi feita uma limpeza dos dados obtidos que levou à exclusão de *outliers* considerando quatro critérios:

⁷ A amostra recolhida é uma amostra de conveniência. A recolha de dados foi feita com 32 sujeitos por experiência, mas o tratamento dos dados levou à exclusão de alguns sujeitos por diversas razões, como, por exemplo, a percentagem de erros nas respostas.

- 1) exclusão das respostas erradas;
- 2) substituição dos valores díspares superiores a 10.000ms e inferiores a 250ms pela média do sujeito na condição;
- 3) substituição dos valores superiores à média em $\pm 2,5$ desvios padrão pela média do sujeito na condição;
- 4) exclusão de todos os valores superiores a 2000ms.

Na tabela seguinte pode verificar-se a percentagem de dados excluídos, tendo em consideração os quatro critérios acima enunciados:

Tabela 2 – Percentagem de dados excluídos

	Tempo de exposição ao <i>prime</i>	Respostas erradas	Valores díspares	Valores > média $\pm 2,5$ DP	Valores >2000ms
Decisão lexical		4,50%	0 %	2,33%	5,04%
<i>Priming</i> morfológico	50 ms	6,50%	0%	2,67%	0,44%
	100 ms	2,00%	0,11%	2,15%	4,08%
	150 ms	7,41%	0,27%	2,40%	6,41%

Fonte: Elaboração própria.

Com esta limpeza pretendeu-se garantir um maior grau de homogeneidade na qualidade das respostas, dado que esta pode ser posta em causa por erros de execução ou por variabilidade inerente aos elementos da amostra.

Foram ainda realizados testes de normalidade da amostra, que não revelaram uma distribuição normal. Esta é uma circunstância esperada, uma vez que a observação incide sobre tempos de reação. Por esta razão, os dados recolhidos apresentam um limite à esquerda, ou seja, existe sempre o limite zero (0) que impede a simetria. Assim, para a análise dos dados serão utilizados testes estatísticos não paramétricos.

Em seguida, serão apresentados os resultados obtidos para cada uma das experiências. Na primeira, discutir-se-á o papel do *priming* no tempo de reação associado à tarefa de decisão lexical; na segunda serão confrontados os resultados obtidos para as três condições morfológicas (‘derivação regular’, ‘alormfia do sufixo’ e ‘alormfia da base’).

Avaliação do papel da condição morfológica

Nesta experiência, avaliam-se os custos de processamento visual das palavras associados às três condições morfológicas: derivação regular, derivação com alormfia do sufixo e derivação com alormfia da base. Os resultados são apresentados por tarefa:

decisão lexical, *priming* com decisão lexical (50ms), *priming* com decisão lexical (100ms) e *priming* com decisão lexical (150ms):

Tabela 3 – Dados descritivos das quatro experiências

<i>Exposição ao prime</i>		Derivação regular	Alomorfia do sufixo	Alomorfia da base
50ms	Média	782	908	861
	(DP)	(220,21)	(310,37)	(229,96)
	Mediana	713,00	795,00	812,50
	1º quartil	662,50	688,00	710,00
	3º quartil	901,00	1075,50	970,00
100ms	Média	879	950	920
	(DP)	(282,86)	(321,40)	(272,15)
	Mediana	802,00	861,00	856,00
	1º quartil	682,00	722,00	731,00
	3º quartil	994,00	1178,00	1028,00
150ms	Média	875	970	921
	(DP)	(328,80)	(342,13)	(290,58)
	Mediana	761,50	886,00	858,00
	1º quartil	659,00	709,00	704,50
	3º quartil	1019,00	1176,00	1032,00
Decisão lexical	Média	932,77	1007,66	993,13
	(DP)	(317,93)	(356,45)	(322,84)
	Mediana	854,00	893,00	896,00
	1º quartil	697,50	727,50	738,00
	3º quartil	1099,00	1270,00	1208,00

Fonte: Elaboração própria.

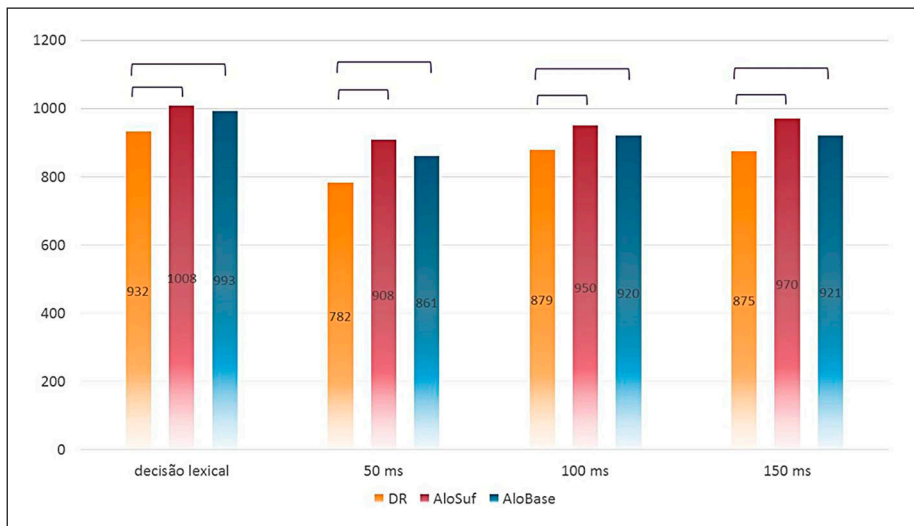
Estas quatro experiências apresentam resultados bastante homogêneos, tendo a condição ‘*derivação regular*’ (=DR) um comportamento sempre distinto das outras duas.

No gráfico 1, pode ver-se que existem diferenças estatisticamente relevantes entre tempos de reação (assinaladas por chavetas), nos seguintes casos:

1. experiência da decisão lexical
 derivação regular - alomorfia do sufixo ($U=24367,5;p=0,034$)
 derivação regular - alomorfia da base ($U=27062,5;p=0,018$)
2. experiência de priming morfológico – 50ms
 derivação regular - alomorfia do sufixo ($U=34023,5;p=0,000$)
 derivação regular - alomorfia da base ($U=34122;p=0,000$)
3. experiência de priming morfológico – 100ms
 derivação regular - alomorfia do sufixo ($U=34308;p=0,009$)
 derivação regular - alomorfia da base ($U=36190;p=0,008$)

4. experiência de priming morfológico – 150ms
 derivação regular - alomorfa do sufixo ($U=23949,5;p=0,000$)
 derivação regular - alomorfa da base ($U=25324;p=0,004$)

Gráfico 1 – Diferenças significativas nas diferentes experiências



Fonte: Elaboração própria.

Os resultados mostram que as estruturas morfológicas composicionais são processadas sempre em menor tempo e que, inversamente, as estruturas que comportam um alomorfe do sufixo são aquelas que requerem maior tempo de processamento.

Avaliação do papel do *priming*

Os resultados seguintes avaliam os custos do processamento em função da existência ou inexistência de *priming* e dos diversos tempos de exposição ao *prime*. Os resultados são apresentados por condição.

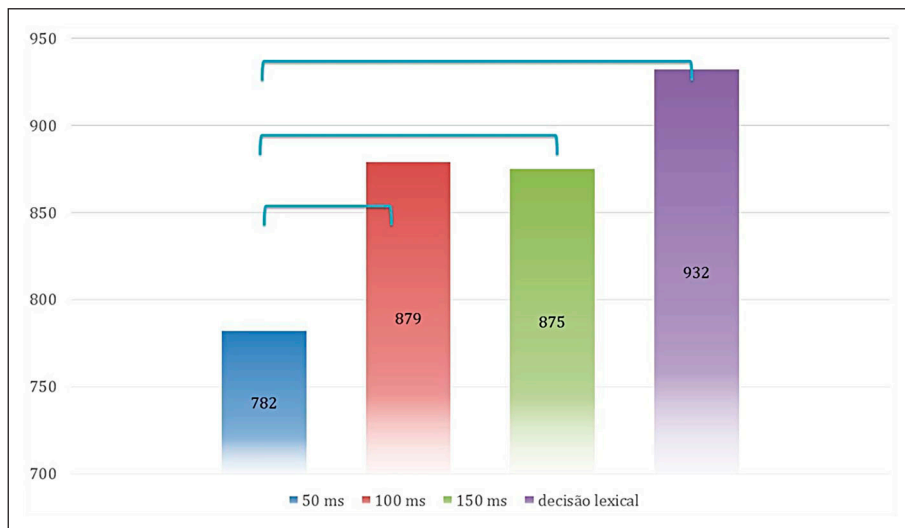
Derivação regular (ex. orgulho/orgulhoso)

Em relação às diferenças estatísticas, verificamos que são significativas nos tempos de reação obtidos nos seguintes casos:

1. duração do *prime* por 50ms – duração do *prime* por 100ms ($U=34644;p=0,000$)
2. duração do *prime* por 50ms – duração do *prime* por 150 ms ($U=34196,5;p=0,005$).
3. duração do *prime* por 50ms – decisão lexical ($U=28155,50;p=0,000$).

O gráfico 2 permite visualizar essas diferenças:

Gráfico 2 – Diferenças significativas na condição ‘derivação regular’



Fonte: Elaboração própria.

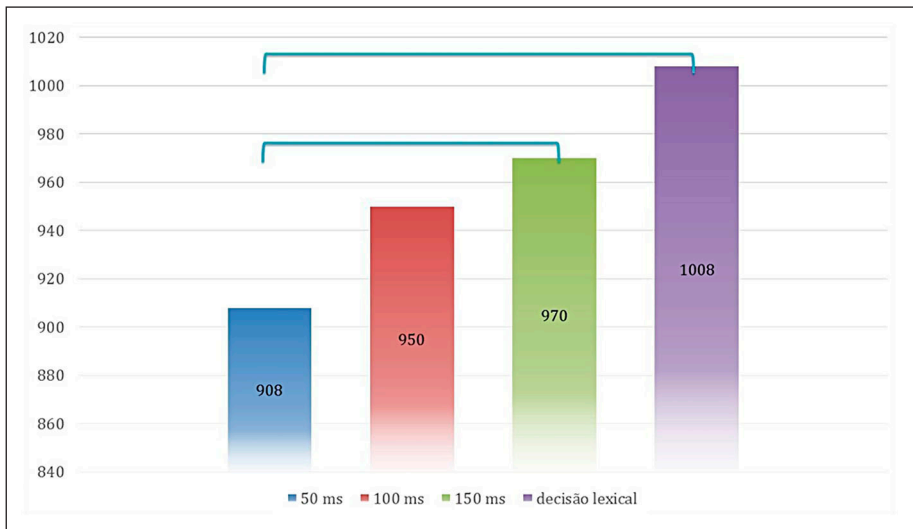
Alomorfia do sufixo (defeito/defeituoso)

No que respeita às diferenças estatísticas, relativas à condição ‘Alomorfia do sufixo’ (=alosság), verifica-se que, tal como na condição anterior, existem diferenças significativas nos tempos de reação obtidos, mas, neste caso, a diferença não é significativa quando se contrasta o tempo de exposição ao *prime* de 50ms com o tempo de exposição ao *prime* por 100ms:

1. duração do *prime* por 50ms – duração do *prime* por 150ms ($U=30093; p=0,034$).
2. duração do *prime* por 50ms – a decisão lexical ($U=26072,5; p=0,002$).

O gráfico 3 permite visualizar essas diferenças:

Gráfico 3 – Diferenças significativas na condição ‘derivação com alomorfia do sufixo



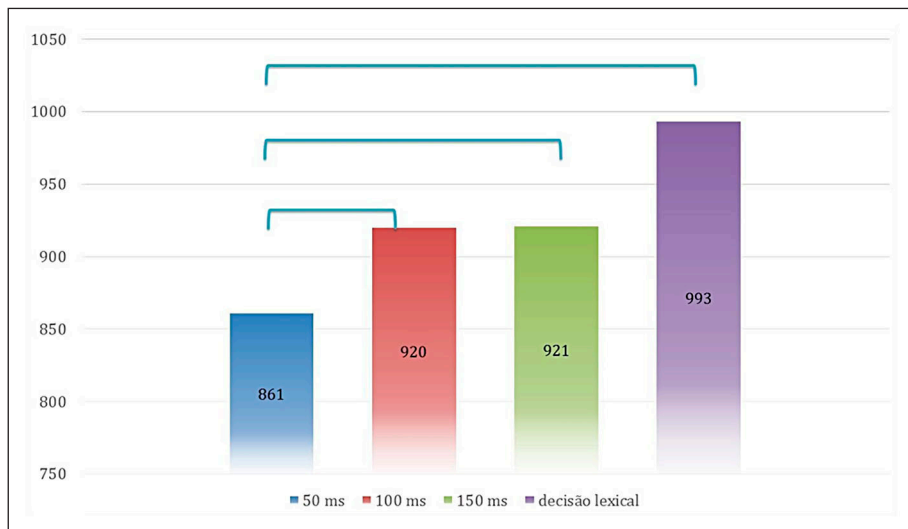
Fonte: Elaboração própria.

Alomorfia da base (lume/luminoso)

No gráfico seguinte apresentam-se os valores relativos à condição ‘Alomorfia da base’ (=alobase), que repetem as tendências encontradas anteriormente. As diferenças estatísticas são semelhantes às da condição ‘derivação regular’, ou seja, existem diferenças significativas nos tempos de reação obtidos nos seguintes casos:

1. duração do *prime* por 50ms - duração do *prime* por 100ms ($U=36610;p=0,007$).
2. duração do *prime* por 50ms - duração do *prime* por 150ms ($U=30360;p=0,03$).
3. duração do *prime* por 50ms - decisão lexical ($U=66474;p=0,038$).

Gráfico 4 – Diferenças significativas na condição ‘alomorfa da base’



Fonte: Elaboração própria.

Leitura dos resultados

Na primeira experiência, onde se avaliaram os custos de processamento visual associados às três condições morfológicas (derivação regular, derivação com alomorfa do sufixo e derivação com alomorfa da base) verificou-se que o processamento visual das estruturas composicionais (geradas regularmente) é mais rápido do que os restantes e verificou-se igualmente que o processamento visual de estruturas com alomorfa do sufixo é o que exige um maior esforço, o que se pode dever a uma menor saliência (talvez semântica, talvez formal, mas os dados não permitem extrair conclusões seguras) dos afixos relativamente aos radicais.

No que diz respeito ao segundo conjunto de experiências, que avaliaram os custos do processamento visual em função da existência ou inexistência de *priming* e dos diversos tempos de exposição ao *prime*, verificou-se que o tempo de reação diminui quando existe *priming*, ou seja, quando se induz a análise morfológica dos derivados, o custo de processamento diminui. Por outro lado, a variação nos tempos de exposição ao *prime* que foram testados (50ms, 100ms e 150ms) também gerou resultados interessantes, dado que o tempo de reação aumenta com o aumento do tempo de exposição ao *prime*, qualquer que seja a condição morfológica em questão.

A leitura combinada de todas estas experiências permite extrair que a leitura de palavras derivadas é facilitada quando a análise morfológica é induzida pela forma derivante e que esse efeito é mais evidente numa fase inicial do processamento destas palavras.

Considerações finais

Este trabalho surgiu de uma reflexão sobre a natureza das estruturas morfológicas e sobre a fragilidade do conhecimento atual sobre o seu grau de complexidade. A oposição entre palavras simples e palavras complexas capta uma distinção básica entre palavras que não são geradas por processos de formação de palavras e palavras derivadas, modificadas ou compostas, mas é fácil demonstrar que nem todas as palavras simples são igualmente simples e que nem todas as palavras complexas são complexas de igual forma.

Do ponto de vista da análise morfológica, é relativamente fácil identificar fatores de avaliação da complexidade, relativamente a cada um dos constituintes das palavras, mas o cálculo do índice de complexidade de cada uma delas não é tão evidente. É esta a razão que leva a recorrer a dados do processamento morfológico como potenciais aferidores das estipulações teóricas. A busca dos procedimentos adequados e a capacidade de recolha de tratamento dos dados impede-nos, para já, de apresentar hipóteses mais abrangentes. Para este trabalho, limitamo-nos a testar o processamento de adjetivos derivados em *-oso*, contrapondo estruturas composicionais (cf. *venenoso*) a estruturas onde ocorrem alomorfes do sufixo (cf. *luxuoso*) e ainda a estruturas onde ocorrem alomorfes da base (cf. *medroso*).

Os resultados obtidos permitem-nos construir a hipótese de que o processamento visual de palavras complexas derivadas por sufixação, no Português, é sensível a propriedades da sua estrutura morfológica. Por um lado, verificou-se que as estruturas composicionais têm um custo de processamento menor do que as estruturas perturbadas por fatores como a alomorfia do sufixo ou a alomorfia da base. E verificou-se, aliás, que o fator de perturbação ‘alomorfia do sufixo’ é aquele que acarreta maiores custos de processamento, sobretudo quando comparado com o fator ‘alomorfia da base’. Esta constatação é, de algum modo, contra-intuitiva, mas pode indiciar que as palavras que integram alomorfes da base (cf. *arenoso*) estão mais próximas da lexicalização que conduz a um acesso lexical direto, do que as palavras que incluem alomorfes dos sufixos (cf. *luxuoso*). Estas últimas, que são as mais pesadas em termos de processamento, parecem manter-se analisáveis, mas requerem um esforço suplementar para que o sufixo seja devidamente reconhecido.

Um outro aspecto que surge da análise dos resultados obtidos diz respeito ao papel do *priming*. Com efeito, dado que o tempo de reação diminui com a existência de exposição a um *prime*, pode admitir-se que o processamento de palavras derivadas por sufixação envolve sempre tarefas de análise morfológica, dado que ela é facilitada pela exposição à forma derivante. Por outro lado, constata-se que o tempo de reação aumenta na proporção direta do aumento do tempo de exposição ao *prime*, o que parece indiciar que o fator de facilitação oferecido pela presença de um *prime* é mais relevante numa janela temporal mais precoce do que nas fases mais tardias do processamento.

Em suma, os resultados obtidos corroboram a hipótese de que as palavras complexas composicionais, que contêm radicais inerentemente intransitivos, ou seja, radicais

que podem ocorrer em palavras simples, e a forma canônica de um sufixo (neste caso, *-oso*) apresentam menores custos de processamento e, portanto, um menor índice de complexidade do que as palavras complexas lexicalizadas. O que os dados também parecem mostrar, e que vai contra a hipótese inicialmente considerada, é que a perda de composicionalidade que se deve a problemas no sufixo é mais gravosa, em termos de processamento, do que a que afeta o radical. Por outras palavras, quando os sujeitos identificam um radical inerentemente transitivo, desistem da decomposição morfológica da palavra, passando a processá-la como uma palavra simples; quando o radical é identificado como o radical de uma palavra simples, mas o sufixo corresponde a um alomorfe com uma distribuição aleatória, a tentativa de decomposição morfológica é mantida durante mais tempo, mas acabará também por ser abandonada. Presume-se, assim, que o índice de complexidade das palavras derivadas é sensível ao lugar onde ocorre o fator de lexicalização.

Não havendo dados comparáveis sobre o Português ou sobre outras línguas, esperamos que os nossos resultados possam vir a ser discutidos futuramente.

VILLAVA, A.; PINTO, C. Morphological Complexity and Lexical Processing Costs. *Alfa*, São Paulo, v.62, n.1, p.151-172, 2018.

- *ABSTRACT: Although many words are formed by more than one morphological constituent, not all of them are complex words. In the framework of morphological analysis, the term 'complex word' usually sets apart words formed by a root, a stem index and inflectional affixes, from words formed by derivation, modification or compounding. This distinction is quite simplistic since all words display a certain degree of complexity. In the literature, there are abundant claims that morphological structure plays an important role in word processing, but the level of morphological complexity is never taken into account. In this paper, we will try to contribute to the discussion of the role of morphological structure for written word processing, namely by taking into consideration the level of morphological complexity of a particular set of Portuguese derived words. We will look at the results of a priming experiment involving a lexical decision task on three sets of derivatives in -oso: the first set is formed by compositional structures; in the second, we have included words that display an allomorph of the suffix (i.e. -oso ~ -uoso); and, in the third set, we gathered words that make use of an allomorphic base. The results of this experiment confirm that derived word processing is sensitive to the morphological structure of the word and they also show that compositional structures involve lower processing costs. Hence, these results allow us to claim that the degree of morphological complexity of complex words needs to be considered for the study of written word processing.*
- *KEYWORDS: Morphological complexity. Visual processing. Derivation.*

REFERÊNCIAS

- BANE, M. Quantifying and measuring morphological complexity. In: CHANG, C. B.; HAYNIE, H. J. (Eds.). **Proceedings of the 26th West Coast Conference on Formal Linguistics**, p.69-76, 2008.
- BECKNER, C.; BLYTHE, R.; BYBEE, J.; CHISTIANSSEN, M. H.; CROFT, W.; ELLIS, N.; SCHOENEMANN, T. Language Is a Complex Adaptive System: Position Paper. **Language Learning**, n.59, p.1-26, 2009.
- FROST, R.; KUGLER, T.; DEUTSCH, A.; FORSTER, K. Orthographic structure versus morphological structure: principles of lexical organization in a given language. **Journal of Experimental Psychology**, n.31, p.1293-1326, 2005.
- GONG, T.; COUPÉ, C. A report on the workshop on complexity in language: developmental and evolutionary perspectives. **Biolinguistics**, n.4, p.370-380, 2011.
- JÄRVIKIVI, J.; BERTRAM, R.; NIEMI, J. Affixal salience and the processing of derivational morphology: the role of suffix allomorphy. **Language and Cognitive Processes**, n.21, p.394-432, 2006.
- LAUDANNA, A.; BURANI, C. Distributional properties of derivational affixes: implications for processing. In: FELDMAN, L. (Org.). **Morphological Aspects of Language Processing**. Hillsdale, NJ: Erlbaum, 1995. p.345-364.
- MCQUEEN, J.; CUTLER, A. Cognitive processes in speech perception. In: HARDCASTLE, W.; LAVER, J.; GIBBON, F. (Eds.). **The Handbook of Phonetic Sciences**. Oxford: Blackwell, 1997. p.489-520.
- MUFWENE, S. S. The emergence of complexity in languages: an evolutionary perspective. In: MASSIP-BONET, A.; BASTARDAS-BOADA, A. (Eds.). **Complexity Perspectives on Language, Communication and Society**, p.197-281, 2012.
- PINTO, C. **O papel da morfologia no processamento da palavra escrita**. 2017. 259 f. Dissertação (Doutoramento em Voz, Linguagem e Comunicação) – Faculdade de Letras da Universidade de Lisboa, Lisboa, 2017.
- RASTLE, K.; DAVIS, M.; MARSLÉN-WILSON, W.; TYLER, L. Morphological and semantic effects in visual word recognition: a time course study. **Language & Cognitive Processes**, n.15, p.507-537, 2000.
- RASTLE, K.; DAVIS, M.; NEW, B. The broth in my brother's brothel: morpho-ortografic segmentation in visual word recognition. **Psychonomic Bulletin & Review**, n.11, p.1090-1098, 2004.
- SIMON, H. A. The architecture of complexity. **Proceedings of the American Philosophical Society**, n.106, p.467-482, 1962.

STEELS, L. The Synthetic Modeling of Language Origins. **Evolution of Communication Journal**, n.1, p.1-3, 1997

_____. Language as a Complex Adaptive System. In: SCHOENAUER, M.; DEB, K.; RUDOLPH, G.; YAO, X.; LUTTON, E.; MERELO, J. J.; SCHWFEL, H.-P. (Eds.). **Proceedings of the 6th International Conference on Parallel Problem Solving from Nature**, n.1917, p.17-26, 2000.

TAFT, M. Recognition of affixed words and the word-frequency effect. **Memory and Cognition**, n.7, p.263-272, 1979.

_____. Interactive-activation as a framework for understanding morphological processing. **Language and Cognitive Processes**, n.9, p.271-294, 1994.

TAFT, M.; FORSTER, K. Lexical storage and retrieval of prefixed words. **Journal of Verbal Learning and Verbal Behavior**, n.14, p.638-647, 1975.

VELAN, H.; FROST, R. Words with and without internal structure: what determines the nature of orthographic and morphological processing? **Cognition**, n.118, p.141-156, 2011.

VILLALVA, A. **Estruturas morfológicas. Unidades e hierarquias nas palavras do Português**. Lisboa: FCG, FCT, 2000.

_____. **Morfologia do Português**. Lisboa: Universidade Aberta, 2008.

_____. Bare morphology. **Revista de Estudos Linguísticos da Universidade do Porto**, v.8, p.121-141, 2013.

VILLALVA, A.; SILVESTRE, J. **Introdução ao estudo do léxico. Descrição e análise do Português**. Petrópolis: Vozes, 2015.

Recebido em 5 de fevereiro de 2017

Aprovado em 6 de agosto de 2017