

MORPHOLOGICAL COMPLEXITY AND LEXICAL PROCESSING COSTS

Alina VILLALVA*
Carina PINTO**

- **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.

Languages are complex systems, formed by complex modules that accommodate complex domains.¹ The complexity of one of these modules, the lexicon, must be evaluated at three distinct levels: the first level concerns the assignment of a complexity index to each lexical unit;² the second level deals with the calculation of the complexity

* University of Lisbon (UL), Faculty of Letters, Linguistics Center, Lisbon - Portugal. Department of General and Romance Linguistics. alinavillalva@campus.ul.pt

** University of Lisbon (UL), Faculty of Letters, Linguistics Center, Lisbon - Portugal. pintocarinaal@gmail.com

¹ Gong and Coupé (2011) present the state of the art on language complexity with regard to other complex systems. See also the discussion of linguistic complexity in Mufwene (2012).

² We assume here, like Villava and Silvestre (2015), that the lexicon contains several kinds of lexical units, namely roots, affixes, words and lexicalised phrases.

of words; and the third level concerns the evaluation of the degree of complexity of lexical relationships (e.g. words that share the same affix; words that make use of competing affixes). These three levels of the analysis of lexical complexity are of great relevance for a better understanding of word processing and lexical access.

In this paper, we intend to contribute to the discussion of the second level by looking at complexity contrasts of a particular type of European Portuguese³ derived words (i.e. adjectives in *-oso*). More specifically, we wish to find out if higher processing costs are somehow correlated to a higher index of word complexity, but in order to confirm or refute this claim, we need to know (i) how to assign a complexity index to a word and (ii) how to measure word processing costs.

The framework that will sustain our morphological analysis is based on the characterization of lexical units presented in Villalva and Silvestre (2015) and in the typology of complex word structures drawn by Villalva (2000, 2008). Experimental evaluation and the study of processing will be mainly based in Taft and Forster (1975), Taft (1979, 1994), Rastle et al. (2000) and Rastle et al. (2004).

Complexity and linguistic complexity

The concept of complexity is often invoked in the analysis of many domains of knowledge, but it is often used in an unprincipled way, which can lead to unpredictable and unsystematic interpretations. Simon's (1962) work probably sets the beginning of the theoretical debate in this field. He argues that complexity is a characteristic of systems that take a hierarchical form: complex systems are formed by subsystems, which in turn have their own subsystems, with no limit established for this chain. It can thus be inferred that, regardless of their specific content, hierarchical systems are always complex systems.

Since languages are hierarchical systems, formed by hierarchical modules that accommodate several hierarchical domains, it can be concluded that linguistic systems and all their subsystems are complex systems. However, applying the concept of complexity to each subject of knowledge raises particular questions, and its application to the particular domain of linguistics is still quite limited.

According to Gong and Coupé (2011, p.370), languages can be seen as complex adaptive systems, because they involve a significant number of units and modules that generate a structural complexity at various levels. Languages are complex adaptive systems because their agents - the speakers - have the ability to change the system itself (cf. STEELS, 1997, 2000; BECKNER et al., 2009).

Another claim on linguistic complexity is offered by Bane (2008, p.69):

³ Considering that the formation of adjectives in *-oso* does not present significant differences in the varieties of contemporary Portuguese, we postulate that the results that we obtained with European Portuguese speakers are not exclusive to this variety of Portuguese.

- i. Linguistic complexity remains constant over time;
- ii. The grammar of a given language is as complex as the grammar of any other language.

These are interesting claims because they allow us to assume that linguistic variation does not affect the degree of complexity of any language, although it may shape it differently. For instance, though the ‘richness’ or ‘poverty’ of inflectional morphological systems is often invoked in correlation with syntactic processes, we can assume that the lower complexity of one subsystem will be compensated by a greater degree of complexity of another subsystem. However, to the best of our knowledge, there is no way of measuring ‘richness’ or, in other words, of measuring the complexity of any of the components of any of the linguistic systems.

Equally interesting is the contribution by Mufwene (2012), who describes the concept of complexity in the field of linguistics, considering the following issues:

- i. Unit complexity (e.g. the phonetic inventory size) and rules of each linguistic subsystem - Mufwene calls it **bit complexity**;
- ii. **Interactive complexity**, which regards the relationship between units and rules, within each module, and also the relationship between different modules.

This is an important consideration because it suggests that the complexity of units differs from the complexity of their interaction and it further claims that the complexity of each subsystem also differs from the complexity of the system globally considered.

A combination of these standpoints allows us to set a hypothesis on linguistic complexity based on the following assumptions:

1. It is mandatory to delimit a stable window over a linguistic system, which can be achieved if the selected sample is consistent;
2. Linguistic systems have a similar degree of complexity, but the complexity of their subsystems is variable; therefore, each assessment should select a subsystem and analyse it independently;
3. Considering that the degree of complexity of a language is based on two vectors (the intrinsic value of each unit and the value of its interaction with other units in a structure), the assessment should be directed to only one of these domains.

Assuming that the analysis of word processing data can bring relevant information to the understanding of the complexity of morphological structures, we will present, in this paper, the results of an experimental study, considering complex denominal adjectives derived by suffixation in *-oso*. This study was carried out on a homogeneous sample of European Portuguese speakers.

Previous attempts to quantify word complexity deal with the description of linguistic structures and their function; the predictability of a given word sequence in speech;

the structural regularities of languages or the size of the phonetic inventory; among others (see GONG; COUPÉ, 2011), but they have produced vague measures, making the measurement of linguistic complexity too subjective. Not much work has been done in the domain of morphological complexity either, and most of the existing research is about inflection⁴. In this context, the research on derived morphological structures complexity is pioneer.

On the other hand, the discussion on complexity makes it possible to understand that the distinction between simple words and complex words, that is usually considered in the domain of morphological analysis, is as important as it is simplistic. In fact, this division merely allows to oppose words consisting of a root and its thematic and inflectional specifiers (if required by the root) and words that combine one or more affixes (derivational or modifiers) or two or more roots. It goes without saying that this is a relevant distinction: the set of simple words is an important set in the lexicon of any given language; and the identification of the set of complex words is equally important, since it provides the evidence required for the description of all the word formation processes of the language (past or contemporary). The problem that the discrimination between simple and complex words does not solve is that not all simple words are equally simple; neither all complex words are equally complex. As far as we know, no lexicological source of information allows us to obtain the set of Portuguese simple words. It is also impossible to gather the set of complex words globally. It is possible to generate subsets of complex words formed by a given prefix or suffix, although the results will need to be filtered, since the search is produced on the basis of a spelling criterion (e.g. words that end in *ity*) and not of a morphological criterion (e.g. words that contain the suffix *-ity*). In the case of simple words, no orthographic strategy allows to produce any results. So, we will now explore the kinds of problems that measuring the complexity of simple and complex words may arise.

The complexity of simple words

The basic issue, in this domain, is to determine if a similar morphological complexity index can be assigned to all the roots that occur in simple words, or if these roots belong to different classes of morphological complexity. We are persuaded that simple word roots are not all alike, but we need to identify which factors will substantiate those differences.

One of these factors is probably related to the identification of the word class of roots. Villalva and Silvestre (2015) establish a basic distinction between roots that can occur in simple words (cf. [*tóxic*]o) and those that cannot (cf. [*aqu*]ífero). The former are classified as inherently intransitive predicates (though in other contexts they may

⁴ The research on morphological complexity has been mainly developed by the Surrey Morphology Group (cf. <<http://www.smg.surrey.ac.uk/>>), within the scope of a project carried out between 2009 and 2015 (see **Morphological complexity**: typology as a tool for delineating cognitive organization).

occur as transitive predicates (cf. *neuro[tóxic]o*), as complements (cf. *[tóxic]idade*) or as modifiers (cf. *[tóxic]o-dependente*). The latter, usually called neoclassical roots because they are Latin and Ancient Greek borrowings, are classified as inherently transitive predicates because they can never occur in simple words: these roots occur as complements in some types of derivatives (cf. *[aqu]oso*) and they also occur in morphological compounds, as its head (cf. *aquí[fer]o*) or as its complement (cf. *[aqu]ifero*).

In this section, we will just consider the first type of roots, that is, those that can occur in simple words.⁵ Although there is no reliable information available, we believe that the amount of simple words that belong to a single word class is quite considerable. Words such as *perna* ‘leg’, *grosso/a* ‘thick’ ou *pedir* ‘ask’ belong unequivocally to the domain of nouns, adjectives and verbs, respectively. The roots that are part of these words will therefore have a single word class specification:

- | | | | |
|-----|------------------|------------|----------|
| (1) | $[pern]_{NR}$ | <i>a</i> | ‘leg’ |
| | $[gross]_{ADJR}$ | <i>o/a</i> | ‘thick’ |
| | $[ped]_{VR}$ | <i>ir</i> | ‘to ask’ |

There is, however, a set of roots that is equally or even more important than the previous one – it is the set of roots that occur in different simple words, as in the following examples:

- | | | | | |
|-----|----|------------------|------------|--------------|
| (2) | a. | $[murch]_{ADJR}$ | <i>o/a</i> | ‘faded’ |
| | | $[murch]_{VR}$ | <i>ar</i> | ‘to fade’ |
| | b. | $[danç]_{NR}$ | <i>a</i> | ‘dance’ |
| | | $[danç]_{VR}$ | <i>ar</i> | ‘to dance’ |
| | c. | $[velh]_{ADJR}$ | <i>o</i> | ‘old’ |
| | | $[velh]_{NR}$ | <i>o</i> | ‘old person’ |
| | d. | $[sec]_{NR}$ | <i>a</i> | ‘drought’ |
| | | $[sec]_{NR}$ | <i>o</i> | ‘dry’ |
| | | $[sec]_{ADJR}$ | <i>o/a</i> | ‘dry’ |
| | | $[sec]_{VT}$ | <i>ar</i> | ‘to dry’ |

These roots have not the same lexical status⁶: some will be simple lexical units, owning an unambiguous categorical specification (e.g. $[murch]_{ADJR}$; $[danç]_{NR}$; $[sec]_{ADJR}$), while others are underspecified roots (e.g. $[velh]_{[±N]R}$); and the remaining are the output of conversion processes (cf. $[murch]_{VS}$; $[danç]_{VR}$; $[sec]_{NR}$; $[sec]_{VR}$). This is not the focus of our discussion here – we merely wish to find out if a different status regarding word class membership has consequences for the assessment of word complexity. For

⁵ Transitive roots will be mentioned in the following section, when we will discuss morphological compounds.

⁶ Cf. Villalva (2013).

the purpose of the present study, we will assume that the least degree of complexity belongs to the unambiguous intransitive roots (cf. [*pern*]_{NR}), and that the highest degree of complexity lies in the roots that were formed by conversion (cf. [*murch*]_{VR}, [*danç*]_{VR}, [*sec*]_{NR}, [*sec*]_{VR}). For this reason, the formation of the lists that we have used for the experimental work, included two series of simple words, but they all contain inherently intransitive and unequivocally noun roots (e.g. *veneno* ‘poison’, *mentira* ‘lie’, *luxo* ‘luxury’, *conflito* ‘conflict’).

On the other hand, some roots present a unique form, whatever structure they occur in, but others do not. The existence of alternating forms may be related to morphophonological issues (cf. 3a) or may arise from lexical idiosyncratic circumstances, such as the introduction of neoclassical loans (cf. 3b):

- | | | |
|--------|---------------|----------|
| (3) a. | <i>cão</i> | ‘dog’ |
| | <i>canil</i> | ‘kennel’ |
| b. | <i>veia</i> | ‘vein’ |
| | <i>venoso</i> | ‘venous’ |

The 1st and 2nd series of data in our experiment are formed by roots that have no alternating forms. Roots that have alternating forms are in the 3rd series.

The complexity of complex words

The evaluation of the complexity of complex words raises the difficulty level referred above even further, since, as we have already mentioned, there is no work done in this domain for Portuguese or other languages. There are many open research lines, such as the comparative weight of affixation vs. compounding⁷, of prefixation vs. suffixation (or other affixation types where available); the comparison of configurations involving different levels of embedding and various kinds of interaction between prefixation, suffixation and compounding; and also the evaluation of the productivity of different word formation processes. However, to the present study, only questions concerning the status of inherently transitive roots and allomorphic derivational suffixes will be taken into account.

Inherently transitive roots have been introduced in the Portuguese lexicon since the 18th century, along with the development of scientific and technical terminologies. Words that carry them often arrive as loans, for instance from French (e.g. *termómetro* ‘thermometer’), though these roots have been borrowed from Latin and Ancient Greek. Consider the words in (4). The root [*pedr*] in (4a) is semantically equivalent to the root of Latin origin [*petr*], which occurs in (4b) and to the root of Greek origin [*lit*] in

⁷ Only morphological compounding (which involves the sequencing of roots, not words) is included here. Morphosyntactic and syntactic compounding are not morphological word formation processes.

(4c). Only the first one appears in a simple word (e.g. *pedra* ‘stone’) that is associated to a semantic value accessible to native speakers. It is assumed that the semantics of the words containing this root are the most transparent and that the processing of their derivatives is compositional. As for the other two roots (cf. 4b and 4c), none of them can be known outside the context of the complex words in which they occur. The meaning of these complex words is often non-compositional.

- | | | |
|--------|-------------------|----------------|
| (4) a. | <i>pedra</i> | ‘stone’ |
| | <i>pedreira</i> | ‘stone quarry’ |
| b. | <i>pétreo</i> | ‘stony’ |
| | <i>petrificar</i> | ‘to petrify’ |
| c. | <i>litografia</i> | ‘lithography’ |
| | <i>megalítico</i> | ‘megalithic’ |

In our experiment, the 3rd series of data contains variants of roots that can only occur in complex words (e.g. *aquoso* ‘watery’ / *água* ‘water’; *medroso* ‘fearful’ / *medo* ‘fear’).

Alternation can also affect affixes. In the specific case of *-oso*, there is an allomorph (*-uoso*) that is unpredictable in contemporary Portuguese. The 2nd series of data is formed by lexicalized words that contain this allomorph (cf. *conflituoso* ‘conflicting’; *luxuoso* ‘luxurious’).

This could be a reliable index for calculating the morphological complexity of complex words. So, our initial hypothesis is that lexicalised words, either because they contain an allomorph of the root (3rd series) or an allomorph of the suffix (2nd series), reveal higher processing costs than compositional words.

Assessment of morphological complexity

The degree of complexity of a word is more than a simple arithmetic sum of the complexity indexes associated to each of its constituents – probably, its assessment requires an algorithmic function that we are far from being able to devise. There are, therefore, two lines of work worth exploring: on one hand, we need to identify what can determine the complexity index of each of the morphological constituent and the algorithm that can allow us to calculate the complexity index of each word; and, on the other hand, we need to find tools that will allow the validation of the stipulations advanced by linguistic analysis.

In the following sections we will explore the possibility of finding a validation tool based on the analysis of the morphological and lexical processing data obtained experimentally.

Morphological complexity and processing

Several experimental studies have already allowed claiming that the morphological structure plays a role in visual word recognition and lexical access. These studies were developed within a wide range of experimental paradigms, as reviewed by McQueen and Cutler (1997). A more recent state of the art can be found in Pinto (2017).

Priming experiments have helped to identify the features that are relevant for lexical activation, as well as their role. The results presented by Frost, Kugler, Deutsch and Forster (2005) and by Velan and Frost (2011) suggest the existence of two hierarchically organized processes:

1. The morpho-orthographic stage characterizes the earlier phases of word recognition - the process is activated when the visual input (the written word) is complex and compositional - in this phase the recognition of forms (base and affix) is activated;
2. The morpho-semantic stage is activated later - at this stage, the previously recognized forms are semantically interpreted.

Other studies (cf. TAFT; FORSTER, 1975; TAFT, 1979, 1994) suggest that all morphological constituents are stored in the mental lexicon as independent lexical inputs and the access to meaning is obtained by decoding the meaning of each individual unit. Morphological priming studies, such as Laudanna and Burani (1995) or Järvikivi et al. (2006), have shown that words containing affixes with various allomorphs have a longer latency response. Probably, the visual processing of morphologically complex words can be affected by different properties of morphological constituents, namely the existence of allomorphs and the productivity of each affix.

Much of the evidence that morphological structure plays a role in the visual processing of complex words is compatible with both a global word processing model and an autonomous processing of word constituents. Nevertheless, it is still unclear when and how morphological analysis comes into play.

Experimental assessment

As a contribution to the discussion on morphological complexity, we have designed and applied four experiments: a simple lexical decision task and three morphological priming tests. By performing these two types of experimental procedures, we intended to find out if the introduction of a prime corresponding to the base form of a derived target facilitates the visual processing of these complex words or not. A positive response may indicate that morphological analysis is available and facilitates the understanding of derivatives.

On the other hand, performing the same priming test with three different exposure intervals will allow us to verify whether morphological processing is performed at a more initial stage or at a later phase of visual processing. We assume here that morphological processing occurs at various stages of word processing, depending on the complexity of the constituents of complex words.

Finally, we have used three series of words derived by the same suffix: the 1st series gathers compositional structures, while the other two include derivatives that present an allomorph of one of the constituents. The hypothesis behind this choice is that words with a compositional structure are accessed more quickly than words with a structure disturbed by the occurrence of allomorphs.

Morphological Data

All the words that we have tested are formed by *-oso*, which is a denominal adjective-forming suffix. We took into account the number of syllables: simple words have two or three syllables (e.g. *veneno* ‘poison’, *luxo* ‘luxury’, *medo* ‘fear’) and the derivatives have four or five syllables (e.g. *venenoso* ‘poisonous’, *luxuoso* ‘luxurious’, *medroso* ‘fearful’). The frequency of use in European Portuguese was equally checked with CRPC⁸, which allowed us to select high frequency words, although derivatives always have a relatively low frequency rate, as an inherent feature.

Finally, and taking into account the morphological structure of the derivatives, we have set three series of ten words, which correspond to the following conditions:

- 1st derivatives with a compositional structure (e.g. *venenoso* ‘poisonous’) - the base is a noun root (*venen-* ‘poison’), the suffix is *-oso*; the derivative is an adjective (presented in the masculine singular form) that can be paraphrased by the expression $X]_{NR} \textit{oso} = \textit{‘that has X’}]_{ADJ}$ (*venenoso* ‘poisonous’ = ‘that has *venen(o)* ‘poison’);
- 2nd derivatives with an allomorph of the suffix (e.g. *luxuoso* ‘luxurious’) - the base is a noun root (*lux-* ‘luxury’), the suffix allomorph is *-uoso*; the derivative is an adjective (presented in the masculine singular form) that can be paraphrased by the expression $X]_{NR} \textit{oso} = \textit{‘that has X’}]_{ADJ}$ (*luxuoso* ‘luxurious’ = ‘that has *lux(o)* ‘luxury’);
- 3rd derivatives with an allomorphy of the base (e.g. *medroso* ‘fearful’) - the base is an allomorph of a noun root (*medr-* ~ *med-* ‘fear’), the suffix is *-oso*; the derivative is an adjective (presented in the masculine singular form) that can be paraphrased by an expression $X]_{NR} \textit{oso} = \textit{‘that has X’}]_{ADJ}$ (*medroso* ‘fearful’ = ‘that has *medo* ‘fear’).

⁸ CRPC is an electronic contemporary Portuguese Corpus that includes more than 1.6 million words. It is available at <www.clul.ul.pt/pt/recurcos/183-reference-corpus-of-contemporary-portuguese-crpc>.

Methodology

All these experiments were carried out with Portuguese subjects. The sample gathered healthy individuals, college students, from Lisbon and Leiria. This study has obtained a favourable opinion from the National Commission of Data Protection (Authorization nº 7788/2013).

In order to exclude cognitive-linguistic changes, the following exclusion criteria were applied to the sample:

- 1) Cerebral Vascular Accident;
- 2) Epilepsy;
- 3) Cranio-Encephalic Trauma;
- 4) Major depression / schizophrenia diagnosed by a medical specialist;
- 5) Uncorrected visual changes;
- 6) Uncorrected auditory changes;
- 7) Changes in written language characterized in DSM IV;⁹
- 8) Serious medical illness that could lead to the appearance of linguistic alterations;
- 9) Drug Addiction / Alcoholism;
- 10) Bilingualism.

In total, we have collected data from 116 subjects. Their informed consent was obtained and all subjects were submitted to an assessment of oral and written language changes. None of the subjects had any deviant behaviour.

Procedure

For the experiments reported in this paper we have used three different prime exposure times: 50 milliseconds (=ms), 100 ms and 150 ms. Experiments were built using the software E-Prime® 2.0. Visual stimuli were presented on a Compaq Presario® computer. The following table shows the distribution of subjects,¹⁰ by experience:

Table 1 – Distribution of subjects by four experiments

Prime presentation during 50ms	32
Prime presentation during 100ms	30
Prime presentation during 150ms	27
Lexical decision	27

Source: Own elaboration.

⁹ DSM IV (*Diagnostic and Statistical Manual of Mental Disorders*) is a manual from the American Psychiatry Association that is used to diagnose mental illnesses.

¹⁰ The collected sample is a convenience sample. The data collection was done with 32 subjects by experience, but data treatment led to an exclusion of some subjects for several reasons, such as percentage of errors in the responses.

The lexical decision test included an experimental list formed by thirty words presented as a target in the priming test (e.g. *venenoso* ‘poisonous’, *luxuoso* ‘luxurious’, *medroso* ‘fearful’). The other three tests included a series of ten pairs of words: the prime is the base word and the target is the derivative (e.g. *veneno* ‘poison’ – *venenoso* ‘poisonous’; *luxo* ‘luxury’ – *luxuoso* ‘luxurious’; *medo* ‘fear’ – *medroso* ‘fearful’). We have also used pseudo-words as fillers. These pseudo-words were obtained by a systematic procedure of replacement of syllables of the target words (e.g. *bexigoso* – *goxiboso*). The fillers are used, canonically, to motivate the lexical decision and conceal the purpose of the test, thus fulfilling the role of distractors.

All the experiments started with six training items (i.e. words that are not part of the test). Both the training items and the experimental items (i.e. test words) appeared in the centre of the screen, in black on a white background, using Times New Roman font size 18, and lowercase. They were preceded by a mask (+), which was displayed in the same position for 500ms, serving as a fixation point. In the case of the morphological priming experiment, the prime (50 ms, 100 ms or 150 ms) appeared immediately after the mask, instantaneously followed by the target word, which was available on the screen until the subjects made the lexical decision, using the computer keyboard.

Results

For the data analysis, we have used the SPSS® software version 20. As usual in similar cases, we have carried out the cleaning of the output data. The outliers were dealt taking into account the following criteria:

- 1) exclusion of wrong answers;
- 2) exclusion of values over 2000 ms;
- 3) replacement of disparate values greater than 10,000 ms and smaller than 250 ms by the mean of the subject in the condition;
- 4) replacement of values above the mean in $\pm 2,5$ standard deviations by the mean of the subject in the condition.

This cleaning aimed to reinforce the quality of the answers that might be threatened by errors of execution or by an inherent variability to the sample elements.

The following table shows the percentage of excluded data, taking into account the above criteria:

Table 2 – Percentage of deleted data

	Prime exposure time	Wrong answers	Disparate values	Values > mean $\pm 2,5$ SD	Values >2000ms
Lexical decision		4,50%	0 %	2,33%	5,04%
Morphological priming	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%

Source: Own elaboration.

Normality tests were also performed -they did not reveal a normal distribution. This is an expected circumstance, since the observation focuses on reaction times - the data collected has a limit on the left, that is, there is always the limit zero (0) that prevents symmetry. Thus, non-parametric statistical tests have been used to analyse the data.

Assessment of the role of the morphological condition

This experiment aimed to evaluate the costs of word visual processing associated to three morphological conditions: regular derivation, derivation with suffix allomorphy and derivation with base allomorphy. Table 3 presents the results by task: lexical decision; priming (50 ms) plus lexical decision; priming (100 ms) plus lexical decision; and priming (150 ms) plus lexical decision.

Table 3 – Descriptive data from the four experiments

<i>Prime exposition</i>		Compositional structure	Suffix allomorphy	Base allomorphy
<i>Lexical Decision</i>	Mean	932,77	1007,66	993,13
	(SD)	(317,93)	(356,45)	(322,84)
	Median	854,00	893,00	896,00
	1st quartile	697,50	727,50	738,00
	3rd quartile	1099,00	1270,00	1208,00
50 ms	Mean	782	908	861
	(SD)	(220,21)	(310,37)	(229,96)
	Median	713,00	795,00	812,50
	1st quartile	662,50	688,00	710,00
	3rd quartile	901,00	1075,50	970,00

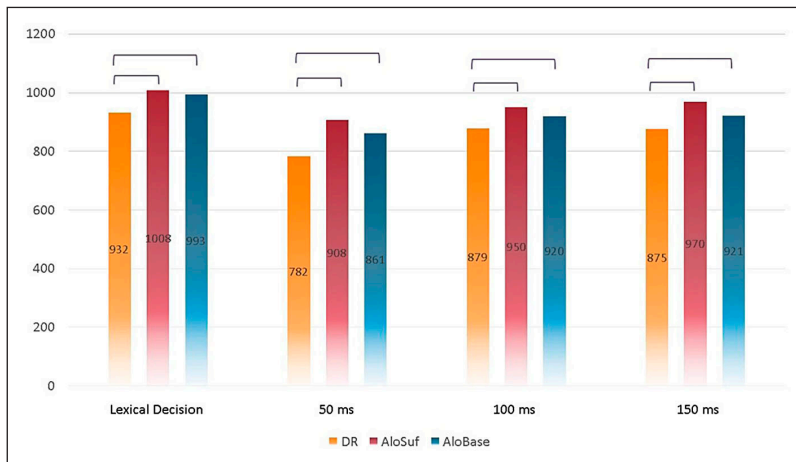
<i>Prime exposition</i>		Compositional structure	Suffix allomorphy	Base allomorphy
100 ms	Mean	879	950	920
	(SD)	(282,86)	(321,40)	(272,15)
	Median	802,00	861,00	856,00
	1st quartile	682,00	722,00	731,00
	3rd quartile	994,00	1178,00	1028,00
150 ms	Mean	875	970	921
	(SD)	(328,80)	(342,13)	(290,58)
	Median	761,50	886,00	858,00
	1st quartile	659,00	709,00	704,50
	3rd quartile	1019,00	1176,00	1032,00

Source: Own elaboration.

These four experiments present fairly homogeneous results, and the condition ‘compositional structure’ (= CS) is always different from the other two. Graphic 1 shows that there are statistically relevant differences between reaction times (marked by braces), in the following cases:

1. Lexical decision experiment
 compositional structure – suffix allomorphy ($U=24367,5; p=0,034$)
 compositional structure – base allomorphy ($U=27062,5; p=0,018$)
2. Morphological priming experiment– 50ms
 compositional structure - suffix allomorphy ($U=34023,5; p=0,000$)
 compositional structure - base allomorphy ($U=34122; p=0,000$)
3. Morphological priming experiment – 100ms
 compositional structure - suffix allomorphy ($U=34308; p=0,009$)
 compositional structure - base allomorphy ($U=36190; p=0,008$)
4. Morphological priming experiment – 150ms
 compositional structure - suffix allomorphy ($U=23949,5; p=0,000$)
 compositional structure - base allomorphy ($U=25324; p=0,004$)

Graphic 1 – Significant differences in the different experiments



Source: Own elaboration.

These results show that compositional structures are always processed in less time and that, conversely, structures that have a suffix allomorph are those that require longer processing time.

Assessment of the role of priming

The following results evaluate the processing costs due to the existence or non-existence of priming and the different exposure time to prime. Results are presented by condition.

1. Compositional structure (*e.g. orgulho 'proud'/orgulhoso 'proud'*)

There are statistically significant differences in the following reaction times:

1. Prime exposure of 50 ms – prime exposure of 100ms ($U=34644$; $p=0,000$)
2. Prime exposure of 50 ms – prime exposure of 150 ms ($U=34196,5$; $p=0,005$).
3. Prime exposure of 50 ms – lexical decision ($U=28155,50$; $p=0,000$).

Graphic 2 – Significant differences in the condition ‘compositional structure’



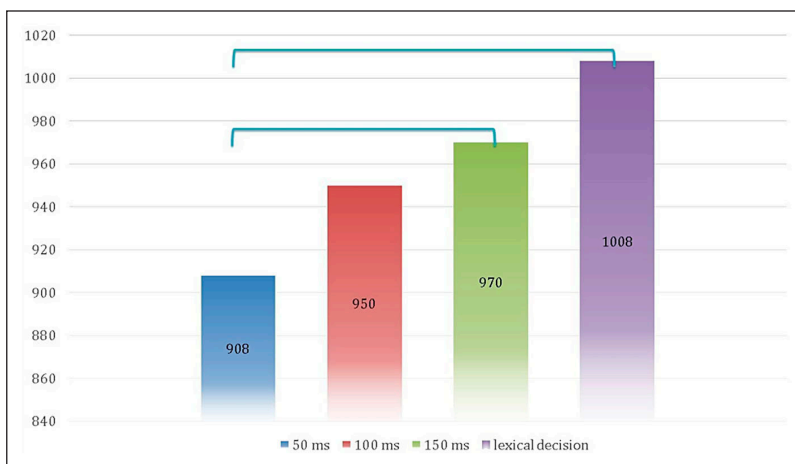
Source: Own elaboration.

2. Suffix allomorphy (e.g. *defeito* ‘defect’ / *defeituoso* ‘defective’)

In this condition there are significant differences in the reaction times, except when the exposure time of 50 ms is contrasted with the exposure time of 100 ms:

1. Prime exposure of 50 ms – prime exposure of 150 ms ($U=30093$; $p=0,034$).
2. Prime exposure of 50 ms – lexical decision ($U=26072,5$; $p=0,002$).

Graphic 3 – Significant differences in the condition ‘derivation with suffix allomorphy’



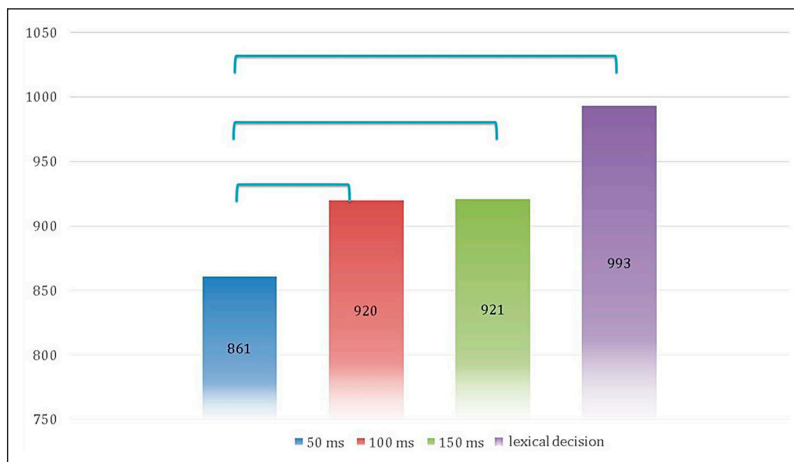
Source: Own elaboration.

3. Base allomorphy (e.g. *lume* ‘fire’ / *luminoso* ‘bright’)

This condition repeats the trends previously found. Significant differences in reaction times are obtained in the following cases:

1. Prime exposure of 50 ms – prime exposure of 100 ms ($U=36610$; $p=0,007$).
2. Prime exposure of 50 ms – prime exposure of 150 ms ($U=30360$; $p=0,03$).
3. Prime exposure of 50 ms – lexical decision ($U=66474$; $p=0,038$).

Graphic 4 – Significant differences in the condition ‘base allomorphy’



Source: Own elaboration.

Discussion

In the first experiment, which evaluated visual processing costs associated with three morphological conditions (regular derivation, derivation with suffix allomorphy and derivation with base allomorphy), our results indicate that compositional structures are processed in less time than lexicalised words, either by an allomorphy of the base or the suffix. We have also found out that visual processing of structures with suffix allomorphy requires a greater effort than all the other, which can be due to a smaller salience of affixes (perhaps semantic, perhaps formal, but the data do not allow to draw safe conclusions) with respect to roots.

The second set of experiments, which evaluated the costs of visual processing due to the existence or non-existence of priming and different prime exposure times, confirmed that reaction time decreases when there is priming, which means that when morphological analysis is induced, processing costs decrease. On the other hand, the contrast of prime exposure times (50ms, 100ms and 150ms) has also generated some

interesting results, since the reaction time increases with the increase of prime exposure time, irrespective of the morphological condition.

The combination of all these observations allows us to conclude that the processing of derived words reading is facilitated when the presentation of their base forms induces morphological analysis, and it also suggests that the effect is more visible in an initial phase of visual word processing.

Final remarks

This study was driven by a discussion on the nature of morphological structures and on the fragility of current knowledge about its degree of complexity. The opposition between simple and complex words captures a basic distinction between words that are not formed by word formation processes and words that are derived, modified, or compounds. However, it can be easily demonstrated that not all simple words are equally simple and that not all complex words are complex in the same way.

From the point of view of morphological analysis, it is relatively easy to set complexity indexes to words constituents, and to proceed to the calculation of the complexity of each word, but this is merely the output of theoretical stipulations. This is why we decided to look at morphological processing data as a means to validate our linguistic assumptions. We have selected a particular word formation process, the formation of *-oso* derivatives, contrasting compositional structures (cf. *venenoso* ‘venomous’) to lexicalised structures, with suffix allomorphy (cf. *luxuoso* ‘luxurious’) and base allomorphy (cf. *medroso* ‘fearful’).

The results obtained in the four experiments allow us to draw the hypothesis that visual processing of complex words derived by suffixation in Portuguese is sensitive to properties of its morphological structure. They also reveal that compositional structures have lower processing costs than structures disturbed by factors such as suffix allomorphy or base allomorphy. Additionally, we found out that ‘suffix allomorphy’ has higher processing costs than ‘base allomorphy’. This finding is somehow counterintuitive, but it may indicate that words that integrate base allomorphs (cf. *arenoso* ‘sandy’) are closer to lexicalization (that leads to a direct lexical access) than words that include suffix allomorphs (cf. *luxuoso* ‘luxurious’). The latter, which are heaviest for word processing, seem to remain analysable, but require an additional effort to ensure that the suffix is duly recognized.

Another aspect that arises from the results analysis concerns the priming role. In fact, since reaction time decreases with the existence of an exposure to a prime, it may be assumed that processing of derived words always involves morphological analysis tasks, since it is facilitated by the exposure to a base form. On the other hand, reaction time increases in direct proportion of the increase of exposure time to a prime, which seems to indicate that the facilitation factor offered by the presence of a prime is more relevant in an earlier time window than in the later stages of processing.

In sum, the results that we have obtained corroborate the hypothesis that complex compositional words, which contain inherently intransitive roots and the canonical form of a suffix (in this case, *-oso*), have lower processing costs and therefore a lower complexity index than lexicalized complex words. However, the second hypothesis initially considered was contradicted by the results that indicate that lexicalization triggered by suffix allomorphy is more onerous for processing than lexicalization triggered by root allomorphy. In other words, when the subjects identify an inherently transitive root, they give up word morphological analysis, and they process it as a simple word; when the root is positively identified as an intransitive root, although the suffix corresponds to an allomorph with a random distribution, the attempt to analyse the word morphologically lasts a bit longer until that process is abandoned. Presumably, then, the index of complexity of derived words is sensitive to the nature of the process of lexicalization.

VILLAVA, A.; PINTO, C. Complexidade morfológica e custos de processamento lexical. *Alfa*, São Paulo, v.62, n.1, p.149-168, 2018.

- *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.*

REFERENCES

- 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.; CHRISTIANSEN, 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-orthographic 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.

STEELS, L. Language as a Complex Adaptive System. In: SCHOENAUER, M.; DEB, K.; RUDOLPH, G.; YAO, X.; LUTTON, E.; MERELO, J. J.; SCHWIFEL, 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, 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.

Received in February 5, 2017

Approved in August 6, 2017