

**The role of the frequency of constituents in compound words: Evidence from
Basque and Spanish**

Jon Andoni Duñabeitia ¹, Manuel Perea ², and Manuel Carreiras ¹

¹ Instituto de Tecnologías Biomédicas, Universidad de La Laguna

² Universitat de València

Running head: Compound words and lexeme frequencies

Word count: 4414

Address for correspondence:

Jon Andoni Duñabeitia

Departamento de Psicología Cognitiva

Universidad de La Laguna

38205-Tenerife (Spain)

phone: +34 922 317515

fax: +34 922 317461

email: jaduna@ull.es

Abstract

Recent data from compound word processing suggests that compounds are recognized via their constituent lexemes (Juhasz et al., 2003). The present lexical decision experiment manipulated orthogonally the frequency of the constituents of compound words in two languages: Basque and Spanish. Basque and Spanish diverge widely in their morphological properties and in the number of existing compound words. Furthermore, the head lexeme (i.e., the most meaningful lexeme related to the whole-word meaning) in Spanish tends to be the second lexeme, whereas in Basque the percentage is more distributed. Results showed a facilitative effect of the frequency of the second lexeme, in both Basque and Spanish compounds. Thus, both Basque and Spanish readers decompose compounds into their constituents for lexical access, and this decomposition is carried out in a language-independent and blind-to-semantics manner. We examine the implications of these results for models of lexical access.

The role of the frequency of constituents in compound words: Evidence from Basque and Spanish

Recent years have witnessed an increasing body of literature supporting the view that the early processing of printed words is influenced by morphology (see Frost, Grainger, & Rastle, 2005; Frost, Kugler, Deutsch, & Forster, 2005). Although most of the research on this issue has studied derivational/inflectional morphology, investigations into compound words also provide important knowledge about morphological and lexical parsing (see Juhasz, Inhoff, & Rayner, 2005).

There is some controversy as to exactly how poly-morphemic words (including compound words) are represented in the mental lexicon. There are two main approaches to this question. On the one hand, a compound (e.g., *weatherman*) could be recognized via its constituents –via a lexeme parsing mechanism that enforces the conjunction of *weather* and *man*. In this view, the recognition of the whole-word would be modulated by the properties of each lexeme (e.g. Andrews, 1986; Juhasz et al., 2005; Libben, 1998; Sandra, 1990; Taft, 1979; Zwitserlood, 1994). On the other hand, some researchers argue that decomposition may not be mandatory to access the meaning of a compound word. Even though the meaning of *weatherman* can be somehow derived from the meaning of its constituents, this must occur in supra-lexical processing of the compound (see Fowler, Napps & Feldman, 1985; Giraudo & Grainger, 2001; Plaut & Gonnerman, 2000). The accumulated evidence for compound word processing strongly supports a third (hybrid) approach that assumes both a whole-word process and a sequential decomposition process (e.g. Bertram & Hyönä, 2003; Inhoff, Radach & Heller, 2000; Pollatsek, Hyönä & Bertram, 2000; Taft, 1994): When a reader encounters a compound word, segmentation mechanisms fire and constituents are progressively recognized, while the meaning of the whole compound is also achieved by a whole-

word recognition pathway.

Lexical decomposition of compound words has often been examined by manipulating constituent frequency, while keeping constant the frequency of the compound word. There is empirical evidence showing that high frequency constituents facilitate the recognition of the compound. For instance, *farmhouse* will be recognized faster than *graveyard* simply because both the first and second constituents are high frequency lexemes in *farmhouse*, and low frequency lexemes in *graveyard* (see Hyönä & Pollatsek, 1998; Juhasz, Starr, Inhoff, & Placke, 2003; Shoolman & Andrews, 2003; Taft, 1979). Nonetheless, there is still a gap to be filled with regard to which lexeme, the first or the second, and therefore which frequency of constituent has a stronger weight in the processing of compounds. To our knowledge, only two papers have tackled this question by reporting an experimental manipulation involving an orthogonal design, while holding the frequency of the whole compound constant: Juhasz et al. (2003) and Andrews, Miller, and Rayner (2004, Experiment 1). It is important to note that this design allows testing of not just the main effects but also the potential interaction between the frequencies of the two lexemes. Juhasz et al. found a facilitative effect of the frequency of the second constituent not only in lexical decision and naming, but also in a silent reading (eye movement) experiment using English compounds. In Experiments 1 and 2 (lexical decision and naming), Juhasz et al. also found a trend towards a frequency effect of the first lexeme. Andrews et al. (2004, Experiment 1) failed to find a significant effect of the frequency of either constituent (although there was a non-significant facilitative trend for both constituents) in an eye movement experiment. All the other papers published manipulated the frequency of one of the lexemes, while maintaining constant the frequency of the other lexeme. For instance, Pollatsek and Hyönä (2005) designed a set of three eye movement experiments

in which the frequency of the second lexeme was kept invariant. They showed a significant first constituent frequency effect. In another set of eye-tracking experiments, Pollatsek, Hyönä, and Bertram (2000) manipulated only the second constituent's frequency, while maintaining fixed the frequency of the initial constituent. They found a robust ending lexeme effect when analyzing late measures such as second fixation duration and gaze duration.

The goal of the present experiment is to determine which lexeme of the compound word exerts a greater influence in its recognition. To that end, we manipulated orthogonally the frequencies of each constituent, while holding the whole-word frequency constant. We did so employing two languages with very different morphological properties, Basque and Spanish. Cross-language comparisons are a key issue to understand the generality of morphological parsing (see Frost et al., 2005, for comparison between Indo-European and Semitic languages). Basque is a pre-Indo-European language which has a very specific morphological structure based on morpheme agglutination at the end of the stem (see Perea & Carreiras, 2006, for a description of the Basque language). In such a strongly agglutinative language, Basque words are inflectionally and derivationally modified by adding morphemes and lexemes to the latter part of the stem, and word compounding is a useful and very common strategy for novel word creation (e.g. *baso*, for *forest*; *basagizona* for *wild man*; or *basozaina* for *gamekeeper*, it is possible to find up to fifty compound words with the root *bas-*). Euskaltzaindia --the entity that regulates the Basque language-- provides, in one of their publications, an appendix with up to 2175 examples of compound words, just by way of illustration (see Zalbide, 1992). Thus, Basque is an excellent language to examine whether words are decomposed in the process of lexical access. In contrast, Spanish follows the morphological rules of Romance languages (e.g., the previous

example *gamekeeper* would be translated as *guardabosque*; only two compound words in Spanish include the lexeme *bosque*, corresponding to *forest*). Spanish is very restricted as to compound word creation: in the Spanish database (Sebastián-Gallés, Martí, Carreiras & Cuetos, 2000) there are less than 425 compound words.

A second goal of the present experiment is to determine whether or not “the second constituent effects reflect the fact that the meaning of compound words is usually determined more by their second than their first constituent” (Andrews et al., 2004, pp. 304). If this view is correct, then robust second lexeme frequency effects would be expected when the materials involve compounds with ending heads (the meaning of the whole compound is directly related to the second lexeme’s meaning; e.g., *blackboard*, see Juhasz et al., 2003, pp. 240), highlighting the relationship between the ending constituent and the whole-word processing . Juhasz et al. discussed the second lexeme effect by arguing that it fitted into theories favoring a late processing of the constituents. As their compounds tended to have an ending head, they argued that their data reinforced the model of Pollatsek et al. (2000), who indicated that ending lexeme activation co-occurs with full compound activation. Interestingly, Spanish compounds are generally composed by an ending head (second lexeme) modulated by an initial lexeme (e.g. *pasatiempo* for *pastime*, *pasa*(*pass*)+*tiempo*(*time*)). In the previously mentioned count of compound words in Spanish, more than the 75% of these had an ending head (323 words). The distribution of the location of the semantic heads in Basque compounds is more ambiguous. Because of the syntactic structure of Basque (footnote 1), it is very easy to find compound words with a first lexeme that is the head of the compound, and a second lexeme that modulates its meaning (e.g. *denborapasa*, meaning *pastime*, composed by *denbora*(*time*)+*pasa*(*pass*)). Nonetheless, due to the wide range of compound words in Basque, one can find interesting examples in which

the semantic head may vary according to the order of the constituents: *esne-behia* (milk+cow, a cow that gives milk) and *behi-esnea* (cow+milk, milk that comes from a cow). In the present experiment, almost 80% of the Basque compounds had an *initial* head, whereas almost 75% of the Spanish compounds had a *final* head. (This was tested by conducting a norming study with 10 Basque and 10 Spanish native speakers; Basque speakers rated 77.5% of the Basque words as initial headed, whereas Spanish speakers rated 73.5% of the Spanish compounds as final headed.) Thus, if we find a similar pattern of results in these two languages, this would provide evidence for a decomposition mechanism that is not influenced by the compound's head position.

In sum, in this experiment, the frequencies of the constituent lexemes of Basque and Spanish compound words were orthogonally manipulated in a lexical decision task. We recruited a group of native speakers of Basque (Basque sub-experiment) and a group of native speakers of Spanish, who did not have any knowledge of Basque (Spanish sub-experiment).

Experiment

Method

Participants. Ninety-six students took part in this experiment. Fifty-two students from the University of the Basque Country, Bilbao, participated in the data collection for the Basque sub-experiment. Forty-four students from the University of La Laguna received course credit for taking part in the Spanish sub-experiment. All the participants had normal or corrected-to-normal vision, and were native speakers of the language they were tested in.

Materials. The Basque materials were composed of a set of forty-four low frequency

Basque compounds, selected from the Perea et al. (in press) Basque database (mean word frequency per million: 3.98; range: 0.28-9.66; mean word length: 8.59; range: 7-10). Each of the Basque compounds permitted the division into its constituents (e.g. a compound like *legegizon*, the Basque for *lawyer*, could be divided into *lege+gizon*, that is, *law+man*). The mean frequency per million of the first constituent was 916.22 (range: 0.83-3251.49), and the mean length was 4.55 (range: 3-6). The overall values for the second lexemes were also matched (mean frequency: 832.87; range: 3.59-2442.28; length: 4.36; range: 3-6). Four groups of compound words were created, using the frequency of the first and second lexeme as variable of distribution. Each group of stimuli was formed by 11 different compound words. The characteristics of these groups are presented in Table 1. Almost 80% of all the Basque compound words had an initial head (i.e., their general meaning could be extracted from the meaning of the first lexeme). An additional set of 44 nonwords was created to make lexical decision possible. Nonwords were obtained by replacing several letters from the original compound words (e.g. *iluspentu*, taken from *ikuspuntu*, the Basque for *point of view*). The mean length of the nonwords was 8.43 (range: 7-11). A large number of filler items (240 word and 240 nonword trials) were added to the experimental items, in order to prevent participants from expecting compound words. The filler items were either morphologically simple or complex (affixed), and did not differ in overall length from the experimental items. Thus, the percentage of compounds was less than 16%.

The Spanish materials included forty-eight targets that were Spanish compound words of six to twelve letters (mean word frequency per one million words in the Sebastián-Gallés, Martí, Carreiras & Cuetos, 2000, LEXESP database count: 2.23, range: 0.18-7.68, and mean word length: 9.81). As in the Basque materials, each compound word (e.g. *guardameta*, the Spanish for *goalkeeper*) was completely decomposable into its

two constituent lexemes without letter addition or deletion (footnote 2) (e.g. *guarda+meta* would be as *keeper+goal*). The frequency of each constituent of the compound word and its length was overall controlled. The mean lexeme frequency of the first constituent was 251.84 (range: 1.61-1594.11), and the mean length was 4.94 (range: 3-7). The mean frequency of the second lexeme was 205.73 (range: 4.46-1378.93), and the mean length was 4.73 (range: 3-8). Combining high and low frequency constituents, four groups of compounds were designed, as shown in Table 1, each group containing 12 different stimuli. Frequency and length was also controlled between groups, resulting in a balanced distribution. Nearly 75% of all the Spanish compounds had an ending head, so that their general meaning was mostly determined by their second lexeme. An additional set of 48 nonwords of six to twelve letters was included for the purposes of the lexical decision task. The nonwords were obtained by replacing two, three or four letters from the original Spanish compound words (e.g. *rosemiscas*, obtained from *tocadiscos*, the Spanish for *record player*). A list of materials was created, by designing a large number of filler items (up to 480), with the same amount of words than of nonwords (240 each). Those filler items were approximately of the same length as the experimental compound stimuli (mean length: 8.89; range: 7-11), and included both morphologically complex (derived by affixes) and simple (non-affixed) words. Therefore, the percentage of experimental compound words in the Spanish sub-experiment did not exceed 10%.

Insert Table 1 around here

Procedure. The experiment was run individually in a quiet room. The stimuli presentation and data collection of the response times were controlled by DMDX software (Forster & Forster, 2003) in a PC compatible computer with a CRT monitor.

The participants had to press one of two labeled buttons on the keyboard in order to indicate the linguistic legality of the letter string presented in the screen ('M' for words and 'Z' for nonwords). All the stimuli were presented centered, in uppercase 12-pt. Courier New font. All the trials were randomized so that there was no trial order repetition across participants. The Basque participants received their instructions and training in Basque, and the group of Spanish speakers received theirs in Spanish.

Results

Incorrect responses (9.7% of the trials) and reaction times beyond the cutoff values 250-1500 ms (7% of the word data) were excluded from the latency analyses. The discarded reaction times due to the cutoff values were equally distributed across languages and conditions. Mean response latencies and percentages of errors associated to each group of compounds and each language are offered in Table 2. ANOVAs on the reaction times and error rates by participants and items were conducted based on a 2 (Frequency of the first lexeme: high, low) x 2 (Frequency of the second lexeme: high, low) x 2 (Language: Basque, Spanish) design.

Insert Table 2 around here

The ANOVA on the response times showed a significant ending lexeme effect, with faster decision times for compounds with a high frequency second constituent (an effect of 50 ms), $F_1(1,94)=44.57$, $p<.01$, $1-\beta=1$; $F_2(1,91)=4.35$, $p<.04$, $1-\beta=.54$. No such effect was found for the first lexeme (a negligible 4 ms effect), both $F_s<1$. The magnitude of the second lexeme effect is higher for the Basque (71 ms) than for the Spanish words (25 ms), but this interaction with the language was significant only in the

analysis by participants, $F(1,94)=16.45$, $p<.01$, $1-\beta=.89$; $F(1,91)=.719$, $p>.39$, $1-\beta=.13$. However, it should be noted that –because of the morphological characteristics of each language– the high frequency lexemes of the Basque compounds are more frequent than the high frequency lexemes of the Spanish compounds, and this could have increased the second lexeme effect for the Basque sub-experiment. (Consistent with this interpretation, a regression analysis on the item response times showed a significant effect of the log of second lexeme frequency, $\beta=-.25$, $p<.025$.) None of the other effects or interactions was significant (all $ps>.25$).

The ANOVA on the error data showed that participants were more accurate when the second constituent was of high frequency than when it was of low frequency, but this effect only reached significance in the analysis by subjects, $F(1,94)=18.70$, $p<.01$; $F(1,91)=1.25$, $p>.15$. Participants made more errors when the targets were Spanish compounds than when they were Basque compounds, $F(1,94)=37.49$, $p<.01$; $F(1,91)=10.17$, $p<.01$. The interactions with language did not approach significance (all $ps>.10$). (footnote 3)

Discussion

The present findings provide compelling evidence for morphological decomposition during the reading of compound words, as deduced from the constituent frequency effects. More specifically, the lexeme frequency effect occurs for the second constituent (but not the first), both in Basque and Spanish compounds. Thus, the main findings can be summarized as follows: i) in a lexical decision task, when the frequency of the constituents of a compound word are orthogonally manipulated, the frequency of the second lexeme modulates response times; ii) this modulation seems to be based on a

language-independent mechanism of morphological decomposition; iii) the position of the lexeme with the greatest semantic contribution to the whole-word meaning is irrelevant for this decomposition process.

The present results suggest that the constituents of a compound do exert an influence on the whole-word processing: low frequency compound words containing high frequency second lexemes were responded to faster than compound words with low frequency second constituents. This frequency-based effect occurs independently of the frequency of the first constituent. These findings converge with recent evidence proposing second lexeme supremacy status (Juhasz et al., 2003). Taken together, all these results pose a problem for proposals favoring first and second constituent-based-decomposition (e.g. Libben, Gibson, Yoon, & Sandra, 2003; Sandra, 1990; Zwitserlood, 1994). Similarly, proposals that present the first constituent as the uniquely relevant one do not seem to be applicable to the processing of Basque and Spanish compounds (e.g., Inhoff, Radach, & Heller, 2000; van Jaarsveld & Rattink, 1983).

Despite the differences in morphological productivity in each of these two languages, we found a similar pattern of results for both languages (i.e., a second lexeme frequency effect). This favors a mechanism of processing compound words that scarcely depends on the properties of the languages. Instead, both Spanish and Basque readers seem to have developed a word-tail-decomposition mechanism, in which the morphological analysis of a poly-morphemic word is mainly centered on the latter part of the word –where morphological cues of words are generally posited (suffixes and/or second lexemes of compounds). This is consistent with the recent findings of Duñabeitia, Perea, and Carreiras (in press), who found a language-independent mechanism of morphological decomposition in transposed-letter effects across morphemes for both Basque and Spanish. In addition, the transposed-letter effect

vanishes when the manipulation occurred within the internal letters of the second constituent (but not the first constituent) of compound words (Duñabeitia, Perea & Carreiras, in preparation). Taken together, these results suggest that there is a morphological decomposition process that co-occurs with whole-word processing.

The converging evidence obtained from our lexical decision task experiment, together with data from other experiments using lexical decision task, naming paradigms, and eye movement measures (see Duñabeitia et al., in preparation, and Juhasz et al., 2003), strongly suggests a view of compound word processing mainly based on the predominance of the second lexeme in the recognition and segmentation processes. The parallel race model of Pollatsek et al. (2000) predicts a supremacy effect for the first lexeme only at initial stages of lexical access (see also Andrews et al., 2004, for similar results in English), whereas there is a supremacy for the second lexeme later in processing. At a very early stage of processing (undetected by lexical decision or naming tasks, but noticeable in first fixation duration in silent reading; see Perea & Pollatsek, 1998), the first constituent is processed. Then, later in processing, the second lexeme of the compound becomes relevant, not only in terms of frequency, but also in terms of syntactic properties: the inflectional morphological properties and the syntactic category of the whole compound word are specified by the second lexeme. The dissociation of the frequency effect for the initial and ending lexeme clearly supports a morphemic/lexical decomposition in lexical access.

The two-stage processing of the compound word could be explained by an activation-verification framework (e.g. Paap, Newsome, McDonald, & Schvaneveldt, 1982; see also Reichle, Pollatsek, Rayner, & Fisher, 1998, for an activation-verification model applied to reading). An activation-verification structure could be (efficiently) adapted to account for the present findings with transparent compounds in the following

way: First, the initial constituent would be treated as a separated unit (i.e. *tea* in *teacup*), activating the corresponding lexeme (*tea*) but also, to some extent, the whole compound (*teacup*). The verification process for the recognition of the whole compound cannot be carried out until the second unit shows up (the second constituent, *cup*). The second lexeme fires a new process, activating its corresponding lexeme and compound word (*cup* and *teacup*). Then, the verification process can be satisfactorily carried out, as the ending lexeme closes the orthographical, morphological, lexical, syntactic and semantic retrieval. Therefore, one could expect the frequency of the ending lexeme to play a large role at later stages (or when the task requires whole-word access, as in lexical decision), given that the second lexeme is the one that accomplishes the activation-verification procedure.

Juhasz et al. (2003) suggested that the second lexeme effect could be due to a (whole-word) meaning extraction procedure, centered in the final part of the compound word, which in English tends to be the head (see Andrews et al., 2004, for the same view). However, both the Spanish compounds (which tend to have an ending head) and Basque compounds (which tend to have an initial head) showed similar pattern of effects. This finding suggests a sequential decomposition of the constituents of the compound, interacting with the whole-word processing, in a blind-to-semantics manner (see Rastle et al., 2003, or Carreiras et al., 2006, for extensive discussion). This decomposition pathway is followed irrespective of the semantic information that can be extracted from the lexemes (no head lexeme effect was shown in the present experiment).

In sum, the present results showed a relevant status of the compound word's second constituent (independently of its status as the head of the compound), emphasizing the mandatory compound word decomposition process. Furthermore, we

found a similar pattern of data in two languages (Basque and Spanish) with very different morphological characteristics. Therefore, the present findings support a language-independent mechanism for the processing of compound words in reading.

References

- Andrews, S., & Davis, C.J. (1999). Interactive activation accounts of morphological decomposition: Finding the trap in mousetrap? *Brain and Language*, 68, 355-361.
- Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects. *Journal of Memory and Language*, 25, 726–740.
- Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal Cognitive Psychology*, 16, 285-311.
- Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from reading short and long Finnish compounds. *Journal of Memory and Language*, 48, 615-634.
- Duñabeitia, J.A., Perea, M., & Carreiras, M. (in press). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition*.
- Duñabeitia, J.A., Perea, M., & Carreiras, M. (in preparation). *Processing compound words: evidence from a multi-task approach*.
- Forster, K.I., & Forster, J.C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35, 116-124.
- Fowler, C. A., Napps, S. F., & Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory and Cognition*, 13, 241-255.
- Frost R., Kugler T., Deutsch A. & Forster K.I. (2005). Orthographic structure

versus morphological structure: principles of lexical organization in a given language. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31, 1293-326.

Frost, R., Grainger, J., & Rastle, K. (2005). Current issues in morphological processing: An introduction. *Language and Cognitive Processes*, 20, 1-5.

Frost, R., Forster, K.I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew: a masked priming investigation of morphological representation. *Journal of Experimental Psychology: Learning Memory, & Cognition*, 23, 829-856.

Giraud, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexicalexical representation of morphology. *Psychonomic Bulletin & Review*, 8, 96-101.

Gómez, P., Ratcliff, R., & Perea, M. (in press). A model of the go/no-go lexical decision task. *Journal of Experimental Psychology: General*.

Hyönä, J., & Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception & Performance*, 24, 1612-1627.

Inhoff, A.W., Radach, R., & Heller, D. (2000). Complex compounds in German: Interword spaces facilitate segmentation but hinder assignment of meaning. *Journal of Memory and Language*, 42, 23-50.

Jaarsveld, H.J. van, & Rattink, G.E. (1988). Frequency effects in the processing of lexicalized and novel compounds. *Journal of Psycholinguistic Research*, 17, 447-473.

Juhasz, B., Starr, M., Inhoff, A.W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions, and eye fixations. *British Journal of Psychology*, 94, 223-244.

Juhasz, B.J., Inhoff, A.W., & Rayner, K. (2005). The role of interword spaces in

the processing of English compound words. *Language and Cognitive Processes*, 20, 291-316

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language*, 61, 30–44.

Libben, G., Gibson, M., Yoon, Y-B., & Sandra D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, 84, 26-43.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word-superiority effect. *Psychological Review*, 89, 573–594.

Perea, M., & Carreiras, M. (2006). Do transposed-letter effects occur across lexeme boundaries? *Psychonomic Bulletin and Review*, 13, 418-422.

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 767-777.

Perea, M., Urkia, M., Davis, C. J., Agirre, A., Laseka, E., & Carreiras, M. (in press). E-Hitz: A word-frequency list and a program for deriving psycholinguistic statistics in an agglutinative language (Basque). *Behavior Research Methods*.

Plaut, D.C., & Gonnerman, L.M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, 15, 445-485.

Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, 20, 261-290.

Pollatsek, A., Hyona, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 820–833.

Rastle, K.G., Davis, M.H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review*, 11, 1090-1098.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125-157.

Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology*, 42A, 529–567.

Sebastián-Gallés, N., Martí, M. A., Carreiras, M., & Cuetos, F. (2000). *LEXESP: Léxico informatizado del español*. Barcelona: Edicions Universitat de Barcelona.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, 7, 263-272.

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, 9, 271-294.

Zalbide, M. (1992). *Hitz-Elkarketa/4. Hitz elkartuen osaera eta idazkera*. Bilbao: Euskaltzaindia.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, 9, 341–368.

Table 1

Mean frequency (per million) and length of the compound words, and mean frequency and length of the constituents of the experimental items in the experiment. Medians are presented in parentheses.

| | <u>Frequency</u> | | | <u>Length</u> | | |
|--------------------------|------------------|-------------------|-------------------|---------------|-----------|-----------|
| | <u>Word</u> | <u>L1</u> | <u>L2</u> | <u>Word</u> | <u>L1</u> | <u>L2</u> |
| <u>Basque Compounds</u> | | | | | | |
| High-High | 4.42 | 1610.55 (1209.82) | 1626.91 (1637.20) | 8.64 | 4.18 | 4.45 |
| High-Low | 4.34 | 1849.09 (732.74) | 96.30 (83.93) | 8.73 | 4.27 | 4.45 |
| Low-High | 3.86 | 114.85 (94.70) | 1441.00 (1613.18) | 8.27 | 4 | 4.27 |
| Low-Low | 3.34 | 90.38 (88.07) | 167.28 (162.34) | 8.73 | 4.45 | 4.27 |
| <u>Spanish Compounds</u> | | | | | | |
| High-High | 2.81 | 540.64 (197.95) | 399.72 (309.83) | 9.08 | 4.5 | 4.17 |
| High-Low | 2.13 | 406.70 (295.36) | 39.07 (41.07) | 10.08 | 4.08 | 5.88 |
| Low-High | 1.79 | 29.52 (24.02) | 343.05 (223.13) | 10 | 5.58 | 4.17 |
| Low-Low | 2.17 | 30.50 (34.38) | 27.19 (24.56) | 10.08 | 5.58 | 4.50 |

Table 2

Mean lexical decision times (in ms) and percentage of errors for compound word targets in the experiment

| | <u>Basque Compounds</u> | | <u>Spanish Compounds</u> | |
|------------------------------|-------------------------|------------|--------------------------|------------|
| | Response Time | Error Rate | Response Time | Error Rate |
| <u>Constituent Frequency</u> | | | | |
| High-High | 870 | 2.1 | 857 | 17.9 |
| High-Low | 930 | 7.2 | 883 | 18.7 |
| Low-High | 849 | 5.9 | 870 | 10.6 |
| Low-Low | 932 | 9.1 | 894 | 14.9 |

Footnote

1. Due to the canonical syntactic order in Basque (Subject+Object+Verb), which is not the same as in Spanish (Subject+Verb+Object), the same compound words behave oppositely in these two languages in many cases. For instance, the Spanish for *birthday*, *cumpleaños* (cumple+años, fulfill+years) when translated to Basque changes into *urtebetetze* (urte+betetze, years+fulfillment).
2. It should be noted that all the constituents of the compounds are meaningful words by themselves. Most of the compound word structures were Noun+Noun, although some were Adjective+Noun or Noun+Adjective. In any case, this fact is not supposed to affect the results (see Juhasz et al., 2005; Juhasz et al., 2003). Only a small percentage of the compounds (13 of 92) was partially opaque (at least one of the constituents was always related to the whole-word meaning), although this fact does not seem to affect response times (Libben, Gibson, Yoon, & Sandra, 2003). These semi-opaque compounds were distributed over all conditions (no more than 4 in each group).
3. As Michael Cortese pointed out, error rates for the Spanish compounds were higher than for the Basque compounds. This difference was probably due to the fact that --unlike Basque-- Spanish has a restricted number of compounds. To test this hypothesis, we conducted an analysis on the error rates for the Spanish filler items. If the proportion of compound words in the language is the factor that leads to a higher error rate, then responses to non-compound filler stimuli should have been more accurate than the responses to compound words. The percentage of errors for the filler words was 3.3%, revealing that Spanish readers correctly recognized non-compound monomorphemic and polymorphemic

words matched in frequency and length to the experimental words. Furthermore, the main finding of the Spanish sub-experiment (i.e., an effect of the frequency of the second lexeme) was recently replicated with a go/no-go lexical decision task (Duñabeitia, Perea, & Carreiras, in preparation). As usual, error rates in the go/no-go task were quite low (Gómez, Ratcliff, & Perea, in press): 3.3% for the High-High compounds, 5.0% for the High-Low compounds, 1.5% for the Low-High compounds, and 0.9% for the Low-Low compounds. This reveals that Spanish readers satisfactorily recognized the same set of materials used in the experiment reported in the present manuscript.

Author's notes

Jon Andoni Duñabeitia was the recipient of a post-graduate grant from the Basque Government. The research reported in this article has been partially supported by Grants SEJ2004-07680-C02-02/PSIC, SEJ2005-05205/EDU, and SEJ2006-09238 from the Spanish Ministry of Education and Science. The authors would like to thank Richard Gerrig, Lee Wurm, Michael Cortese, Barbara Juhasz, and Itziar Laka for their helpful comments on earlier drafts. Thanks are also due to Edurne Laseka for selecting the Basque stimuli and conducting the Basque sub-experiment.