COMPOUND WORD PROCESSING: DECOMPOSING, COMPETING AND RECOMPOSING

Cannarella Roberto (matr. 616400)

Compound words (or *compounds*) can be defined as lexemes or lexical units that are internally composed of two (or more) lexemes already attested in a language, that is, that are composed of at least two bases being themselves words (Culpeper 2009). *Compounding* is the name of the morphological process through which a new compound word is derived. The *head* (or *head*-word) of a compound is the one word, among those composing the compound: *(a)* which determines the lexical category of the whole unit (e.g., the head of [to] sleepwalk is [to] walk, which is a verb, thus [to] sleepwalk is a verb in turn; sleepwalker, instead, is a noun, as is its head walker); *(b)* which is placed, in English, in the 'right-side' of the compound (Libben, Gagné & Dressler 2020); *(c)* which, in the so-called *endocentric compounds*, is the 'carrier' of the basic meaning which is being modified by the non-head word (the concept of sleepwalking is a troponomy of the concept of walking, which is being modified in the fact that it occurs during one's sleep). In such cases, the meaning of an endocentric compound can be inferred through the re-composition of the meanings of the composers: such compounds are said to be *semantically transparent*. They are distinguished from *exocentric compounds*, whose meaning is usually not compositional, as in the case of four-eyes. These are said to be *semantically opaque*.

Libben *et al.* (2020) use the term *dual nature* for referring to the fundamental feature of compounds – namely, the fact that compounds contain "very recognizable sub-elements" (regardless of their being transparent or opaque) and are, at the same time, used as structures "with specific whole-word meanings" (*ibidem* 2020: 338). The link (in particular, the cognitive *interference* or *aid*) between these two aspects (the easily recognisable elements composing the whole, and the whole itself) is a crucial topic in psycholinguistic research on compounds, and the effect of semantic transparency on the ease with which the processing of compounds occurs has been (and is being) widely studied. Libben *et al.* (2020) state that, in general, compounds that are semantically transparent appear to be easier to process (i.e., they are processed more quickly[1]) than opaque ones. Starting from this generally accepted idea, research has focused on exploring compounds processing in more depth. For instance, it has focused on determining the 'nature' of the influence that the components have in the processing of the compound. More precisely, the question is: does accessing the components helps the speaker in processing the whole unit purely thanks to the lexical representation the components provide, or is a semantic/conceptual influence to be taken into consideration? If semantic representation has a role in compound processing, does it occur for *any* kind of compound (i.e., for both transparent and opaque compounds)? Considering that (as already stated) some compounds are easier to process than others, how can this variable difficulty be quantitatively measured, and what does it depend on?

THE COST OF DECOMPOSING AND RECOMPOSING

Ji, Gagné & Spalding (2011) underline the importance of taking into consideration the relationship between the *cost* of the *lexical decomposition* of multimorphemic words (in our case, compound words[2]) and *lexical storage*. In fact, even though it is certain that decomposition involves some kind of additional effort and cost,

[1] The processing time is estimated, for instance, by considering gaze durations in eye-tracking studies (with the gazes lasting more for opaque compounds), or by considering the latencies in lexical decision tasks (with the latencies being longer for transparent compounds).

[2] Ji, Gagné & Spalding (2016) give two reasons why focusing on compound words can be useful for inferring information on the decomposition of multimorphemic words in general: the first reason is that compound words tend to be low frequency, thus speakers are more likely to access them by decomposition; the second reason is that this same decomposition is likely to be more effortful than it is when it comes to, for instance, derived words, because in the latter case speakers can rely on recurrent morphemes (e.g., a suffix) which appear systematically in several derivatives, thus 'guiding' the speaker in processing the whole unit.

several studies have suggested that, despite this, complex words *are* processed through it. Studies that have compared the speed of compound and monomorphemic words have shown that, in some cases, the processing of the former appears to be quicker and easier, which also confirms that, despite the aforementioned cost, decomposition can yield processing advantages.

It should be also noted that the processing of a compound word involves not only the decomposition of it but also the consequent *integration* between its components. Whereas decomposition "involves accessing the constituent word forms", integration "involves integrating the constituents" (Ji *et al.* 2011: 408), that is, it involves re-combining them. Studies suggest that the process of integration could correspond to a process of *semantic integration*, which means that the inference of the whole unit's meaning could be drawn through a process of semantic composition based on the semantic traits of the constituents. In this case, the processing benefits should be available only for transparent compounds (whose processing, as already mentioned in [§2], is generally considered to be faster), whereas the processing of opaque compounds should have a higher cost, which would result in longer reaction times (RTs). The other possible type of integration is the purely *lexical* one, in which case "facilitating decomposition should yield a processing benefit for *both* semantically transparent and opaque compounds" (*ibidem*), since it may rely on the co-activation of only the constituents' lexical representations, and not of their meanings/semantic traits. The six experiments conducted by Ji *et al.* (2011) compare the processing of (two-constituent) compounds with the processing of monomorphemic words in order to identify which word type is processed more easily[3] (in this way focusing on the benefits and cost of compound processing) and to further investigate the nature of the integration.

In the following table, a brief recap of the results is presented:

***N* of Exp.**	**Aim(s) and method of the specific experiment**	**Results**
Exp. *1*	- It compares the processing time of transparent compound words with the processing time of monomorphemic words. - According to the *dual-route theory* (described, for instance, in Baayen *et al.* 1997), both the decomposition route and the whole-word route are available during the processing of a compound, and the quickest route is the one generally followed by the speaker. Those compounds where both access routes are available equally (i.e., transparent compounds) should correspond to those situations where the processing is particularly eased and notably quicker than the processing of monomorphemic words. The experiment explores the validity of this notion.	- Compounds were responded to more quickly than monomorphemic words. - The advantage of compounds does not appear to be limited to the low-frequency ones. This finding contrasts with both the dual-route theory and Libben's (1998) theory, that posited that low-frequency compounds should be more positively influenced in their processing than the high-frequency ones.
Exp. *2*	- It compares the processing time of transparent compound words with the processing time of monomorphemic words. - In Exp. 1, the filler items were constructed by combining a word and a nonword, so the first might have influenced the participants in their responses. In this experiment, filler items have been constructed from two words, in order to investigate this potential influence.	- The type of filler items does not have an influence on the results: even in this case, compounds were responded to more quickly than monomorphemic words.

[3] It should be noted that both the word frequency and the word length are the same for each pair of words (monomorphemic word-compound word), in order to focus more precisely on the dynamics of their processing.

Exp. *3*	- It compares the processing time of compound words with the processing time of monomorphemic words. - In the first two Exp., the compounds were all transparent. Here, the comparison is between monomorphemic words, transparent compounds and opaque compounds.	- Transparent compounds were responded to more quickly than opaque compounds, which in turn were responded to more quickly than monomorphemic words. - Constituent frequency also appears to be effective, and to ease the processing of both transparent and opaque compounds. This indicates that the responses reflected an influence *on the lexical level* (and not on the semantic one).
Exp. *4*	- It focuses on how the meaning of a compound is cognitively composed starting from its constituents. In particular, it concentrates on the cost of semantic composition, which should be influential especially for opaque compounds. - To test and stimulate semantic composition, a space has been inserted between the constituents.	- Transparent compounds were responded to more quickly than monomorphemic words, whereas opaque compounds were processed as quickly and as accurately as monomorphemic words (consequently, transparent compounds were responded to more quickly than opaque compounds). Additionally, first constituents being high frequency boosted the response to transparent compounds and slowed the response to opaque compounds. The experiment hints that semantic composition, when encouraged (in this case, by inserting a space between the constituents), is especially effortful for opaque compounds.
Exp. *5*	- It focuses on how the meaning of the compound is cognitively composed starting from its constituents. In particular, it concentrates on the cost of semantic composition, which should be influential especially for opaque compounds. - To promote semantic composition, the two constituents of compounds have been presented in different colours.	- In this experiment too, transparent compounds were responded to more quickly than monomorphemic words, opaque compounds, instead, lost the processing advantages of the first experiments, thus showing that the 'forced' involvement of the cognitive and semantic level makes their processing more effortful.
Exp. *6*	- As in Exp. 4-5, it focuses on how the meaning of the compound is cognitively composed starting from its constituents and concentrates on the cost of semantic composition, which should be influential especially for opaque compounds. - The experiment does not use any visual manipulation for stimulating morphological decomposition and the consequent semantic composition. In fact, the aim is to detect whether the mere inclusion of nonword compound fillers (constructed from two real words) is sufficient for removing the processing advantage that opaque compounds presented in Exp. 3.	- This experiment confirms the findings observed in Exp. 4-5: the processing of transparent compounds is achieved more quickly than the processing of opaque compounds, which in turn have lost the processing advantage over monomorphemic words attested in Exp. 3.

Considered all together, the results confirm the importance of "distinguishing between lexical and semantic levels of representation" (Ji, Gagné & Spalding 2011: 419). Exp. 3, in particular, highlights that the access to the *lexical* entries of constituents occur for both opaque and transparent compounds. Yet, when *semantic* composition is encouraged (by dividing the two constituents in Exp. 4, by presenting the constituents in different colours in Exp. 5, through the insertion of nonword compound fillers alone in Exp. 6), the difficulty in re-creating the meaning of the unit starting from its constituents leads to a slower processing of the opaque units, which confirms that compound processing is accomplished through some sort of representation at the semantic level, too.

The explanation provided by the authors stresses the importance of the *morphological computation* in compound processing: in fact, the fundamental idea is that there is *always* an attempt to compute a compound's meaning starting from its constituents. Such attempted meaning (temporally) co-exists with the activation of the lexicalized meaning of the unit, with which it is in competition[4]. In the case of transparent compounds, the derived meaning and the lexicalized meaning are likely to overlap, that is, the competition should be minimal, which should explain the general ease in their processing. In the opposite case of opaque compounds, constructed meanings, after having been computed, should be 'put aside' in favour of the lexicalised meaning. In particular, when semantic composition is encouraged and the access to the lexicalised meaning is consequently discouraged (Exp. 4, 5, 6), the computed meaning of an opaque compound is likely to be more distant from the lexicalised meaning (because it has been inferred without its influence, in that it has been computed from the mere combination of the two constituents). This framework is, according to the authors, compatible with both the recent studies focusing on the role of computation and the studies that acknowledge the evocation, during processing, of multiple interpretations.

COMPETING AND COMPUTING

The previous section has presented a framework that gives an explanation to the different 'dynamics' of compound processing, positing that both the lexicalised meaning of the unit and its computed meaning have a crucial role and are assessed during compound processing. In this section, the description of the work by Schmidtke *et al*. (2016) will allow to explore *how* the computed meaning is cognitively inferred and computed. The focus is, in particular, on *relational structures*. Relational structures are the cognitive-semantic links that possibly occurs between lexemes and, in the case of compounds, between constituents. Examples of this kind of relation include, for instance, the FOR relation (such as in teapot, which is a pot FOR making tea). According to Schmidtke *et al.* (2016), research suggests that the availability of relational structures influence the processing of compounds, and that "during the interpretation of modifier-noun […] compounds, relational structures *compete* for selection" (*ibidem*: 558, stress added). The term 'competition' implies that the stronger it [*the competition*] is (between the relational structures that one may 'attach' to a compound), the harder it is to process the compound.

The measure used by Schmidtke *et al.* (2016) for formalising the degree of competition is the information-theoretic measure of *Entropy* (*H*), which is defined as follows (*i* stands for the paradigm of semantic relations):

$$H = -\sum p_i \, log_2 \, p_i$$

It is a measure that can be used for estimating the degree of competition between the possible semantic relations associated to a compound because its value increases with the number of multiple distinct relations and when the possibility of their occurring is more similar. The higher the Entropy, thus, the higher the competition and the uncertainty in computing, and the harder the processing. When it comes to using Entropy for relation-based competition, the focus has been, for instance, in assessing whether the 'richness' of the relational structures attested for a modifier (such as TEA in TEAPOT) influences the processing (which is the case, since more relational structures equal to higher Entropy, thus to harder processing). It should be noted that Entropy has also been used for measuring the role of family size in tasks such as the lexical decision task, where lexical access is not dependant on semantics. In these experiments, Entropy has confirmed one typical assumption of morphological processing, that is, that morphological (and lexical) families/paradigms with several members being highly frequent (i.e., families having higher Entropy) are easier to process. In general, in psycholinguistic research, Entropy can be calculated using data from a *possible relation task*, which is a task in which

[4] The authors underline an important difference with the aforementioned theories describing two possible routes of access: in their framework, the ones competing are the possible *meanings* that have been computed or lexically retrieved, not two distinct *routes*.

participants are presented pairs of nouns and have to choose the most likely meaning for each of it, choosing among a set of potential relational structures.

Data derived from this kind of experiment have been used for two measures. The first is *Relational Diversity Ratio* (RDR), defined "as the number of distinct relational interpretations given to a compound" (*ibidem*: 560), the second is the *Relational Relative Entropy* (RRE), which is a measure for defining how probable the set of relations identified for *one* specific compound is, compared to "the probability distribution of relations estimated across a larger set of compounds" (*ibidem*). These measures have been used, for instance, for research on transparent compounds. Transparent compound processing has been shown to be harder when both RDR and RRE are high, because their being high corresponds to the subject having to choose between a *large* set of possible relations, that are also all *strong* candidates. Nevertheless, both RDR and RRE differ from Entropy in that the latter is used by Schmidtke *et al.* (2016) for exploring how the competitivity *between the relations* of a compound influences its processing. This occurs by taking into consideration, for instance, how wide the frequency gap between the chosen relations is: compounds with one particular relation having been systematically chosen in a possible relation task should have lower Entropy than compounds for whom multiple relations have been chosen with the same frequency – the former should, thus, be easier to process.

The results of the analysis by Schmidtke *et al.* (2016) confirm this assumption. In general, compounds for which more relations have been chosen in possible relation tasks (i.e., with higher *H*) need more time for their processing. According to the authors, Entropy is also slightly more precise as a measure for predicting differences in RTs than Relational Diversity. The authors also state that the influence of Entropy (that is, the 'activation' of a competition between relations) is tangible in those cases when relations have been chosen more than once and are, quantitatively, between 4 and 16.

Considered together, the work by Ji *et al.* (2011) and Schmidtke *et al.* (2016) shows the peculiarity of the cognitive representation of compounds, words that are "both greater than the sum of their parts and greater than the division of their wholes" (Libben *et al.* 2020: 349). The results hint that the processing of compound words seems to be based on both a lexical and a semantic representation, the latter resulting as more difficult and effortful in those cases where more possible kinds of relations compete one with the other (that is, in the cases of opaque compounds). The 'level' of this competition can be measured through a measure taken from information theory, i.e. Entropy, which has been shown to be a reliable measure of how the co-activation of multiple different semantic relations can inhibit and slow down the processing. Focusing on the semantic relations evoked by a compound entails trying to work at a theoretical and highly conceptual level of semantic representation (Schmidtke *et al.* 2016: 568 talk of semantic relations as a "high-level semantic process"), and will involve exploring the kinds of relations and cognitive 'constructions' between lexemes that can and cannot be established.

REFERENCES

Baayen, R.H., Dijkstra, T. & Schreuder, R. (1997). *Singulars and plurals in Dutch: Evidence for a parallel dual-route model*. Journal of Memory and Language, 37, 94-117.

Culpeper, J., Katamba F., Kerswil, P., Wodak, R. & McEnery, T. (eds). (2009). *English Language. Description, Variation and Context*. Hampshire/New York: Palgrave Macmillan.

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

Libben, G., Gagné, C. & Dressler, W.U. (2020). *The representation and processing of compound words*. In V. Pirrelli, I. Plag & W.U. Dressler (eds.) Word knowledge and word usage: a cross-disciplinary guide to the mental lexicon, 336-352.

Schmidtke, D., Kuperman, V., Gagné, C.L. & Spalding, T.L. (2016). *Competition between conceptual relations affects compound recognition: The role of entropy*. Psychonomic Bulletin & Review 23(2). 556–570.

Ji, H., Gagné, C.L. & Spalding, T.L. (2011). *Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds*. Journal of Memory & Language, 65, 406–430.