

V International Conference on Corpus Linguistics (CILC2013)

## Incorporating frequency information in a collocation dictionary: Establishing a methodology

Orsolya Vincze, Margarita Alonso Ramos\*

*Universidade da Coruña, Campus da Zapateira, A Coruña 15071, Spain*

---

### Abstract

This paper describes the methodology applied to incorporate frequency information in the *Diccionario de Colocaciones del Español*, a dictionary developed within the framework of *Explanatory and Combinatorial Lexicology*. The main reasons for including lexical frequency information in DiCE were to enrich the dictionary content with information potentially useful for language learners, and to filter currently listed collocates. Obtaining and providing frequency information on collocations is not a straightforward task. In our case, we have opted for assigning a frequency score to each individual lexical unit which constitutes the base of a collocation, while collocations are assigned a frequency score taking into account both the absolute frequency of the collocation itself and the frequency of the lexical unit constituting its base. As a result, collocates are presented in the dictionary under the lexical entry of the base organized according to their syntactic pattern and their meaning, and ordered according to their frequency. Through assigning frequency scores to collocations we also foresee obtaining a frequency list of all collocations included in DiCE, valuable for language teaching and testing purposes.

© 2013 The Authors. Published by Elsevier Ltd.  
Selection and peer-review under responsibility of CILC2013.

*Keywords:* lexical frequency; collocations; collocation dictionary

---

### 1. Introduction

The present paper describes the methodology applied in order to incorporate frequency information in the *Diccionario de Colcaciones del Español* (DiCE, Alonso Ramos, 2004; Alonso Ramos et al., 2010; Vincze et al., 2011). DiCE has been developed within the framework of *Explanatory and Combinatorial Lexicology* (ECL,

---

\* Corresponding author. Tel.: +34-981-167-000, ext. 1776  
E-mail address: [ovincze@udc.es](mailto:ovincze@udc.es)

Mel'čuk et al., 1995). In ECL, collocations are defined as phraseological units constituted by two elements, 1) the *base*, which is chosen freely by the speaker, according to its meaning, and 2) the *collocate*, chosen in accordance with the restrictions imposed by the base. For instance, as it can be observed in (1), the prototypical adjectives used to express the meaning 'intense' together with the nouns *fiebre* 'fever', *enfermedad* 'illness', and *dolor* 'pain', are not necessarily interchangeable.

- (1) a) *fiebre alta* 'high fever', *enfermedad grave* 'serious illness', *dolor profundo* 'deep pain'  
 b) \**fiebre grave* 'serious fever', \**enfermedad profunda* 'deep illness', \**dolor alto* 'high pain'

The content of DiCE has been derived from corpus data, specifically, from the *Corpus de Referencia del Español Actual* (CREA). However, when deciding on which lexical combinations to include in the dictionary, frequency information was not taken into account. As a result, while certain pieces of linguistic data incorporated in DiCE are representative of Spanish language, others may be the result of the mere stylistic choice of the author of a specific text found in the corpus. Such is the case, for instance, with some of the 33 adjectives currently indicated by DiCE to express the meaning 'intense' in combination with the noun *odio* 'hatred'. Consequently, the reason why we opted for incorporating frequency data to the content of DiCE was exactly that of organizing combinatory information in a more informative way. In addition, information on lexical frequency not only allows us to present collocations in a more useful way to dictionary users, but it also provides information for the development of teaching material on collocations, and can be applied as an objective filter of dictionary content.

In what follows, we briefly discuss the importance and usefulness of lexical frequency information in language teaching and learning, as well as certain questions that should be raised when it comes to measuring lexical frequency, especially that of collocations. After this, we introduce the methodology we have established in order to obtain frequency information both in the case of lexical units (LU) that appear as bases of collocations and in the case of collocations included in DiCE. Finally, we conclude the paper with brief considerations concerning the state of language resources, specifically corpora, available for Spanish, and its implications for obtaining linguistic data such as lexical frequency information.

## 2. Lexical frequency and collocations

Frequency dictionaries and lists of frequent words are in general associated with language teaching. Given that it is commonly believed that if a word is more frequent, it is also more useful from the point of view of the language learner, frequency lists have often been used for determining which words a learner should acquire at a given level (e.g. Thorndike & Lorge, 1944; West, 1953; Hindmarsh, 1980; Coxhead, 2000 for English and García Hoz, 1953 for Spanish). From our point of view, in order for the language learner to communicate efficiently, in addition to learning the most frequent words, it is also necessary to know how they are used, which implies being familiar with what other words they can be combined with. That is why we consider that it is important to develop a frequency list of collocations.

We agree that the frequency of use of a lexical item constitutes an especially valuable piece of information when it comes to developing teaching material. However, there are a number of factors to be considered when assessing the extent to which frequency information should be given preference over other criteria. Firstly, lexical frequency is generally measured in corpora, consequently, the content of the corpus has a direct impact on the data obtained. Accordingly, it is important to take into consideration the representativeness of the corpus, not only with respect to the language as a whole, but also the specific language variant or register to be taught to learners. In this respect, we can say that the teaching or lexicographical aims not necessarily coincide with the objective data extracted from the corpus, in other words, it is not sufficient to offer a mere x-ray photograph of the corpus itself, without observing other criteria. Consider that in the corpus *esTenTen* (Kilgarriff et al., 2004) we have found 37 cases of *miedo infundado* 'unfounded fear', 194 cases of *miedo atroz* 'terrible fear' and 156 occurrences of *miedo profundo* 'deep fear', all of them being collocations that express a very recurrent meaning as compared to *miedo escénico* 'stage fright', which expresses a much more specific meaning, and, nevertheless appears with a much greater frequency; 618 cases, in the same corpus. We believe that in view of this specific example, we need to raise

the question of whether it is legitimate to use lexical frequency as a unique criterion to determine what is more useful for the language learner, or what combinatory data to include in a collocation dictionary, and, for that matter, what the role of the teacher's or the lexicographer's intuition should be in deciding on such matters (see e.g. *formula teaching worth score* used by Simpson-Vlach & Ellis, 2010).

Secondly, when it comes to developing a frequency list, we also have to address the question of what types of elements we should include in given list. Probably for practical reasons, the majority of existing frequency lists provides information on the frequency of *word forms* or *lemmas*, instead of *lexical units* (LU); some exceptions are West (1953), Hindmarsh (1980) and Capel (2010, 2012). The reason for this is certainly that it is easier to count occurrences of word forms or even lemmas in a corpus, than those of each meaning of a given lemma. However, if a wordlist is aimed for teaching purposes, it is desirable that it show information pertaining to individual LUs. If we assign a single frequency score to, for instance, the lemma *cólera*, we avoid providing information with respect to its different meanings referring to 1) the emotion 'anger' and 2) the disease 'cholera'. In fact, in certain cases, the lack of disambiguation of lemmas can lead to false conclusions. For instance, both in Almela et al. (2005) and Davies (2006) the form *vez* 'time, occasion' appears among the most frequent nouns. This surprising result might be a consequence of the fact that this word appears very frequently as part of the phraseological expressions *a la vez* 'at the same time', *de una vez* 'once and for all', *de vez en cuando* 'sometimes', *tal vez* 'maybe', etc. Consequently, it seems to be clear that in spite of the complications implied, it is important to take into account the frequency of individual LUs when developing a frequency list, however, we should note that, this again not necessarily can be taken as a definitive criterion when it comes to determining the order in which we should present LUs to learners. For example, in the case of the adjective *ligero* 'light', intuition can lead us to choose the meaning 'weighing little' as the more basic sense to be taught first (cf. Casso, 2010: 82), in spite of the fact that, as shown in Alonso Ramos (2012), the meaning 'not intense' (e.g. *ligero retroceso* 'light set-back/relapse', *aumento ligero* 'light growth', etc.) appears frequently in corpus data.

Finally, in the specific case of obtaining frequency information on collocations, naturally, we have to take into account the working definition of *collocation* being used. Shin and Nation (2007), for instance, present a list of the 100 most frequent collocations of English; however, from our point of view, their list contains a set of expressions with very different characteristics: free combinations, such as *this morning*, *I think*, etc. or expressions that can be considered discourse markers, such as *you know*, *thank you*, etc. The content of this list, as we have suggested, corresponds to the specific definition of collocation of the authors, which includes the condition that the elements of a collocation must be immediate constituents, which for the most part means that they are linearly adjacent in the text. Within our framework, however, collocations are defined as restricted lexical combinations with no regard to whether their elements constitute an adjacent string. As we will see below, this last consideration has significant implications when it comes to extracting frequency information on collocations.

### 3. Assigning a frequency score to bases of collocations

In the course of incorporating frequency information in DiCE, as a first step, we assigned a frequency score to the bases of collocations. This task was carried out using CREA, more specifically, a portion of this corpus containing texts from the decade between 1990 and 2000 coming from Spain.

#### 3.1. Manual queries in CREA

When carrying out manual queries using the web search interface of CREA, we encountered a number of difficulties due to the absence of lemmatization in the corpus, such as the lack of resolution of syntactic category ambiguity in the case of certain bases (e.g. the form *odio* can correspond with the noun 'hatred' or the first person singular of the verb 'hate'), and the fact that some of them can appear as part of an idiom (e.g. the form *pesar* which corresponds to the noun 'regret' often appears in the idiom *a pesar de* 'in spite of'). As a consequence, in the search terms used in the queries (see 3) we had to represent all possible forms of a lemma – including singular and plural beginning in upper case or lower case – and, at the same time, we excluded frequent strings where the same form did not correspond to the lemma of the base, such as in the case of idioms.

- (3) pesar O pesares O Pesar O Pesares Y NO a pesar de Y NO A pesar de Y NO a pesar del Y NO A pesar del

This way we acquired information with respect to the absolute frequency of the lemmas found in DiCE in the CREA subcorpus. In addition, we used the filter options of the corpus query interface in order to obtain a sample containing approximately 100 concordances in the case of each lemma.

### 3.2. Manual semantic disambiguation

As we have mentioned earlier, we consider that, from the point of view of language teaching, frequency information included in a dictionary or in a frequency list should concern individual lexical units, not lemmas or word forms. Given that at present automatic semantic disambiguation delivers rather poor results, in order to obtain information on the frequency of the different meanings of a lemma, we necessarily have to resort to manual annotation. Consequently, we manually disambiguated the concordance samples of each lemma, obtaining the frequency of each individual LU in a reduced random sample. This piece of data was used to estimate the frequency (see 4) of the LU in the complete subcorpus.

- (4) Estimated frequency of LU =  $(\text{FreqLU}_{\text{sample}} / \text{sample size}) \times \text{FreqLemma}_{\text{subcorpus}}$

### 3.3. Frequency bands

As a final step, we assigned each LU to one of the frequency bands established by Almela et al. (2005). In Table 1 we show the criteria used in the case of each frequency band, together with the number of corresponding LUs found in DiCE.

Table 1: The number of LUs in DiCE assigned to each of the five frequency bands defined by Almela et al. (2005)

Frequency band	Criterion in relative freq./million words	Number of LUs
low	$\text{Freq} < 3$	309
moderate	$3 \leq \text{Freq} < 11$	157
prominent	$11 \leq \text{Freq} < 26$	50
high	$26 \leq \text{Freq} < 75$	28
very high	$75 \leq \text{Freq}$	6



As Table 2 shows, frequency information is displayed in the entry of each LU in DiCE via indicating the name of the corresponding frequency band.

Table 2: Frequency information of different LUs of *amistad* ‘friendship/friend’ in DiCE

Lexical Unit	Quasi-synonyms	Example sentence	Frequency
amistad 1	<i>camaradería</i> ‘friendship’	En Sahagún contaba con la <i>amistad</i> y la hospitalidad de Martín y de Zulema. ‘In Sahagún he could count on Martín's and Zulema's <i>friendship</i> and hospitality.’	high
amistad 2a	<i>amigo</i> ‘friend’	Me presentó a una <i>amistad</i> de la infancia. ‘He introduced me to a childhood <i>friend</i> .’	moderate
amistad 2b	<i>contactos</i> ‘contact’	Tiene amistades en el ministerio que lo apoyarán. ‘He has contacts in the ministry who will support him.’	low

#### 4. Assigning a frequency score to collocations

The retrieval of frequency information concerning collocations requires a large corpus which should also be lemmatized. For this reason, we opted for using the web corpus *esTenTen* available through the web interface of the Sketch Engine (Kilgarriff et al., 2004), consisting of more than two billion words and lemmatized with Tree Tagger (Schmid, 1994). In order to make sure that the content of this corpus is compatible with the CREA subcorpus we used in the case of LUs (see above), we compared the relative frequency in one million words of the lemmas of DiCE in the two corpora, and concluded that in effect they were comparable.

##### 4.1. Searching for collocations in corpus

Collocations are multiword expressions, and, as such, they pose certain difficulty when it comes to identifying them in corpus. On the one hand, as it can be seen in (4a), the words constituting the elements of a collocation are not necessarily adjacent, and, on the other hand, the same words not necessarily form a collocation, see (4b).

(4a) Sólo Dios sabe el *miedo* que les *entró*. lit. ‘God only knows the *fear* that *entered* them.’

(4b) Hay personas que anteponen el *miedo* a *entrar* a un quirófano al complejo que les ocasiona el aspecto de su nariz. ‘There are people who place their *fear* of *entering* the operating room before their complex about their nose.’

A possible solution to this problem is the use of grammars in the corpus queries, which in fact constitutes the strategy used by the Sketch Engine itself for obtaining Word Sketches, in other words, lists of words of a given syntactic category that co-occur with a specific word. The Sketch Grammars integrated with different corpora on the Sketch Engine interface work rather well when it comes to retrieving combinations of two words joined by a particular syntactic relation (subject-verb, verb-object, noun-modifier, etc.). However, in terms of recall, these grammars are somewhat deficient, given that they allow for very few variations in the syntactic pattern. We believe that this is so, because, by definition, Sketch Grammars always contain an unknown variable, in the sense that they serve to find, for example, which verbs co-occur frequently with the noun *amistad* ‘friendship’ as subject. On the contrary, in our case, we are interested in finding the greatest number of possible examples in the case of each particular collocation. In other words, both elements of the collocation are already known, and we are interested in how many times a given combination, e.g. the noun *amistad* ‘friendship’ as subject of the verb *unir* ‘bind’ appears in the corpus. For this reason, we developed our own grammars, which allowed us to retrieve examples of a collocation that correspond to a wider variety of patterns. For instance, in the case of collocations of the type subject + verb, we allow patterns where the subject precedes the verb (5a), where the verb appears in a subordinate clause (5b), and where the subject follows the verb in the linear order of the sentence.

(5a) Una recíproca *amistad* *unirá* ya siempre dos almas gemelas. ‘A reciprocal *friendship* will always *bind* these twin souls together.’

(5b) Pero ante todo la *amistad* fraternal e inquebrantable que me *une* al cantante de la banda. ‘But especially the brotherly and unbreakable *friendship* that *binds* me to the band’s singer.’

(5c) Telmo Rodríguez es mi socio y me *une* a él una muy buena *amistad*. ‘Telmo Rodríguez is my associate and there *binds* me to him a very good *friendship*.’

##### 4.2. The process of assigning frequency information to collocations

The process of assigning frequency information to the collocations included in DiCE consists of three main steps. Firstly, once the grammars used for querying the corpus have been formulated, we launch automatic queries in *esTenTen* through the Sketch Engine API. The results of the queries are saved in plain text files convertible to tables in order to facilitate reading, in the following step.

In the second step, the manual revision, we limit ourselves to resolve problematic cases we foresee since the development of the grammars. Sometimes it was necessary to manually modify the query grammars in order to adopt them to specific problematic cases, such as with the noun *odio* ‘hatred’ or the adjective *malvado* ‘evil’, which are not correctly lemmatized in the corpus, and, therefore, it is not possible to search them via the lemma, but we need to carry out a query specifying the possible word forms. Another problematic case is that of the noun *celo* because there is a lemmatization mismatch between the corpus, where it corresponds to one single lemma and DiCE, where it is split into two lemmas *celo* ‘zeal, enthusiasm, rut, heat’ and *celos* ‘jealousy’. In the second case, we modified the query grammar, in order to recover only the plural forms of the noun. A third example is that of frequent phraseological expressions which include one of the members of a collocation, such as in the case of *tener en cuenta* ‘take into consideration’, *tener que* ‘have to’, *debido a* ‘due to’, *puesto que* ‘given that’. The three verbs *tener* ‘have’, *poner* ‘take’ and *deber* ‘must’ appear as collocate verbs in a great number of collocations included in DiCE, as a consequence, we paid special attention to filtering these phraseological expressions, adopting the query grammars.

In the third and final step, once the manual revision of data is concluded, frequency data concerning collocations will be included in DiCE. When accessing the list of collocations of a certain lemma, the user will be able to choose between the options of viewing collocates in an alphabetical order or ordered according to their frequency. Therefore collocates are listed in such a way that within the entry of the base, the group of collocates with the same syntactic pattern and expressing approximately the same meaning, will be ordered from higher to lower frequency, see Table 3.

Table 3: Collocations of the noun *miedo* ‘fear’ corresponding to the syntactic pattern noun + adjective, semantically grouped and ordered according to lexical frequency

MIEDO ‘FEAR’ + ADJECTIVE		
Gloss	Collocate	Frequency
‘intense’	atroz ‘atrocious’	194
	profundo ‘deep’	155
	intenso ‘intense’	147
	cerval ‘deer-like’	85
	horrible ‘horrible’	60
	espantoso ‘awful’	33
	visceral ‘visceral’	28
	...	
‘rational’	verdadero ‘true’	131
	justificado ‘justified’	6
	fundado ‘established’	3
‘irrational’	injustificado ‘unjustified’	56
	infundado ‘unestablished’	37

Finally, we also foresee the creation of a frequency list containing all the collocations included in DiCE. We believe that in order to assign a frequency value to a collocation, it is necessary to give prominence to the frequency of the base itself, given that, from the point of view of language teaching, we can say that the more frequent the base is, the more probable it is that a learner will know it, and intend to use it. As a consequence, they will have to be familiar with its collocations. If, for instance, the collocation *miedo atroz* ‘terrible fear’ occurs the same number of times in a corpus as *abatimiento profundo* ‘utter dejection’, we propose that when it comes to assigning a frequency measure to these collocations, one should reflect that the noun *miedo* ‘fear’ is considerably more frequent than *abatimiento* ‘dejection’, a fact that indicates the greater necessity of being able to use and understand the collocation *miedo atroz* on the part of the language learner.

## 5. Conclusion and future work

In this paper we have presented the methodology applied for obtaining frequency information on collocations in a dictionary oriented to language learners. As we have shown, frequency information relative to LUs constituting the bases of collocations is already available on the DiCE web interface. The frequency band (Almela et al., 2005) to which each LU has been assigned is shown in the corresponding lexical entry. As for the collocations, we are talking about a work in progress, which will conclude with the presentation of collocates in the dictionary ordered according to their frequency within the corresponding subgroup of collocations inside the entry of the base. In addition, as we have said, we also foresee the creation of a frequency list of all collocations included in DiCE, which we believe to be especially useful in the field of language teaching as well as language testing, such as testing collocational competence or the grading teaching material on collocations.

We have already mentioned that in the course of obtaining frequency information on collocations from corpus, we have encountered a series of difficulties, mainly due to the deficiencies of the linguistic resources themselves, i.e. Spanish language corpora. We have seen that CREA, which is a balanced corpus containing good quality selected texts, could be an especially useful tool. However, the lack of lemmatization renders it inadequate for searching collocations. In addition, the query interface of the corpus obligatorily prompts for applying an automatic filter when a given query returns a high number of results, hence it does not allow retrieving all concordances. When it comes to freely available corpora, we have found that those being lemmatized and containing morphological annotation, such as Corpus del Español (Davies, 2002-), do not have a sufficient size that would make them suitable for deriving frequency information on collocations. Finally, concerning *esTenTen*, a lemmatized corpus with considerable size, we have to note that, being a webcorpus, it is less reliable in terms of representativeness and the quality of the texts it contains. Furthermore, as we have seen, the lemmatization of this corpus suffers from certain deficiencies, such as the case of missing lemmas like *odio* ‘hatred’ and *malvado* ‘evil’, tagged as if they were inflected forms of the verbs *odiar* ‘hate’ and *\*malvar*. In conclusion, given the lack of corpora of sufficient size and with quality lemmatization easily accessible to the linguist, the inclusion of frequency values in a collocation dictionary, necessarily requires a considerable amount of manual labor, which makes this task especially an expensive and slow process.

## Acknowledgements

This work has been supported by the Spanish Ministry of Economy and Competitiveness (MINECO) and the FEDER Funds of the European Commission under the contract number FFI2011-30219-C02-01, and the Spanish Ministry of Education under the FPU grant AP2010-4334. We would also like to thank Miguel Ramos Naveira for his collaboration in computational tasks and the collaborators of the DiCE group, Iria Bello Viruega, Patricia Castelo, Estela Mosqueira Suárez, Catharina Müller, and Ana Orol González, for the disambiguation of samples and revision of frequency data.

## References

- Almela, R., Cantos, P., Sánchez, A., Sarmiento, R. & Almela, M. (2005). *Frecuencias del español: Diccionario y estudios léxicos y morfológicos*. Madrid: Universitat.
- Alonso Ramos, M. (2004). *Diccionario de colocaciones del español*. Available at: [www.dicesp.com](http://www.dicesp.com).
- Alonso Ramos, M. (2012). Explorando la frecuencia léxica para el Diccionario de colocaciones del español. In T. Jiménez Juliá, B. López Meirama, V. Vázquez Rozas & A. Veiga (Eds.), *Cum corte et in nova grammatica: Estudios ofrecidos a Guillermo Rojo* (pp. 19–40). Santiago de Compostela: Universidade de Santiago de Compostela.
- Alonso Ramos, M., Nishikawa, A. & Vincze, O. (2010). DiCE in the web: An online Spanish collocation dictionary. In S. Granger & M. Paquot (Eds.), *eLexicography in the 21st century: New challenges, new applications. Proceedings of eLex 2009* (pp. 367–368). Louvain-la-Neuve: Presses Universitaires de Louvain.
- Capel, A. (2010). A1–B2 vocabulary: insights and issues arising from the English Profile Wordlists project. *English Profile Journal*, 1(01), e3.
- Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 vocabulary. *English Profile Journal*, 3, e1.
- Casso, J. (2010). *Análisis y revisión crítica de los materiales de evaluación de la competencia léxica. Elaboración de un test de vocabulario de nivel umbral*. Master's project. Madrid: Universidad Antonio de Nebrija.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- CREA = Real Academia Española. *Corpus de referencia del español actual*. Available at: <http://www.rae.es>.
- Davies, M. (2002-). *Corpus del Español: 100 million words, 1200s–1900s*. Available at: <http://www.corpusdelespanol.org>.

- Davies, M. (2006). *A frequency dictionary of Spanish: Core vocabulary for learners*. New York: Routledge.
- García Hoz, V. (1953). *Vocabulario común, vocabulario usual y vocabulario fundamental*. Madrid: SCIC.
- Hindmarsh, R. (1980). *Cambridge English lexicon a graded word list for material writers and course designers*. Cambridge: Cambridge University Press.
- Kilgarriff, A., Charalabopoulou, F., Gavrilidou, M., Johannessen, J. B., Khalil, S., Kokkinakis, S. J., Lew, R., Sharoff, S., Vadlapudi, R. & Volodina, E. (to appear). Corpus-based vocabulary lists for language learners for nine languages. *Journal of Language Resources and Evaluation*.
- Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.), *Proceedings of the Eleventh EURALEX International Congress, Euralex 2004* (pp. 105–116). Lorient: Université de Bretagne-Sud.
- Mel'čuk, I., Clas, A., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.
- Shin, D., & Nation, P. (2007). Beyond single words: the most frequent collocations in spoken English. *ELT Journal*, 62(4), 339–348.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.
- Simpson-Vlach, R. & Ellis, N. (2010). An academic formula list: New methods in phraseology research. *Applied Linguistics* 31/4: 487-512.
- Thorndike, E. & Lorge, I (1944). *The teacher word's book of 30,000 words*. New York: Teachers College, Columbia University.
- Vincze, O., Mosqueira, E., & Alonso Ramos, M. (2011). An online collocation dictionary of Spanish. In I. Boguslavsky & L. Wanner (Eds.), *Proceedings of the 5th International Conference on Meaning-Text Theory* Barcelona, September 8-9, 2011 (pp. 275–286). Barcelona.
- West, M. 1953. *A general service list of English words*. London: Longman.