

# Web-Based Model for Disambiguation of Prepositional Phrase Usage\*

Sofía N. Galicia-Haro<sup>1</sup> and Alexander Gelbukh<sup>2</sup>

<sup>1</sup> Faculty of Sciences UNAM University City, Mexico City, Mexico  
sngn@ciencias.unam.mx

<sup>2</sup> Center for Computing Research, National Polytechnic Institute, Mexico  
gelbukh@cic.ipn.mx,  
www.Gelbukh.com

**Abstract.** We explore some Web-based methods to differentiate strings of words corresponding to Spanish prepositional phrases that can perform either as a regular prepositional phrase or as idiomatic prepositional phrase. The type of these Spanish prepositional phrases is preposition–nominal phrase–preposition (P–NP–P), for example: *por medio de* ‘by means of’, *a fin de* ‘in order to’, *con respecto a* ‘with respect to’. We propose an unsupervised method that verifies linguistic properties of idiomatic prepositional phrases. Results are presented with the method applied to newspaper sentences.

## 1 Introduction

There exist certain word combinations of the type preposition–nominal group–preposition (P–NP–P) that can be idiosyncratic in nature syntactically, or semantically, or both; we call them  $EXP_{PNP}$ . Automatic determination of such  $EXP_{PNP}$  groups can help in different tasks of natural language processing.

Spanish has a great number of prepositional phrases of the type P–NP–P more or less fixed. Among them: *a fin de* (in order to), *al lado de* (next to), *en la casa de* (in the house of), etc. The  $EXP_{PNP}$  (*a fin de*, *al lado de*) define three or more simple forms (since the nominal group can contain more of a simple form) as one lexical unit. Specifically, such combinations are frequently equivalent to prepositions, i.e., they can be considered as one multiword preposition: e.g., in order to is equivalent to ‘for’ (or ‘to’) and has no relation with order.

In opposition, regular P–NP–P ( $REG_{PNP}$ ) are analyzed considering the initial combination P–NP like a unit, and the second preposition as a one introducing a complement, not always linked to the preceding NP. Therefore, it is necessary to distinguish which of the Spanish P–NP–P should be analyzed as an  $EXP_{PNP}$  and which should be analyzed as a  $REG_{PNP}$ .

The PNP-EXP groups are used frequently in everyday language, therefore natural language applications need to be able to identify and treat them properly. Apart from syntactic analysis the range of applications where it is necessary to consider their

---

\* Work partially supported by Mexican Government (CONACyT, SNI, COFAA, SIP-IPN).

specific non compositional semantic is wide: machine translation, question answering, summarization, generation, etc.

There is no a complete compilation of the EXP<sub>PNP</sub> groups. [10] compiled the widest list but he himself considers that a prepositive relation study is something incomplete “susceptible of continuous increase”. In addition, the main Spanish dictionaries [6], [11] do not contain the information necessary for a computational resource. Even if we could compile an exhaustive EXP<sub>PNP</sub> list, they present different uses and their meaning agree with context.

In this work, we mainly investigated web-based methods to disambiguate the use of prepositional phrases as fixed forms from the literal ones, for example:

1. Idiomatic expression: *a fin de obtener un ascenso* ‘to obtain a promotion’ (literally ‘at end of’),
2. Part of a larger idiom: *a fin de cuentas* ‘finally’ (literally ‘at end of accounts’),
3. Free combination: *a fin de año obtendrá un ascenso* ‘at the end of the year she will be promoted’.

Identifying non-compositional or idiomatic multi-word expressions is an important task for any computational system, in recent years attention has been paid to practical methods for solving this problem [5], [1], and specially for prepositional phrases [13]. In this work, we analyze those linguistic properties of EXP<sub>PNP</sub> that differentiate them from REG<sub>PNP</sub> and idioms. For each property, we propose a web-based method that finally could contribute with a specific measure value in order to decide the nature of the P–NP–P phrase according to such linguistic properties.

In Section 2 we present the linguistic characteristics of the EXP<sub>PNP</sub>. In Section 3 we present the method by which we detect P–NP–P phrases and their context. Section 4 describes the association measures that we considered to disambiguate the three uses of P–NP–P. In Section 5, we present the obtained results.

## 2 Some Linguistic Characteristics of the EXP<sub>PNP</sub>

In this section we discuss some linguistics properties that place EXP<sub>PNP</sub> in a different group from the regular prepositional phrases and from idioms. In Spanish grammar, EXP<sub>PNP</sub> groups are denominated adverbial locutions [12] or prepositional locutions [10] according to their function. [12] indicates that locutions could be recognized by its rigid form that does not accept modifications and the noun that shows a special meaning; or by its global meaning, that is not the sum of the meanings of its components. For word combinations lexically determined that constitute particular syntactic structures, [9] indicate their properties: restricted modification, non composition of individual senses and the fact that nouns are not substitutable.

In [8] we present the linguistic properties of idiomatic P–NP–P Spanish phrases, we take in account the more general characteristics of them that are the following:

1. **Restricted modification.** Many of the nouns found in EXP<sub>PNP</sub> groups cannot be modified by an adjective. For example: *por temor a* (to avoid, literally: by fear of) vs. *por gran temor a* (by great fear of). In some cases, the modification forces

to take a literal sense of the prepositional phrase, for example in the following sentences:

- ... *por el gran temor a su estruendosa magia* (by the great fear to its uproarious magic)
  - ... *denegó hoy la libertad bajo fianza por temor a una posible fuga*. (today denied the freedom on bail to avoid a possible flight)
2. **Non substitutable nouns.** The noun inside the EXP<sub>PNP</sub> cannot be replaced by a synonym. For example in the phrase: *se tomará la decisión de si está a tiempo de comenzar la rehabilitación* (the decision will be taken on if it is the right time to begin the rehabilitation), where *a tiempo de* cannot be replaced by *a período de* (on period of), *a época de* (time of).
  3. **Part of idioms.** Some EXP<sub>PNP</sub> groups initiate fixed phrases or can be literal phrases according to the context, in addition to their use as idiomatic expression.

Examples:

“*al pie de*” (lit. ‘to the foot of’) It appears as idiomatic expression in:

*La multitud que esperó paciente al pie de la ladera de la sede de la administración del canal, corrió hacia arriba ...* (The patient multitude that waited at the base of the slope of the seat of the channel administration, run upwards ...)

“*al pie de*” It initiates a larger idiom: *al pie de la letra* (exactly) in:

*Nuestro sistema de procuración de justicia se ha transformado y en vez de observar al pie de la letra las leyes ...* (Our system of justice care has been transformed and instead of observing the laws exactly ...)

“*al pie de*” It initiates a free combination in:

*El anillo estaba junto al pie de María* (The ring was next to the foot of Maria)

### 3 Structure of the Prepositional Phrases and Their Context

We wrote a Perl program to automatically determine each P–NP–P and its right and left contexts. We used the AGME<sup>1</sup> tool to define the POS of each word and a very wide list of prepositions (P) obtained from [10] that includes prepositions with liberality.

The grammar defining the structure of the P–NP–P consists of the following rules:

PP → P NP P

NP → N | D N | V-Inf | D V-Inf

where PP stands for prepositional phrase, N for noun, D for determinant, and V-inf for infinitive verb (in Spanish, infinitives can be modified by a determinant: *el fumar está prohibido*, literally ‘the to-smoke is prohibited’).

<sup>1</sup> <http://www.cic.ipn.mx/~sidorov/agme/>

For example, in the following sentence the program detects six P–NP–P phrases (*de residencia a, a los inmigrantes para, a la falta de, de mano de, de obra en, según cálculos del*):

*El presidente cree necesario que el Gobierno agilice los permisos de residencia a los inmigrantes para poder responder a la falta de mano de obra en el país que, según cálculos del Departamento de Trabajo, impide cubrir unos 23,000 empleos en diversos sectores.*

To determine the right and left context of each P–NP–P for the analysis, we considered the following characteristics to delimit them, in the following order: (1) Punctuation, (2) Conjunctions, (3) Articles, adjectives and noun phrases.

For the same paragraph, the P–NP–P phrases and their reduced context were: *permisos de residencia a los inmigrantes, residencia a los inmigrantes para poder esponder, responder a la falta de mano, falta de mano de obra, mano de obra en el país, según cálculos del Departamento*. For the phrase (*mano de obra en el país*) the first cut was due to punctuation (,) the second cut was due to a conjunction (*que*) for the right context. For the left context, the article was the cut to leave a noun.

## 4 Web-Based Testing of Linguistic Properties

In this section we analyze the linguistic properties that could help to differentiate  $EXP_{PNP}$  groups from the  $REG_{PNP}$  and from idioms. For each property we try to find examples in Internet to decide if such property is accomplished.

We used the Internet as corpora. Internet was accessed using the Google search engine limited to Spanish language. Google statistics of co-occurrences of any two strings with any  $N$  intermediate words can be gathered by queries in quotation marks containing these strings separated with an asterisk, e.g., “*pie de \* página*”. The search engine returns phrases as: “*pie de cada página*”, “*pié de cada página*”, “*pie de lucha · Página*”, “*Pie de Página de la página*”, etc.

### 4.1 Restricted Modification

Since the nouns in  $EXP_{PNP}$  groups cannot be modified by an adjective we searched for possible noun phrase modifications in each  $P_1$ –NP– $P_2$  with its specific context, i.e., we searched for phrases of the type “context  $P_1$  \* NP \*  $P_2$  context”.

For example, if we have the  $EXP_{PNP}$  phrase “*con motivo de*” (that can be replaced by another single preposition: *por* ‘by’) and we want to find examples of restricted modification of a complete sentence containing it, we need to consider its specific context:

*Está bien que con motivo de fin de año todos tomen vacaciones, pero ¿todos a la vez?* (It is well that in occasion of end of year all people take vacations, but all at the same time?)

We did not find examples for the complete example in the Internet. Using asterisk in Google search and a reduced context we obtained 36 pages in the Spanish language search of “*con \* motivo \* de fin de año*”, but only three different types:

1. The prepositional phrase appears without modification (NOTHING)
2. The prepositional phrase is broken and parts of it correspond to different parts of a long phrase (PHRASE)
3. The noun of the prepositional phrase is modified (MODIF)

In order to obtain more examples we reduced the context without changing the prepositional phrase meaning. We limited the phrase to “*con \* motivo \* de fin*”, and we obtained 55 examples. In the following table we present the three different observed types, an example for each type and the total number of obtained items:

NOTHING	... que se celebran en fechas connotadas, generalmente relacionadas <b>con celebraciones históricas o con motivo de las fiestas de fin</b> de año. ... ‘that are celebrated on the corresponding dates, usually related <b>with historical celebrations or on the occasion of the holidays of end</b> of year ...’	34
MODIF	<b>Con el motivo adicional de fin</b> de año, esperamos desde ya su asistencia y participación Jorge Castro, Jorge Raventos, Pascual Albanese 27/11/2006 ... ‘On the additional occasion of the end of year, we expect from right now his presence and participation Jorge Castro, Jorge Raventos , Pascual Albanese 27/11/2006 ...’	4
PHRASE	... Orizaba, Tampico, etcétera, etcétera, han estado cambiando, <b>con este último motivo, telegramas para ponerse de acuerdo a fin</b> de llevar una manifestación ... ‘Orizaba, Tampico, etc., etc. has been changing, on the occasion of the latter, telegrams to agree in order to organize a manifestation ...’	17

We observed that very few snippets for restricted noun modification were obtained for the example when the prepositional phrase corresponded to a  $EXP_{PNP}$  phrase and we expect that much more snippets should be obtained when the prepositional phrase corresponds to a  $REG_{PNP}$ .

Considering the regular phrase *motivo de fin de año* ‘of end of year’, we made a Google search of the string “*de \* fin de año*”. The search gave 651,000 examples. We intended to determine frequencies of the different types of examples obtained. Because of the volume, we cannot determine them thoroughly. To evaluate the true portion of the different types of examples in the automatically counted amounts, we looked through the first hundred snippets and manually analyzing their syntax to prove the type.

In 100 snippets, we manually found three different types: eight snippets for MODIF, 89 for PHRASE and 3 snippets for NOTHING. We expected much more cases of MODIF type, but we found less cases of the prepositional phrase as it. It seems that the relations are inverted between MODIF and NOTHING.

## 4.2 Non-substitutable Nouns

Considering that the noun phrase inside the  $EXP_{PNP}$  cannot be replaced by a synonym, we tried to find how noun phrase could be substitutable inserting its synonyms from a dictionary and searching the modified phrase in the Internet.

The initial phrase “*que con motivo de fin de año*” obtained 220 pages in Google search. We kept the same context and for the phrase “*que con motivo de SINONYM*”

*de año*” we did not obtain results for any synonym. However, reducing the context we obtained:

---

“ <i>con motivo de SINONYM de año</i> ”:	<i>final</i> – 2 pages, <i>terminación</i> – 3 pages
“ <i>motivo de SINONYM de año</i> ”:	<i>final</i> – 2 pages, <i>terminación</i> – 3 pages
“ <i>de SINONYM de año</i> ”:	<i>final</i> – 157000 pages, <i>cabo</i> – 121 pages <i>remate</i> – 12 pages, <i>terminación</i> – 4 pages
	<i>propósito</i> – 132 pages, <i>meta</i> – 4 pages

---

We could observe that for  $REG_{PNP}$  we found many combinations and we did not expect to obtain results for the  $EXP_{PNP}$ .

Considering the same phrase, we searched “*que con SYNONYM de fin de año*” and we did not find results using the synonyms of *motivo* (*causa, razón, pretexto*, etc.). We reduced the context and we only found one page for “*con pretexto de fin de año*” and two pages for “*con pretexto de fin*”.

### 4.3 Part of Idioms

As it is reported in [4], idiomatic expressions combine with a more restricted number of neighboring words than free combinations. They computed three indices on the basis of three-fold hypothesis: a) idiomatic expressions should have few neighbors, b) idiomatic expressions should demonstrate low semantic proximity between the words composing them, and c) idiomatic expressions should demonstrate low semantic proximity between the expression and the preceding and subsequent segments

We consider the first hypothesis since the others hypothesis imply semantic measures. For example, the P–NP–P *a fin de* has more neighbors than the phrase *a fin de cuentas*. Searching for “*a fin de \**” we obtained 1,340,000 pages, manually we observed that a noun was its immediate neighbor and that the noun was linked to the P–NP–P in most of the cases. We obtained 1,220,000 pages for *a fin de cuentas* and manually we observed that the immediate neighbor was not linked to the phrase in most of the cases.

## 5 Association Measures

Diverse statistical measures have been used to identify lexical associations between words from corpus [3, 7]. To formally define and to numerically test the relation between words we did not take the criterion based only on frequencies, since these frequencies depend on the corpus size.

The results that we present in the following subsections were obtained with a collection of examples obtained from 1000 sentences that have one  $EXP_{PNP}$  according to [10]. For each sentence, we extracted each  $P_1$ –NP– $P_2$  with its context.

## 5.1 Noun Modification

For each  $P_1$ -NP- $P_2$  we got the  $P_1 * NP * P_2 + context$  search where asterisk could represent punctuation signs, one word or several words. We wrote a PERL program to determine the type of the obtained string, instead of the asterisk, by means of POS and very simple rules. We classified such type as: ‘modif’, ‘phrase’, ‘nothing’ and ‘doubt’. Doubt means that our POS tool and the heuristics could not assign a type to that string. Considering 100 snippets, we automatically calculate the true portion of each type.

We obtained the results for five hundred examples considering only the MODIF and PHRASE types. We could observe that several of them had the same values since the delimitation of context made equal phrases. Nevertheless, we could observe that a delimitation could be defined for MODIF from zero to 0.22% value.

We did not obtain clearly differentiated values for MODIF and PHRASE types in one thousand examples of  $REG_{PNP}$ .

## 5.2 Substitutable Nouns

The measure that we want to obtain to determine if it is possible to substitute the noun inside the  $P_1$ -NP- $P_2$  is how many combinations exist, i.e., how many of the synonyms are combinable with  $P_1$  and  $P_2 + context$ .

We wrote a program that use a Spanish Synonym Dictionary to substitute each one of their synonyms instead of the NP, then it searches the Internet and it outputs the number of synonyms that obtained a result. Some examples are listed in the following table where we show: the original phrase with the  $P_1$ -NP- $P_2$  between hyphens, an example of the reduced context, the number of synonyms that obtained a result and the number of pages.

Original phrase	Example	#SYN	Pages
propuesta - de acuerdo con - sus intereses	de compromiso con sus intereses	1	4
fin - de llegar a - una posible reglamentación al respecto	de aproximarse a una posible	2	17
calidad - de vida de - sus habitantes	de ocupación de sus habitantes	11	159
fin - de trabajar de - manera coordinada y enfrentar cualquier acto de	fin de luchar de manera	46	1685

The values obtained for the prepositional phrases that obtained results substituting the noun by a synonym are shown in Figure 1. We eliminated such sentences that had a NP without an entry in the Synonym Dictionary, phrases that obtain zero results in the Web search and phrases without context. We could observe that  $REG_{PNP}$  have values between 1 and 25. We could observe that  $EXP_{PNP}$  have less number of combinations. We can neglect the peaks since they correspond to a P-NP-P that is part of an idiom for example: “*a fin de cuentas*” and “*a fin de que*”. We could determine that  $REG_{PNP}$  have differentiated values between 4 and 25 combinations, where  $EXP_{PNP}$  have near zero examples.

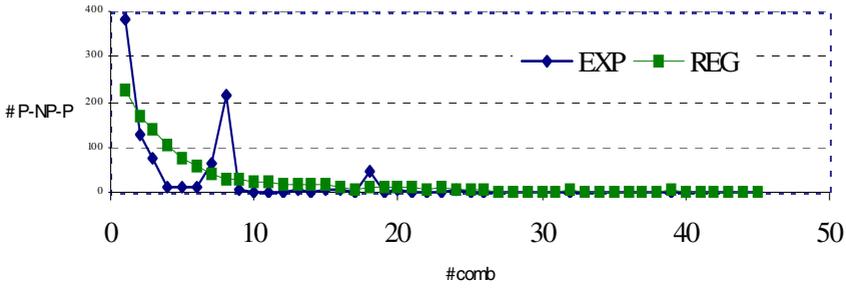


Fig. 1. Number of combinations obtained for the substitution of NP by its synonyms

### 5.3 Lexical Association for Idioms

We use similar criteria of Mutual Information, the Stable Connection Index [2] in the shape

$$SCI(P_1, P_2) \equiv 16 + \log_2 \frac{N(P_1, P_2)}{\sqrt{N(P_1) \times N(P_2)}}$$

where the constant 16 and the logarithmic base 2 are chosen in a way that the majority of the results are in the interval 0 to 16. To calculate *SCI*, we do not need to know the steadily increasing total volume under the search engine’s control. *SCI* reaches its maximally possible value 16 when  $P_1$  and  $P_2$  always go together. *SCI* retains its value when  $N(P_1)$ ,  $N(P_2)$ , and  $N(P_1, P_2)$  change proportionally. This is important since all measured values fluctuate quasi-synchronously in time.

We obtained the *SCI* values for pairs of word groups using the Google search engine considering that  $P_1$  is P–NP–P and  $P_2$  is the right context to evaluate the possibility of P–NP–P being the initial part of an idiom.

For example, we took 219 sentences of newspaper texts where the P–NP–P *al pie de* occurs. We extract each P–NP–P with its context and their *SCI* value was obtained. The range for *SCI* values goes from –2.84 to +13.13 Thirteen phrases have negative values; 35 phrases have zero *SCI* value since the string of words  $P_1 P_2$  got zero occurrences, 9 phrases have *SCI* values between 12.11 and 14.33 All the rest have *SCI* values between 0.16 and 10.8

We show some examples in the following table:

Phrase	#Pages	Context	# Pages	SCI
al pie de la sede	31	la sede	9900000	-1.44
al pie de la Suburban	0	la Suburban	940	0
al pie de un cactus	9	un cactus	51000	0.57
al pie de una cruz	205	una cruz	964000	2.96
al pie de unas colinas	90	unas colinas	9930	5.07
al pie de la torre	13500	la torre	7210000	7.55
al pie del cerro	52600	cerro	6460000	9.59
al pie de la montaña	49500	la montaña	4480000	10.06
al pie del cañón	140000	cañón	2290000	12.98
al pie de la letra	767000	la letra	10200000	14.33

The group with values greater than 12.98 correspond to idioms, for example: *al pie del cañón*, *al pie de la letra*. Other five cases showed that idioms obtained SCI values greater than 12.5

We also obtained the Google statistics for the phrases: NP and P<sub>2</sub> + context, to prove the combinability of the noun with the second preposition and the right context. It seems that a high value of SCI<sub>NP-P2+context</sub> and much higher than SCI<sub>P-NP-P+context</sub> correspond to a REG<sub>PNP</sub>

#### 5.4 Heuristics for Use Disambiguation of Prepositional Phrases

The heuristics that we deduce from the previous results are:

1. If the number of synonyms is greater or equal to four and lower than 25, it is probable that the P–NP–P is a REG<sub>PNP</sub>.
2. If modifiers detected like adjectives exist and their values are greater than zero and smaller than 0.2 it is probable that the P–NP–P is an EXP<sub>PNP</sub>.
3. If the SCI values for the complete phrase and noun both are greater than 12.5 and very closer it is probable that the P–NP–P is an idiom.
4. If the noun SCI<sub>NP-P2+context</sub> is greater than the SCI of the complete phrase and the values have a difference greater than 3.0 then it is probable that the P–NP–P is a REG<sub>PNP</sub>.

## 6 Results

We extracted sentences of a newspaper corresponding to 22/12/04. For each sentence we obtained the prepositional phrases P–NP–P + context. For each one we obtained the SCI values for the complete group and for the noun and P<sub>2</sub> + context. We also obtained the modifier types (PHRASE, MODIF, NOTHING, DOUBT) of the noun with their True Portion and the statistics of synonymous substitution from Google.

After applying the heuristics described in the previous section, we obtained the following results:

Type	Precision	Recall	# detected	# correct	#correct detected
EXP <sub>PNP</sub>	46	-	39	18	18
IDIOM	100	80	4	5	4
REG <sub>PNP</sub>	99	82	99	120	98

where:

Precision is the number of correct P–NP–P detected / # of P–NP–P detected,

Recall is the number of correct P–NP–P detected / # of P–NP–P manually labeled.

The results show the values obtained only for the P–NP–P that got at least one value, since our Synonym dictionary have few thousands of entries, some contexts were not reduced and the Google search engine did not find hits for them.

We found that among the 21 REG<sub>PNP</sub> bad recognized as EXP<sub>PNP</sub> eight of them correspond to cases where the context coincides exactly with a named entity. Other eight

cases correspond to P–NP–P that got two opposite values, for example, the test for noun modification detects an EXP<sub>PNP</sub> and the test for non-substitutable noun detects a REG<sub>PNP</sub>. When the context is a named entity the previous noun is highly tied to the name, for example “*el penal de La Loma*” (La Loma prison) so few modifiers and synonyms are used. If we include a heuristic regarding named entities, precision for EXP<sub>PNP</sub> increases to 56%. For REG<sub>PNP</sub> there was only one case where the P–NP–P got two opposite values. When only two opposite values are obtained the heuristics assign first the value for the noun modification test.

Based on these results, future work will consider to apply a classification method instead of heuristics but more examples will be needed for training them. The very simple rules to determine the type of the retrieved strings of noun modification should be changed to improve the “*modif*” and “*phrase*” recognition.

In addition, a refinement is necessary in the delimitation of context, since one word could be the difference to obtain results from Google. Having results from the search engine the next improvement should deal with inclusion of a bigger quantity of snippets to have a more accurate determination of modifier types.

## 7 Conclusions

Idiomatic word combinations of EXP<sub>PNP</sub> type, usually functioning as compound prepositions, have linguistic properties distinct from those regular. In particular, they combine with a greater number of words than usual idioms. For their unsupervised determination from Internet, we proposed to obtain their combinability through Google searching. We presented an approach to differentiate them from fixed phrases and free combinations by means of some association measures according to linguistic properties.

We used as association measures the combinations with synonyms of the noun to prove non-substitutable nouns, the noun modifications by means of adjectives to prove restricted modification and a like-mutual information to deduce the strength of links between the prepositional phrase and the context, and between the noun and the right prepositional phrase. The values obtained for such measures in a test collection were adapted to some heuristics.

The preliminary results obtained from the application of heuristics give some directions to improve their adequacy to determine the degree to which the prepositional phrases are compositional.

## References

1. Baldwin, T., Bannard, C., Tanaka, T., Widdows, D.: An Empirical Model of Multiword Expression Decomposability. In: Dignum, F.P.M. (ed.) ACL 2003. LNCS (LNAI), vol. 2922, pp. 89–96. Springer, Heidelberg (2004)
2. Bolshakov, I.A., Galicia-Haro, S.N.: Detection and Correction of Malapropisms in Spanish by means of Internet Search. In: Matoušek, V., Mautner, P., Pavelka, T. (eds.) TSD 2005. LNCS (LNAI), vol. 3658, pp. 115–122. Springer, Heidelberg (2005)

3. Church, K.W., Hanks, P.: Word association norms, mutual information, and lexicography. In: Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics, pp. 76–83 (1989)
4. Degand, L., Yves, B.: Towards automatic retrieval of idioms in French newspaper corpora. *Literary and Linguistic Computing* 18(3), 249–259 (2003)
5. Dekang, L.: Automatic identification of noncompositional phrases. In: Proc. of the 37th Annual Meeting of the ACL, College Park, USA, pp. 317–324 (1999)
6. de María Moliner, D.: *Diccionario de Uso del Español*. Primera edición versión electrónica (CD-ROM) Editorial Gredos, S. A. (1996)
7. Evert, S., Krenn, B.: Methods for the Qualitative Evaluation of Lexical Association. In: Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics. Toulouse, France, pp. 188–195 (2001)
8. Galicia-Haro, S.N., Gelbukh, A.: Towards the Automatic Learning of Idiomatic Prepositional Phrases. In: Gelbukh, A., de Albornoz, Á., Terashima-Marín, H. (eds.) *MICAI 2005*. LNCS (LNAI), vol. 3789, pp. 780–789. Springer, Heidelberg (2005)
9. Manning, C.D., Schütze, H.: *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge (1999)
10. Nañez Fernández, E.: *Diccionario de construcciones sintácticas del español*. Preposiciones. Madrid, España, Editorial de la Universidad Autónoma de Madrid (1995)
11. Real Academia Española.: *Diccionario de la Real Academia Española*, 21 edición (CD-ROM), Espasa, Calpe (1995)
12. Seco, M.: *Gramática esencial del español, introducción al estudio de la lengua*, Segunda edición revisada y aumentada, Madrid, Espasa Calpe (1989)
13. Villada, B., Tiedemann, J.: Identifying idiomatic expressions using automatic word-alignment. In: Proceedings of the EACL 2006 Workshop on Multiword Expressions in a Multilingual Context, Trento, Italy (2006)