

CALLEX-ESP: a software system for learning Spanish lexicon and collocations

Boguslavsky, I.¹, Barrios Rodríguez M.², Diachenko, P.³

1 Universidad Politécnica de Madrid, Campus de Boadilla del Monte, Madrid, Spain.

2 Universidad Complutense de Madrid, Avda. Del Paraninfo s/n, Madrid, Spain.

3 Institute for Information Transmission Problems RAS, 19, B. Karetnyj, GSP-4, Moscow, Russia.

The paper introduces *CALLEX-ESP*, an adaptation for Spanish of the *CALLEX* system developed for Russian and English at the Institute for Information Transmission Problems of the Russian Academy of Sciences. The system is intended for learning Spanish as a foreign language, mostly in the aspect of the lexicon and collocations. We show some of the exercises developed within *CALLEX-ESP*.

Keywords computer-assisted language learning, educational software, lexicon, lexical functions, collocations

1. Introduction

Below we will summarize the methodology adopted for the development of *CALLEX-ESP*. This system is an adaptation for Spanish of a software system, *CALLEX*, developed at the Institute for Information Transmission Problems of the Russian Academy of Sciences. So far, *CALLEX* has been developed for Russian, English and, to a lesser extent, for German (in cooperation with German and Austrian partners within the INTAS program). The system is intended for both foreign language learners and those native speakers who wish to improve and enrich their mother tongue mastery. The dictionary of *CALLEX* counts more than 3,000 lexical entries for Russian and more than 1,500 entries for English. They are studied in various aspects, both syntagmatic (synonymy, antonymy, hyperonymy, etc.) and paradigmatic (collocations). *CALLEX* allows the learners to extend their lexical knowledge by means of five types of games. More details on *CALLEX* are given in the paper by P. Diachenko in the present volume. Another product of the Russian group – the ETAP-3 machine translation system – also makes heavy use of LFs for disambiguation and finding idiomatic translations (Apresjan et al., 2003a, 2003b).

The system is based on the theoretical foundations of the Meaning – Text Theory (MTT) by I. Mel'čuk and the Theory of Systemic Lexicography by Ju. Apresjan. The current version of *CALLEX-ESP* concentrates on lexical knowledge formalized by means of lexical functions (LF) (see below, section 2). Some tools developed in the framework of MTT (Alonso Ramos: 2004; Polguère: 2005) have shown that lexical functions provide not only a powerful apparatus for theoretical and practical lexicography, but are also instrumental in the learning of collocations. Many of the lexical functions are used by Alonso Ramos in her *Diccionario de colocaciones del español*. The entries of this dictionary cover the lexical field of emotions and are accompanied by a large set of exercises to be done by the learner. While *CALLEX-ESP* is far from being completed, we have a right to claim that although standard lexical functions do not cover all possible lexical relations, they are sufficiently rich to be extremely useful for language learners.

The purpose of this presentation is to show some examples of the learning activities included in *CALLEX-ESP* and explain why LFs are so convenient for learning languages and for software applications dealing with natural language processing.

Section 2 gives a brief introduction into LFs and demonstrates their usefulness for learning languages. In Section 3, we show some exercises of *CALLEX-ESP* based on standard lexical functions, while in section

¹ Contact author: igor@opera.dia.fi.upm.es

4 we propose some exercises around non-standard lexical functions. Finally, some conclusions are formulated.

2. Lexical functions: a tool for the formalization of natural language

Learning a new language at an intermediate or advanced level implies more than a grammar and a lexicon: at these levels students usually have problems with collocations. Spanish collocations have been studied by Higuera García (2004) and Pejovic (2003) who present a rich variety of exercises in order to make students learn collocations by heart, especially when Spanish expressions differ from those in the student's mother tongue.

For example, if an English speaker tries to say *to have a shower* in Spanish and translate word by word, he would say *#tener una ducha*, but not *darse una ducha* (*#to give a shower to himself*;) which is the correct Spanish expression. A French student would have the same problem because of *prendre une douche* (*take a shower*), or a Russian student because of *принимать душ* (*#receive a shower*); also in the case of Italian where the literal translation of *fare la doccia* is *#hacer una ducha* (*#to do a shower*).

To account for this type of phenomena, the concept of lexical function (LF) has been introduced and developed within the Meaning – Text theory (Mel'čuk, 1995; Apresjan, 2002). More than 100 different LFs have been discovered in natural languages. Most of them are universal across languages. To come back to the example above, the general sense of the verb in the expression *to have a shower* and its equivalents in other languages is called **Oper₁**, that is an LF that allows us to translate *to have a shower* automatically into different languages, as shown in (1):

(1) **Oper₁** (ducha) = darse, **Oper₁** (shower) = to have, **Oper₁** (douche) = prendre, **Oper₁** (doccia) = fare, **Oper₁** (душ) = принимать

The **Magn** LF covers adjectival correlates of the meaning 'intense', as in the case of *heavy rain* that cannot be translated into Spanish as *#lluvia fuerte* but as *lluvia intensa* (*#intense rain*) (cf. (2)). Yet another LF, **S_i** (the name of the participant of the situation) applies to syntagmatic relations such as *school – teacher – student* (cf. (3)):

(2) **Magn** (lluvia) = intensa; **Magn** (rain) = heavy

(3) **S₁**(colegio) = alumno; **S₂**(colegio) = profesor; **S₁**(school) = student; **S₂**(school) = teacher

LFs can also appear in combinations (complex and compound LFs), thus covering most of the lexicon and its functional possibilities in any language.

The formalization of Spanish lexical relations with LFs and their combinations is useful not only for developing the educational software system itself but for understanding the meaning of the expressions involved; every LF has one or more paraphrases in natural language. For instance, **CausFunc₀** means 'to cause something to exist' and it can be paraphrased as 'to create', 'to produce', 'to make' or 'to invent'. *CALLEX-ESP* makes use of these paraphrases and offers the students the possibility of understanding their meaning and learning other predicates which are equivalent to them, even if they appear in different combinations, such as *to build a house, to create a danger, to unleash a war, to put forward an idea, to spread a rumour*, etc.

3. CALLEX-ESP exercises with standard LFs

CALLEX contains several types of lexical games. One of them confronts the user with the semantic definition of a word and asks him/her to guess what word is denoted by this definition. Another game deals with translating words (into Russian, if it is an option for English-speaking learners, and into English, if it is intended for speakers of Russian). Some of the games make use of lexical functions to train the user in relations between the words. Some of the games have a “prompt” version, in which the foreign language learner is given a pointing in his/her native language.

Each game related to lexical functions has three levels depending on the complexity of educational material. The first level is composed of the simplest functions such as **A₀** and **Adv₀**; lexical functions of medium complexity such as **A₁** and **A₂**, **Incep** and **Fin** are placed at the second level; the most complicated functions belong to the Real-Fact family, such as **Labreal** or **CausFact** and form the third level of complexity.

For each correct answer, the user is given a certain number of points. The achievement of the user is determined not only by the number of correct answers and not only by the complexity level at which he/she is playing but also by his/her ability to find alternative answers. A lexical function may belong to a low complexity level but have a large number of values for the same argument, some of them being more difficult for the learner than others. The more variants he/she can find, the more points are given. If the learner is in a difficulty, he/she may choose to ask the system for the correct answer. He will be shown not only all possible variants of the answer but also the number of points he could have won, should he give the correct answer by himself.

The Spanish version of the game, *CALLEX-ESP*, will have a slightly different form of presenting material. The games will transform into exercises which will make use of the GLOSSES of lexical functions, that is, their rendering in natural language.

For example, adjectival lexical function **AntiVer** can be roughly glossed as ‘no auténtico’ (‘not authentic’). The exercise for training this function can look like (6), and correct answers are given in (7):

(6)

cuadro (‘no auténtico’) = cuadro _____
 acusación (‘no auténtica’) = acusación _____
 gobierno (‘no auténtico’) = gobierno _____
 prueba (‘no auténtica’) = prueba _____
 moneda (‘no auténtica’) = moneda _____
 demostración (‘no auténtica’) = demostración _____
 zumo (‘no auténtico’) = zumo _____

(7) **AntiVer** = ‘no auténtico’

AntiVer(cuadro) = falsificado
AntiVer(acusación) = gratuita
AntiVer(gobierno) = ilegítimo
AntiVer(prueba) = infundada
AntiVer(moneda) = falsa
AntiVer(demostración) = inconsistente
AntiVer(zumo) = artificial

CALLEX-ESP allows the user to expand his/her vocabulary. Some exercises, for example, recur to incomplete collocations. The user sees a keyword followed by a paraphrase or a gloss of a LF. His task is

to find a lexical unit that co-occurs with the keyword and has the meaning corresponding to the gloss; cf. (4):

(4) Imaginación ('grande') = imaginación _____

The paraphrase of the lexical function in single quotes serves to give the students a guiding line for finding collocates. Usually, a native speaker is not able to come up with all possible variants. Only some of them usually come to mind, for example, *imaginación calenturienta, fértil* or *poderosa*. After trying to recollect as many collocates as he is able to, the learner is allowed to have a look at the full set in (5). This exercise shows that both for language learners and for native speakers, collocations turn to be much easier to understand passively than to use actively.

(5) **Magn** = 'grande'

Magn(imaginación) = viva, fértil, fecunda, prolífica, caudalosa, rica, desbordante, ardiente, encendida, exuberante, contagiosa, portentosa, poderosa, potente, febril, desaforada, desenfrenada, exagerada.

4. CALLEX-ESP exercises with compound lexical functions

Many lexical functions can combine with each other to form compound LFs. Paraphrases of compound LFs can often be obtained as combinations of paraphrases of their components. For example, if **Func**₀ is paraphrased as 'existir' ('exist'), **Incep** as 'empezar' ('begin'), **Cont** as 'continuar' ('continue'), **Fin** as 'terminar' ('finish'), then the paraphrase of the compound LF **IncepFunc**₀ would be 'empezar a existir' ('begin to exist'), the paraphrase of **ContFunc**₀ would be 'seguir existiendo' ('continue to exist') and the paraphrase of **FinFunc**₀ - 'finish to exist'. This type of paraphrases underlies exercises such as (9):

(9)

'empezó a existir' una flor 'a flower began to exist' = _____ *una flor*
 'empezó a existir' una duda 'a doubt began to exist' = _____ *una duda*
 'empezó a existir' una tormenta 'a storm began to exist' = _____ *una tormenta*
 'empezó a existir' el silencio 'the silence began to exist' = _____ *el silencio*
 'empezó a existir' la guerra 'a war began to exist' = _____ *la guerra*
 'siguen existiendo' las precipitaciones 'rainfall continues to exist' = _____ *las precipitaciones*
 'sigue existiendo' la sensación 'sensation continues to exist' = _____ *la sensación*
 'dejó de existir' el viento 'the wind finished to exist' = _____ *el viento*
 'dejó de existir' el temporal 'the storm finished to exist' = _____ *el temporal*
 'dejó de existir' la niebla 'the mist finished to exist' = _____ *la niebla*
 'dejó de existir' la multitud 'the crowd finished to exist' = _____ *la multitud*
 'dejó de existir' la fiebre 'the fever finished to exist' = _____ *la fiebre*

The right answers to this exercise are given in (10):

(10)

IncepFunc₀ = 'empezar a existir'

ContFunc₀ = 'seguir existiendo'

FinFunc₀ = 'dejar de existir'

IncepFunc₀(flor 'flower') = brotar 'sprout'

Brotó una flor

IncepFunc₀(duda 'doubt') = surgir 'arise'

Surgió una duda

IncepFunc₀(tormenta 'storm') = desatarse 'burst'

Se ha desatado una tormenta

IncepFunc₀(silencio 'silence') = hacerse 'ensue'

Se hizo el silencio

IncepFunc₀(guerra 'war') = estallar 'break out'

Estalló la guerra

ContFunc₀(precipitación) = persistir 'persist'

Persisten las precipitaciones

ContFunc₀(sensación 'feeling') = permanecer 'remain'

Permanece la sensación

FinFunc ₀ (viento ‘wind’) = amainar ‘tail away’	<i>Amainó el viento</i>
FinFunc ₀ (temporal) = remitir ‘subside’	<i>Remitió el temporal</i>
FinFunc ₀ (niebla ‘mist’) = disiparse ‘lift, clear away’	<i>Se disipó la niebla</i>
FinFunc ₀ (multitud ‘crowd’) = dispersarse ‘disperse’	<i>Se dispersó la multitud</i>
FinFunc ₀ (fiebre ‘fever’) = remitir ‘decline’	<i>Remitió la fiebre</i>

5. Conclusions

The state-of-the-art of the study of lexical functions permits to corroborate the hypothesis brought forward by some authors that lexical functions are useful for language learning. In case of Spanish, rendering of LF symbols by means of natural language glosses (paraphrases) facilitates the comprehension of the LF meaning. We showed examples of exercises intended to train adjectival and compound verbal LFs.

References

- Alonso Ramos, M. (2005). “Semantic description of collocations in a Lexical Database” en Kiefer F. et al. (eds.), *Papers in Computational Lexicography COMPLEX 2005* Budapest,): Linguistics Institute and Hungarian Academy of Sciences, pp. 17-27.
- Alonso Ramos, M. (2004). *DiCE, Diccionario de Colocaciones del Español*. Universidad de la Coruña. [Diccionario disponible en: <http://www.dicesp.com>]
- Apresjan et al. (2002) Ju. D. Apresjan, I. M. Boguslavsky, L. L. Iomdin, L. L. Tsinman, Lexical Functions in NLP: Possible Uses, Computational Linguistics for the New Millennium: Divergence or Synergy? Proceedings of the International Symposium held at the Ruprecht-Karls-Universität Heidelberg, 21-22 July 2000, Manfred Klenner / Henriëtte Visser (eds.), Frankfurt am Main, 2002, pp. 55-72.
- Apresjan et al. (2003a). Ju. D. Apresjan, I. M. Boguslavsky, L. L. Iomdin, A. V. Lazurskij, L. G. Mitjushin, V. Z. Sannikov, L. L. Cinman . ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the Meaning ⇔ Text Theory. *Proceedings of the 1st International Conference on Meaning-Text Theory*. Paris, Ecole Normale Supérieure, June 16–18; 279-288, 2003 [Traductor automático ruso-inglés disponible: <http://proling.iitp.ru/etap/index.html>]
- Apresjan et al. (2003b). Ju. Apresjan, I. Boguslavsky, L. Iomdin, L. Tsinman. Lexical Functions as a Tool of ETAP-3. *Proceedings of the 1st International Conference on Meaning-Text Theory*. Paris, Ecole Normale Supérieure, June 16–18; 2003.
- Bosque, I. (2006) *Diccionario práctico combinatorio del español contemporáneo*, Madrid: ed. SM.
- Bosque, I. (2004). *Redes. Diccionario combinatorio del español contemporáneo*, Madrid: ed. SM.
- Higueras García, M. (2006). *Las colocaciones y su enseñanza en la clase de ELE*. Madrid: Arco/Libros.
- Higueras García, M. (2004). *La enseñanza-aprendizaje de las colocaciones en el desarrollo de la competencia léxica en el español como lengua extranjera*. Madrid: Universidad Complutense de Madrid. [Tesis inédita]
- Mel’čuk, I. (2003). “Collocations dans le dictionnaire”. *Les écarts culturels dans les dictionnaires bilingues*. Ed. T. Szende. París: Honoré Champion. 19-64.
- Mel’čuk, I. (1998). “Collocations and Lexical Functions”. *Phraseology. Theory, analysis and applications*. Ed. A. P. Cowie. Oxford: Clarendon Press. 23-53.
- Mel’čuk, I., Clas., A. y Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Duculot : Lovaina la Nueva.
- Pejovic, Anđelka. (2003). Las colocaciones con verbo como elemento nuclear en español contemporáneo: aspectos descriptivos aplicados. Granada: Universidad de Granada [Tesis inédita]
- Polguère, A. (2005). *DiCouébe, Dictionnaire en ligne de combinatoire du français*. Universidad de Montreal. [Diccionario disponible en: <http://olst.ling.umontreal.ca/dicouebe>]
- Polguère, A. (2000). “Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French”. *Proceedings of EURALEX 2000*. 517-527.