

ANGLOGERMANICA ONLINE 2006. Períñán-Pascual, Carlos and Arcas-Túnez, Francisco:
Reusing computer-oriented lexica as foreign-language electronic dictionaries

Reusing computer-oriented lexica as foreign-language electronic dictionaries¹

Carlos Períñán-Pascual, Universidad Católica San Antonio (Spain)

Francisco Arcas-Túnez, Universidad Católica San Antonio (Spain)

Index

[1 Introduction](#)

[2 FunGramKB](#)

[2.1 Overview](#)

[2.2 The lexical entry](#)

[2.2.1 Part of speech](#)

[2.2.2 Number](#)

[2.2.3 Gender](#)

[2.2.4 Degree](#)

[2.2.5 Adjectival position](#)

[2.2.6 Verb paradigm](#)

[2.2.7 Pronominalization](#)

[2.2.8 Voice and tense constraints](#)

[2.2.9 Countability](#)

[2.2.10 Predicate frame](#)

[3 Benefits of FunGramKB lexica as electronic learners' dictionaries](#)

[3.1 Retrieving grammatical information efficiently](#)

[3.2 Providing complete and sufficient grammatical information](#)

[3.2.1 Grammatical information in monolingual electronic learners' dictionaries](#)

[3.2.2 Grammatical information in bilingual electronic learners' dictionaries](#)

[3.3 FunGramKB as a bilingualized electronic dictionary](#)

[4 Conclusion](#)

[References](#)

1 Introduction

Building a large-scale natural language processing (NLP) knowledge base is costly in time and effort, so it is recommendable to design reusable resources, so that they can be easily implemented in other projects. Many researchers distinguish between two notions of "reusability":

- (i) Developing resources for multiple use
- (ii) Using resources which already exist, but for a different purpose

Both approaches of reusability can be adopted with FunGramKB (Functional Grammar Knowledge Base), since our resource (i) has been primarily developed to be reused in various NLP tasks—e.g. information retrieval and extraction, machine translation, or dialogue-based systems, and (ii) can also be used by human users (such as linguists, translators or language learners) in order to retrieve relevant lexical information through a web-based interface. In other words, the same resource can be suitably reused by computers and people. It is this second sense of reusability which this paper examines.

It is well-known that using electronic dictionaries can bring some benefits over paper

¹ This paper is an outcome of the research projects "Construcción de un lexicón multipropósito para el procesamiento del lenguaje natural" (PMAFI-PI-03/1C/03) and "Implementación de una base de conocimiento para el razonamiento del sentido común" (PMAFI-PI-11/1C/05), both of them funded by Universidad Católica San Antonio, Spain.

dictionaries (Nesi 1998; De Schryver 2003):

- (i) powerful search capabilities
- (ii) rapidity of data retrieval
- (iii) removal of linear text restrictions (through hyperlinking)
- (iv) ever-increasing storage capacity

However, it is time to move to a new generation of dictionaries by attempting to adapt them to the electronic medium, and not developing just computerized versions of hardcopies (Perry 1997; De Schryver 2003).

The hypothesis of this paper is that FunGramKB lexica can be reused as human-oriented dictionaries which can make up what electronic learners' dictionaries² lack. More particularly, the purpose of this paper is two-fold: to present a description of the grammatical features and values in FunGramKB lexica (section 2), and to justify why these data can make our knowledge base become a more efficient electronic lexical resource (section 3).

2 FunGramKB

2.1 Overview

FunGramKB is a complex knowledge base which comprises two comprehensive information levels, where several independent modules are interrelated:

Lexical level (i.e. linguistic knowledge):

- The lexicon stores morphosyntactic, pragmatic and collocational information of lexical units.
- The morphicon helps our system to handle cases of inflectional morphology.

Cognitive level (i.e. non-linguistic knowledge):

- The ontology is presented as a hierarchical structure of well-defined concepts used by ordinary humans when talking about everyday situations.
- The cognicon stores procedural knowledge by means of cognitive macrostructures, i.e. script-like schemata in which a sequence of stereotypical actions is organised on the basis of temporal continuity.
- The onomasticon stores information about instances of entities, such as people, cities, products, etc.

The differentiation between lexical and cognitive modules in FunGramKB is not based on the traditional distinction between linguistic and encyclopaedic knowledge respectively, but on the borderline between knowledge about words and knowledge about the model of the world, being the latter more restricted than encyclopaedic knowledge. The main consequence of this two-level design is that every lexical module is language-specific, while every cognitive module is shared by all languages. In other words, computational lexicographers must develop one lexicon and one morphicon for English, one lexicon and one morphicon for Spanish and so on, but knowledge engineers build just one ontology, one cognicon and one onomasticon to process any language input cognitively.

Nowadays language engineers usually develop resources for a particular NLP

² Along this paper we use the term "electronic dictionaries" for human-oriented dictionaries in machine-readable format (e.g. PEDs, dictionaries on CD-ROM/DVD and on-line dictionaries), and the term "learners' dictionary" to refer to any type of dictionary (i.e. monolingual, bilingual or multilingual) written specifically for L2 learners.

application in a particular domain. This *modus operandi* leads to greater efficiency in knowledge representation, but the main drawback is the lack of flexible portability to other domains or tasks when adjusting existing NLP systems to meet the new requirements of other applications (Lenci 2000). Since building a large-scale NLP knowledge base is costly in time and effort, it is recommendable to design reusable and updatable resources, so that they can be easily maintained or improved in different projects along the time (Floridi 1999). FunGramKB is multipurpose in the sense that it is both multifunctional and multilingual. In other words, FunGramKB can be reused in various NLP tasks (e.g. information retrieval and extraction, machine translation, dialogue-based systems, etc) and with several natural languages.

2.2 The lexical entry

With the purpose of designing robust lexica for FunGramKB, we chose to follow EAGLES³ (1993, 1996a, 1996b, 1999; Underwood and Navarretta 1997) and OLIF⁴ (Lieske *et al.* 2001, McCormick 2002, McCormick *et al.* 2004) guidelines for lexicographic standards. Taking them as the starting point of FunGramKB lexical level, we soon realised that, although facilitating a homogeneous and consistent lexicographic work, we could not confine ourselves to their recommendations, particularly regarding full-featured predicate frames and meaning postulates.

Currently lexical modules for English and Spanish are being populated in FunGramKB. Table 1 contains the types of grammatical features being present in our lexical entries:⁵

Table 1. Grammatical features in FunGramKB lexica

	Nouns	Adjectives	Verbs
PART OF SPEECH	en/sp	en/sp	en/sp
NUMBER	en/sp	sp	
GENDER	sp	sp	
DEGREE		en	
ADJECTIVAL POSITION		en/sp	
VERB PARADIGM			en/sp
PRONOMINALIZATION			sp
VOICE/TENSE CONSTRAINTS			en/sp
COUNTABILITY	en/sp		
PREDICATE FRAME:			
- SEMANTIC PATTERN			en/sp
- PREFERENCES			en/sp
- PHRASE PATTERNS			en/sp
- SYNTACTIC PATTERNS			en/sp
- ALTERNATIONS			en/sp

³ EAGLES (The Expert Advisory Group on Language Engineering Standards) is an initiative sponsored by the European Commission which aims to provide recommendations for the standardization of the language technologies field. More particularly, the Computational Lexicons Interest Group is in charge of analysing the main practices in lexicographic encoding by comparing computational lexical resources available in European languages such as Catalan, Danish, Dutch, English, French, German, Greek, Irish, Italian, Portuguese, Spanish and Swedish.

⁴ OLIF (Open Lexicon Interchange Format) is an XML-compliant lexical and terminological exchange format specifically designed for users of language technology tools.

⁵ The “en” and “sp” tags represent English and Spanish languages respectively.

Computationally speaking, FunGramKB lexical entries take the form of feature-value data structures formatted in XML. XML was chosen as the formal language for knowledge representation since it helps the system transfer structured data faster, thus facilitating the access to information. Grammatical features in Table 1 are described more accurately in the next sections.

2.2.1 Part of speech (ptOfSpeech)

FunGramKB lexica only store content words (i.e. nouns, adjectives and verbs). This part-of-speech information can be automatically inferred from the root metaconcept which the basic concept linked to the entry belongs to.⁶ In other words, lexical units mapped to metaconcepts #ENTITY, #EVENT and #QUALITY own values “noun”, “verb” and “adj” respectively in feature <ptOfSpeech>. For example:

```
(1) <entry ConceptUserId="+TABLE_00">
      <mono>
        <keyDC>
          <canForm>table</canForm>
          <ptOfSpeech>noun</ptOfSpeech>
        </keyDC>
      </mono>
    </entry>
```

2.2.2 Number

Feature <number> in FunGramKB lexical entries can have various values: “dreg” (dual-regular),⁷ “dirreg” (dual-irregular), “sgt” (*singulare tantum*), “plt” (*plurale tantum*) and “invar” (invariant). For example:

```
(2) <entry ConceptUserId="+TABLE_00">
      <mono>
        <keyDC>
          <canForm>table</canForm>
          <ptOfSpeech>noun</ptOfSpeech>
        </keyDC>
      </mono>
    </entry>
```

⁶ Our ontological model distinguishes three different conceptual levels, each one of them with concepts of a different type, i.e. metaconcepts, basic concepts and terminals:

- (i) Metaconcepts constitute the upper level in the taxonomy. The analysis of the main upper-level linguistic ontologies led to a metaconceptual model whose design contributes to the integration and exchange of information with other ontologies. Moreover, our metaconcepts play the role of “hidden categories”, that is, concepts which aren’t linked to any lexical unit so that they can serve as hidden superordinates and avoid circularity. Root metaconcepts are #ENTITY, #QUALITY and #EVENT.
- (ii) Basic concepts are used in FunGramKB as defining units enabling the construction of meaning postulates for basic concepts and terminals, as well as taking part as selection preferences in thematic and predicate frames.
- (iii) Finally, the terminal level is not hierarchical structured. The borderline between basic concepts and terminals is based on their definitory potential to take part in meaning postulates.

⁷ Some grammarians use the term “dual” for those lexical units where exactly two people or objects are meant. In FunGramKB the term is used to refer to the singular-plural duality in which lexical units can be involved.

```
</keyDC>
<monoDC>
  <monoMorph>
    <number>dreg</number>
  </monoMorph>
</monoDC>
</mono>
</entry>
```

Values “dreg” and “dirreg” correspond to nouns or adjectives belonging to the classic binary opposition singular-plural. In the case of “dreg”, every language owns a set of expression rules which let the system construct plural signifiers by assigning morphological markers. In the case of “dirreg”, no expression rule can be applied for plural constructions, because plural signifiers cannot be formed out of singular forms. Hence plural forms are stored in the morphicon.

On the other hand, “sgt”, “plt” and “invar” are used for those nouns or adjectives where the plurality does not come from any inflectional process but from syntactic markers. Unlike “invar” (e.g. *species* [en] or *análisis* [sp]), values “sgt” (e.g. *dust* [en] or *salud* [sp]) and “plt” (e.g. *trousers* [en] or *alicates* [sp]) refer to lexical units which are not involved in the singular-plural duality.

2.2.3 Gender

In FunGramKB, feature <gender> presents the following values: “dreg” (dual-regular), “dirreg” (dual-irregular), “m” (masculine), “f” (feminine), “c” (common) and “am” (ambiguous). For example:

```
(3) <entry ConceptUserId="+CAT_00">
  <mono>
    <keyDC>
      <canForm>gato</canForm>
      <ptOfSpeech>noun</ptOfSpeech>
    </keyDC>
    <monoDC>
      <monoMorph>
        <gender>dreg</gender>
      </monoMorph>
    </monoDC>
  </mono>
</entry>
```

Values “dreg” and “dirreg” are assigned to nouns and adjectives taking part in the masculine-feminine opposition. In the case of “dreg”, the system can build the feminine form by applying a language-specific set of regular expressions. In the case of “dirreg”, the gender duality of the signifier cannot be derived automatically through regular expressions, so the feminine form must be stored in the morphicon.

Values “m” and “f” are assigned to nouns which only appear as masculine (e.g. *árbol*) or feminine (e.g. *tribu*) respectively.

Other nouns and adjectives can belong to the “c” category, where the same word form is used for both genders (e.g. *artista*, *verde*).

Finally, “am” is assigned to those nouns which admit both genders without a shift in form or meaning (p.ej. *mar*).

2.2.4 Degree

Feature <degree> enables the system to know if comparative and superlative forms of English adjectives⁸ are built in an inflectional or periphrastic way, since FunGramKB is not ready to find out the number of syllables in lexical units. For example:

```
(4) <entry ConceptUserId="+GENEROUS_00">
    <mono>
        <keyDC>
            <canForm>generous</canForm>
            <ptOfSpeech>adj</ptOfSpeech>
        </keyDC>
        <monoDC>
            <monoMorph>
                <degree regularEn="peri">reg</degree>
            </monoMorph>
        </monoDC>
    </mono>
</entry>
```

The morphicon stores all those cases where the lemma is not used as the root of the comparative or superlative forms.

2.2.5 Adjectival position (synType, synPosition)

FunGramKB lexica store information about the standard position of adjectives within a phrase, distinguishing three different values: “attrib” (just attributive), “pred” (just predicative) and “attrib-pred” (attributive or predicative). For example:

```
(5) <entry ConceptUserId="+BEAUTIFUL_00">
    <mono>
        <keyDC>
            <canForm>handsome</canForm>
            <ptOfSpeech>adj</ptOfSpeech>
        </keyDC>
        <monoDC>
            <monoSyn>
                <synType>attrib-pred</synType>
            </monoSyn>
        </monoDC>
    </mono>
</entry>
```

In the case of Spanish adjectives in attributive function, a further specification is made regarding their occurrence as “prenoun” (just premodifiers), “postnoun” (just postmodifiers) or “pre-post” (premodifier or postmodifier). For example:

```
(6) <entry ConceptUserId="+GOOD_00">
    <mono>
        <keyDC>
            <canForm>buen</canForm>
            <ptOfSpeech>adj</ptOfSpeech>
        </keyDC>
```

⁸ This feature is not pertinent to Spanish adjectives, since they are constructed periphrastically. Exceptional instances taking the organic comparative and superlative (e.g. *mejor*, *peor*, *superior*, *inferior*, etc.) are stored in the morphicon.

```
<monoDC>
  <monoSyn>
    <synPosition>prenoun</synPosition>
  </monoSyn>
</monoDC>
</mono>
</entry>
```

2.2.6 Verb paradigm

Lexical entries of verbs also specify if any irregularity occurs in their inflectional paradigm. For example:

```
(7) <entry ConceptUserId="+SELL_00">
  <mono>
    <keyDC>
      <canForm>buy</canForm>
      <ptOfSpeech>verb</ptOfSpeech>
    </keyDC>
    <monoDC>
      <monoMorph>
        <verbParadigm>irreg</verbParadigm>
      </monoMorph>
    </monoDC>
  </mono>
</entry>
```

In any case, FunGramKB relies on the morphicon for the construction of the inflectional forms of lexical units.

2.2.7 Pronominalization (synType)

Pronominalization covers the phenomenon involving reflexive clitic variations of the headword. For example:

```
(8) <entry ConceptUserId="+MOVE_00">
  <mono>
    <keyDC>
      <canForm>ir</canForm>
      <ptOfSpeech>verb</ptOfSpeech>
    </keyDC>
    <monoDC>
      <monoSyn>
        <synType>refl-o</synType>
      </monoSyn>
    </monoDC>
  </mono>
</entry>
```

The four different values are described as follows:

- "refl-n" (never reflexivized): No reflexive pronoun can be used with the verb, e.g. *parir*.
- "refl-a" (always reflexivized): A reflexive pronoun is obligatorily cliticised to the verb, e.g. *jactarse*.
- "refl-o" (optionally reflexivized): The verb can be reflexively marked, but the

presence of the reflexive clitic subtly shades the meaning of the verb's root (Robertson and Turley 2003), e.g. *ir(se)*.

- "refl-g" (grammatically reflexivized): This contextual variant of *se* is traditionally conceived as a grammatical device which affects the canonical transitivity of predicates according to the following cognitive criteria:

- (i) transitive verbs which can be reflexively marked in order to establish a relationship of identity between the Theme and Referent (or Target) arguments of the predicate frame:

(a) *Se miró en el espejo.*

- (ii) transitive verbs which can be reflexively marked in order to force the Theme argument into the background—including cases such as (b) passive, (c) decausative verbs or (d) indeterminate reflexives (Robertson and Turley 2003):

(b) *Muchas pirámides se construyeron en el México antiguo.*

(c) *El vaso se rompió.*

(d) *Por aquí se come mucho helado.*

In FunGramKB, pronominalization is not pertinent to those cases where the presence of the reflexive clitic alters the meaning of the verb, e.g. *acordar* (= agree) and *acordarse* (= remember), since they are conceived as two different headwords in FunGramKB lexical entries.

2.2.8 Voice and tense constraints

FunGramKB is also able to cover constraints on voice and tense, because there are verbs whose morphological paradigm lacks some inflectional forms. For example, *soler* is not conjugated in the future and conditional tenses.

2.2.9 Countability (synType)

Availability of noun countability information is essential to explain the syntactic behaviour of both English and Spanish nouns—although less attention has been paid to this type of classification by Spanish lexicographers (Bosque Muñoz 1999). Countability is linguistically realized by means of morphosyntactic contrasts, particularly affecting subject-verb concordance and the use of some determiners. The analysis of the world model does not provide enough information as to how the categorization of nouns can be inferred, so the value of this feature cannot be derived from the location that the concept of a noun takes in the ontology. In fact, there is nothing in the make-up of things that can explain why some are perceived as mass and others as individual entities. Therefore, words must set these distinctions because they form part of our knowledge on language and not of the reality denoted by language (Bosque Muñoz 1999). Various languages can even categorize the same entity in a different way: e.g. in Spanish *consejo* and *mueble* are countable but their English equivalents *advice* and *furniture* are uncountable.

Although six degrees of countability can be found in English (Downing and Locke 1992), FunGramKB provides just three values for this feature: a noun is always countable ("cnt"), a noun is always uncountable ("mass") or a noun can sometimes behave as countable and other times as uncountable ("mass-cnt"). For example:

(9) `<entry ConceptUserId="+FLOUR_00">
 <mono>
 <keyDC>
 <canForm>flour</canForm>
 <ptOfSpeech>noun</ptOfSpeech>
 </keyDC>
 <monoDC>
 <monoSyn>
 <synType>mass</synType>
 </monoSyn>
 </monoDC>
 </mono>
</entry>`

The “mass-cnt” category covers cases of “recategorization” (Lyons 1968), where Spanish is one of the languages which shows more facility for this phenomenon (Bosque Muñoz 1999). Psychologically, it is more valid to treat these cases of lexical recategorization by relating countable and non-countable uses of dual nouns in the same lexical entry, instead of creating an entry for a noun denoting a class of objects and another entry for a noun referring to the material or substance from which a well-delimited unit is extracted.

2.2.10 Predicate frame

In computational linguistics, predicate frames include key information which lets the machine build the underlying predication of an input text. To illustrate, we take the predicate frame of *eat* in FunGramKB, presented according to Dik’s Functional Grammar model (10) and the XML format (11):⁹

(10) break_{verb} (x₁: +HUMAN_00, +ANIMAL_00)_{NP / Subj / Theme} (x₂: +SOLID_00)_{NP / DO / Referent}

(11) `<entry ConceptUserId="+BREAK_00">
 <mono>
 <keyDC>
 <canForm>break</canForm>
 <ptOfSpeech>verb</ptOfSpeech>
 </keyDC>
 <monoDC>
 <monoSyn>
 <frame>
 <pattern>
 <slot phrase="NP" syn="S" sem="Theme" />
 <slot phrase="NP" syn="DO" sem="Referent" />
 </pattern>
 <alternation>Inchoativity</alternation>
 </frame>
 </monoSyn>
 </monoDC>
 </mono>
</entry>`

The predicate frame is a structural scheme stating the quantitative and qualitative valences of the verb: two subcategorized arguments with the semantic functions

⁹ Although nouns and adjectives can also have predicate frames, the current version of FunGramKB only includes those of verbs.

Theme and Referent and the selection preferences HUMAN, ANIMAL and SOLID respectively.¹⁰ The order of presentation of arguments in predicate frames is not subject to the linear order of the constituents making up the linguistic realization, since the latter is constructed from expression rules of the target language. In fact, it is an arbitrary canonic order based on the list of semantic functions.

Regarding qualitative valence, it is worthwhile to mention the following features:

- A single predicate frame cannot contain more than one argument with the same semantic function.
- Selection preferences are represented through basic concepts from the ontology

Apart from the valences, FunGramKB predicate frames are also enriched with information about subcategorization patterns and diathetic alternations. Subcategorization patterns state the phrase realizations and their syntactic behaviour in which the verb can co-occur, taking into account only those complements which are semantically necessary. For instance, subcategorization describes that *button* is a transitive verb, whose arguments are realized by noun phrases with the syntactic functions Subject and Object. On the other hand, the diathetic alternations covered by FunGramKB are inchoativity and causativity—the two alternations which involve a shift of transitivity in the verb.

3 Benefits of FunGramKB lexica as electronic learners' dictionaries

Inquiring into the suitability of FunGramKB lexical model for human look-up requires to define previously the general purpose of dictionaries. Many lexicographers mistakenly think that the function of an ideal dictionary is to list all lexical entries in a language, as well as providing every type of information (e.g. morphological, syntactic, semantic, stylistic, etc) about the behaviour of lexical units in the linguistic system. In practice, however, not all dictionaries should be identical, since they must meet the particular communicative needs of their users. In other words, certain types of information should be restricted to certain types of dictionaries. For example, phonetic transcriptions are more suitable for learners' dictionaries, but translation equivalents are characteristic of bilingual dictionaries. On the other hand, learners' dictionaries should also adapt to users' command of the language: beginners are more willing to deal with translations, while advanced students need more accurate grammatical and pragmatic information in order to develop their communicative competence (Castillo Carballo and García Platero 2003). Therefore, we redefine the main goal of dictionaries as "to provide complete and sufficient lexical information efficiently". Two distinctive parts of this definition must be highlighted, since each one of them raises some lexicographical problems:

- (i) Dictionaries should provide "complete and sufficient" specifications, which implies that (a) dictionaries should focus on codification and decodification functions in a balanced way, and (b) lexicographers should be able to adapt dictionaries to users' profiles. On the one hand, although electronic dictionaries can store much more data than printed dictionaries, paradoxically they do not hold all the information necessary for language production—a particularly serious problem in the case of monolingual dictionaries for Spanish (Moreno Fernández 2000; Castillo Carballo and García Platero 2003). On the other hand, abundance of lexical information can

¹⁰ As you can see in (11), selection preferences are not stored in lexical entries. Because of their cognitive burden, selection preferences are part of thematic frames in FunGramKB ontology. However, since predicate frames are derived from thematic ones, selection preferences can definitely take part in the full construction of our predicate frames.

slow down the look-up process, making dictionaries less efficient (Hernández 2000). Unfortunately, contents in current paper- or computer-based dictionaries do not suitably adapt to users' profiles.

- (ii) Retrieving information “efficiently” implies obtaining the expected results in the shortest period of time possible (Roberts 1997). Unlike paper-based dictionaries, one of the main benefits of electronic dictionaries is that data are stored in an accessible way, so you can retrieve information very easily and quickly. However, querying systems in current electronic dictionaries are rather primitive.

These two problems are mainly due to the fact that most electronic dictionaries are still based on printed dictionaries, so both of them store the same type and amount of information. In fact, they share the same microstructure, since most electronic dictionaries even attempt to reproduce the external image of their paper counterparts (Gelpí Arroyo 2003). Many researchers think that the solution lies on the development of electronic dictionaries from scratch and not as adaptations of printed sources (Nesi 1998). We agree that computerizing paper-based dictionaries is the root of the problem; yet rather than developing the resource *ex professo*, we choose to reuse NLP lexica (i.e. computer-oriented) as electronic dictionaries (i.e. human-oriented). More particularly, in the following sections we justify why reusing FunGramKB lexica as electronic learners' dictionaries can resolve deficiencies related to (i) and (ii) from a grammatical view.

3.1 Retrieving grammatical information efficiently

The most outstanding characteristics of today's electronic dictionaries are their search capabilities and the speed with which data are retrieved. For example, information on any word in a definition can be typically obtained by double-clicking on that word—an interactive process called “hyperlinking” (Nesi 1998). Lexical accessibility is also enhanced by allowing complex search routes including wildcards and/or Boolean operators. However, the ideal approach would be to use search routes which could find exactly what you need and no more (Roberts 1997). In other words, we must give up designing electronic dictionaries as information retrieval systems, whose search engine returns whole lexical entries, and start to conceive them as information extraction systems, which would provide just relevant information to users' queries.

To illustrate, let us examine feature <verbParadigm>, for which NLP lexica prove to be more convenient than standard electronic dictionaries, particularly when handling languages with a high degree of inflection. For example, Spanish dictionaries usually link every verb headword to a conjugation model, from which users can infer the headword's whole verb paradigm—as you can see in (12) from *Diccionario Salamanca de la Lengua Española*:

- (12) **oler**
verbo transitivo, intr.
...
Conjugación: 55

With reference to this feature in FunGramKB lexicon for Spanish, the difference is not found in the quantity of information but in the way this information can be retrieved. For example, FunGramKB interface can let users retrieve a particular verb form as long as its person, number, tense and mood features are provided. Moreover, users could also be interested in knowing just the irregular forms of a particular verb, and not its whole paradigm.

3.2 Providing complete and sufficient grammatical information

As we explained above, lexicographical information involves two troublesome issues: completeness and sufficiency. Firstly, regarding adaptability to users' profiles (i.e. sufficiency), a single learners' dictionary is inexorably restricted to one or two languages (monolingual and bilingual respectively), and to a certain learning level—or to advanced L2 learners by default, as this is typically the case for Spanish.¹¹ The potential of the electronic medium should alter the present-day lexicographical scene; dictionaries could be designed as dynamic resources which would evolve along with the user's learning process. In fact, different types of dictionaries are preferred depending on users' knowledge on the foreign language, e.g. bilingual dictionaries for beginners, monolingual learners' dictionaries for intermediate students, and monolingual general-purpose dictionaries for advanced learners (Stein 1990). However, dynamic selection of information within a single resource in accordance with the user's knowledge can only be performed if data are systematically stored in a highly-structured fashion, as in the case of computer-oriented lexica. More particularly, De Schryver (2003) suggests a pop-up adaptive multimedia Internet dictionary as a sample of lexicography in the third millennium. This type of dictionary would have to reject the traditional model of "electronic dictionaries as adaptations of printed sources" because of the need to integrate NLP techniques. On the one hand, the "pop-up" functionality implies that it is a context-sensitive electronic dictionary, which would require formal syntagmatic and paradigmatic analyses of the source lexical unit. On the other hand, "adaptive" dictionaries imply that the system can adapt dictionary data to the automatically-derived user profile, which would require the construction of a user model characteristic of intelligent tutorial systems. Therefore, next-generation dictionaries would be developed as systems implementing Human Language Technologies. To this respect, FunGramKB lexica can be found advantageous, since there is no need to transduce machine-readable entries into a lexical database with the risk of inheriting all deficiencies from printed dictionaries.

Secondly, FunGramKB can make up for the lack of robust lexicographical resources aimed at both language comprehension and production (i.e. completeness). According to Moreno Fernández (2000), the commonest features of monolingual learners' dictionaries are:¹²

- (i) selection of lexical units according to coherent systematic criteria,
- (ii) easily-understandable definitions, with or without a closed list of defining terms,
- (iii) abundant exemplification of definitions, created *ad hoc* or taken from corpora,
- (iv) and grammatical and pragmatic specifications of lexical units.

Currently FunGramKB can provide (i) and (ii). Feature (iii) will soon be satisfied, since we intend to carry out a full syntactic and semantic tagging of a corpus using our knowledge base; thus it wouldn't be very difficult to automatically extract examples which could illustrate not just meaning postulates in the ontology but also predicate frames in the lexicon. Following the aforementioned lexicographical model of new

¹¹ Remember that most dictionaries are commercial products subject to the editorial policy, so lexicographical work is usually aimed at finding a considerable niche of market in order to make dictionaries profitable.

¹² L2 dictionaries for Spanish do not meet all these features, unlike English learners' dictionaries, whose tradition in pedagogical lexicography can be traced back to the 40's with A. S. Hornby, E. V. Gatenby and H. Wakefield's *A Learner's Dictionary of Current English* (1948).

generation, dictionaries should also be “multimedia”, which means that users should be allowed to access multimedia corpora—consisting of text, audio and video—from lexical entries (De Schryver 2003).

3.2.1 Grammatical information in monolingual electronic learners’ dictionaries

Regarding feature (iv), table 2 describes how grammatical features in table 1 are treated in the monolingual electronic learners’ dictionaries of our study:¹³

Table 2. Comparative analysis of FunGramKB lexica and monolingual electronic learners’ dictionaries¹⁴

FunGramKB	LDOCE	OALD	DSLE
Part of speech	=	=	=
Number ¹⁵	=	=	—
Gender			=
Degree ¹⁶	—	=	
Adjectival Position	=	=	?
Verb paradigm	=	=	=
Pronominalization			=
Voice/Tense constraints	=	=	=
Countability ¹⁷	=	=	—
Predicate frame:			
- Semantic pattern	?	?	?
- Preferences	—	—	—
- Phrase patterns	=	=	—
- Syntactic patterns	?	?	?
- Alternations	=	=	=

Some researchers (Porto Dapena 2002) miss the presence of other categorizations

¹³ Electronic learners’ dictionaries studied in this paper include:

(i) English monolingual

* *Longman Dictionary of Contemporary English Online* (<http://pewebdic2.cw.idm.fr>) [LDOCE]

* *Oxford Advanced Learners’ Dictionary Online*

(<http://www.oup.com/elt/catalogue/teachersites/oald7/?cc=global>) [OALD]

(ii) Spanish monolingual

* *Diccionario Salamanca de la Lengua Española* (<http://fenix.cnice.mec.es/diccionario>) [DSLE]

(iii) Bilingual

* *Diccionario Cambridge Klett Compact Español-Inglés English-Spanish*

(<http://dictionary.cambridge.org/Default.asp?dict=S>) [Klett]

* *Diccionario Richmond Español-Inglés Inglés-Español* (<http://fenix.cnice.mec.es/richmond>) [Richmond]

¹⁴ For each grammatical feature, we indicate if information provided by the electronic dictionaries is the same (=) or less accurate (-) than in FunGramKB lexica. When a grammatical specification is not explicitly present in the lexical entry, but it was expected to be there, the question mark (?) is used.

¹⁵ DSLE shows <number> of nouns, but not of adjectives.

¹⁶ LDOCE shows irregular comparative forms of adjectives, but not regular ones.

¹⁷ DSLE does not describe cases of lexical recategorization.

which can condition the grammatical behaviour of nouns (e.g. animate/inanimate, human/non-human, etc) or verbs (e.g. perfective/imperfective, punctual/durative, etc). Indeed, FunGramKB can supply this information about lexical units; however, it is not really stored in the lexicon but in the ontology: strictly speaking, these categorizations involve cognitive specifications of entities and events, so they cannot become grammatical features of lexical units but properties of concepts.

The remainder of this section describes the findings of our analysis with respect to the predicate frame. Here are some examples:

(13) [LDOCE]

rec•om•mend [transitive]

1 to advise someone to do something, especially because you have special knowledge of a situation or subject

recommend (that)

I recommend that you get some professional advice.

*Doctors **strongly recommend** that fathers should be present at their baby's birth.*

recommend doing something

I would never recommend using a sunbed on a regular basis.

Sleeping tablets are not recommended in this case.

*It is dangerous to exceed the **recommended dose**.*

2 to say that something or someone is good, or suggest them for a particular purpose or job:

I recommend the butter chicken - it's delicious.

Can you recommend a good lawyer?

recommend something to somebody

I recommend this book to anyone with an interest in chemistry.

recommend something for something/somebody

Which type of oil do you recommend for my car?

recommend somebody for something

I have decided to recommend you for the directorship.

highly/thoroughly recommend

The hotel is highly recommended.

(14) [OALD]

rec•om•mend verb

1 [VN] ~ **sb/sth (to sb) (for / as sth)** to tell sb that sth is good or useful, or that sb would be suitable for a particular job, etc.: *Can you recommend a good hotel?* ◊ *I recommend the book to all my students.* ◊ *She was recommended for the post by a colleague.* ◊ *The hotel's new restaurant comes **highly recommended** (= a lot of people have praised it).*

2 to advise a particular course of action; to advise sb to do sth: [VN] *The report recommended a 10% pay increase.* ◊ *It is dangerous to exceed the recommended dose.*

◊ *a recommended price of \$50* ◊ [V (**that**)] *I recommend (that) he see a lawyer.* ◊ (BrE also) *I recommend (that) he should see a lawyer.* ◊ [VN (**that**)] *It is **strongly recommended** that the machines should be checked every year.* ◊ [VN to inf] *We'd recommend you to book your flight early.* ◊ [V -ing] *He recommended reading the book before seeing the movie.* [also VN -ing, V wh-]

(15) [DSLE]

recomendar

verbo transitivo

1 Aconsejar o indicar < una persona > a [otra persona] [una cosa]: *El médico le recomendó tranquilidad.*

2 Hablar < una persona > en favor de [otra persona] ante [otra tercera persona] para ayudarla en [una cosa]: *Te recomendé al director para ocupar el puesto.* **Conjugación:**

58

A synonymic model of definition is more rigorously applied in (15); in the case of

verbs, meaning postulates highlight prototypical neighbouring words making up the syntagmatic axis of the *definiendum* by enclosing them within angle or square brackets, which are deemed to serve as grammatical markers (Porto Dapena 2002). Therefore, (15) pinpoints more clearly the quantitative valence than (13-14), where the number of subcategorized arguments of a verb is inferred from its phrase patterns.

On the other hand, selection preferences of verb arguments can be identified in the phrase patterns (13-14), or inside the definition text but separated from the real *definiens* by brackets (15). However, these selection preferences are usually underspecified through generic or broad-meaning lexical units—e.g. *somebody*, *something* (13-14), or *persona*, *cosa* (15).¹⁸

Concerning phrase patterns, English learners' dictionaries are found to be very valuable, although formal linguistic terminology is usually avoided. On the contrary, Spanish dictionaries do not cover this grammatical feature so adequately, where this information can only be inferred from the exemplifications of the headword. In none of our dictionaries are phrase realizations tagged syntactically anyway.

Finally, (16-18) show that the dictionaries of our study let users recognize inchoative alternations, although not before getting acquainted with selection preferences of arguments and reading the exemplifications of the headword:

(16) [LDOCE]

break

1 SEPARATE INTO PIECES

a) [transitive] if you break something, you make it separate into two or more pieces, for example by hitting it, dropping it, or bending it:

I had to break a window to get into the house.

Don't lean on the fence like that - you'll break it!

break something in half/two

He broke the biscuit in half and handed one piece to me.

Break the chocolate **into** small **pieces** and melt it over a gentle heat.

b) [intransitive] if something breaks, it separates into two or more pieces:

He kept pulling at the rope until it broke.

The frames are made of plastic and they tend to break quite easily.

(17) [OALD]

break

verb

IN PIECES

1 ~ (sth) (in / into sth) to be damaged and separated into two or more parts, as a result of force; to damage sth in this way: [V] *All the windows broke with the force of the blast.*

The bag broke under the weight of the bottles inside it. She dropped the plate and it

broke into pieces. [VN] *to break a cup / window* *She fell off a ladder and broke her arm. He broke the chocolate in two.*

(18) [DSLE]

romper

verbo transitivo

1 Hacer < una persona o una cosa > trozos irregulares de [una cosa]: *El balón rompió el cristal. He roto el jarrón al tirarlo al suelo.*

2 Hacer < una persona, un animal o una cosa > un agujero o una raja en [un material flexible]: *Los niños rompieron el balón con una navaja. Los gatos han roto las cortinas de la ventana.*

...

verbo pronominal

¹⁸ Traditionally, selection preferences in Spanish dictionaries have been scarce, and when appearing, they have been coded through the formula *Referido a...*

- 1 Partirse < una cosa > en trozos irregulares: *Él se ha roto la pierna por tres sitios. El vaso se rompió al caerse.*
2 Hacerse un agujero o una raja en < un material flexible >: *Este papel se rompe fácilmente. Las medias se me han roto.*

In short, the above entries from learners' dictionaries prove that abundant exemplification of the headword together with users' language competence are essential to complete the insufficient information about predicate frames. Moreover, information is so dispersedly arranged within lexical entries that learners perceive that the development of their language production skills is dramatically hindered. Indeed, human-oriented lexicographical resources are lacking in standardization when displaying information on predicate frames, not only concerning the quantity of data to be stored but also the manner of presenting these data within the lexical microstructure. FunGramKB lexica provide more robust predicate frames, which means that they are more informative (particularly in semantic and syntactic patterns) as well as presenting information in a more accurate and integrated way. The benefit of this well-structured model of predicate frame lies on the enhanced efficiency of information retrieval. For example, with no need to browse whole lexical entries, dictionary readers can be able to know if a particular a verb argument in a particular syntactic position must be headed by a preposition (and if so, which one), or if that argument should be linguistically realized by lexical units from a particular cognitive dimension, or if the verb can be detransitivized by promoting the object to subject.

3.2.2 Grammatical information in bilingual electronic learners' dictionaries

In the consultation of any bilingual dictionary, just one of its two languages is the user's mother tongue, so this type of resources should contain more grammatical descriptions for lexical items in the L2 language. Naturally, since lexicographers don't know the dictionary reader's native language beforehand, the most sensible approach would be to provide both sections of bilingual dictionaries with sufficient grammatical information. However, as you can see in table 3, bilingual dictionaries suffer from more deficiencies in grammatical knowledge than monolingual resources.

Table 3. Comparative analysis of FunGramKB lexica
and bilingual electronic learners' dictionaries

FunGramKB	Klett	Richmond
Part of speech	=	=
Number ¹⁹	—	—
Gender	=	=
Degree ²⁰	—	—
Adjectival Position	?	?

¹⁹ Klett and Richmond show <number> of nouns in English and Spanish, but no information is provided about this feature in the case of Spanish adjectives.

²⁰ In Klett and Richmond, degree information is restricted to irregular comparative and superlative forms of English adjectives.

Verb paradigm ²¹	—	—
Pronominalization	=	=
Voice/Tense constraints	?	?
Countability	?	?
Predicate frame:		
- Semantic pattern	?	?
- Preferences	?	—
- Phrase patterns	?	?
- Syntactic patterns	?	?
- Alternations	?	?

Grammatical features in table 3 have been examined in both sections of the dictionaries (i.e. Spanish-English and English-Spanish). Our study reveals that when relevant information is supplied, this information only corresponds to the headword and not to its translation equivalents. This lexicographical trend results from space restrictions in paper-based dictionaries, on which most electronic dictionaries are based. Consequently, this layout of bilingual dictionaries does not favour language production, since learners have to look up translation equivalents in the other section of the dictionary, or even in monolingual dictionaries—due to incomplete information for correct usage of words found in current bilingual dictionaries. This loss of grammatical knowledge in bilingual dictionaries is more remarkable particularly with predicate frames, as it is shown in (19-20):²²

- (19) [Klett]
[a] recomendar <e -> ie> vt *to advise*; nos recomendó no salir de casa *he/she advised us not to leave the house*

[b] recommend vt *recomendar*; it is not to be ~ed *no es recomendable*

- (20) [Richmond]
[a] recomendar
 verbo
una película, un libro, un hotel, a una persona
to recommend
 te la recomiendo, es muy buena
 I recommend it, it's very good aconsejar
to recommend
 el médico le recomendó reposo absoluto
 the doctor recommended complete rest
 te recomiendo que esperes
 I recommend that you wait

²¹ Lexical entries in Klett and Richmond include all irregular forms of English verbs. In the case of Spanish verbs, however, inflectional paradigms are not displayed in the same way; Richmond gives no description at all, and Klett applies inconsistent criteria when handling irregular verb inflections, by pointing out (i) where the irregularity lies exactly or that (ii) there is simply some kind of irregularity:

- (i) [Klett]
soler <o -> ue> vi ~ hacer *to be in the habit of doing*;
 ...

- (ii) [Klett]
dormir irr I vi 1. (descansar) *to sleep*;
 ...

²² We have introduced tags [a] and [b] in (20) in order to report that the headword was looked up in the Spanish-English and English-Spanish sections of the dictionary respectively.

[b] recommend

verb

recommends, recommending, recommended

1 *a place, a person, a film*

recomendar

she recommended this course to me

me recomendó este curso

2 *to advise*

aconsejar

recomendar

he recommended a few days' rest

aconsejó recomendó unos días de descanso

I recommend that you make a copy

te aconsejo te recomiendo que hagas una copia

As we have proved, FunGramKB lexica provide more accurate and consistent grammatical information than current learners' dictionaries, because the latter inherit the incompleteness of information from their paper counterparts—partly due to space restrictions in books. Although learners' dictionaries have a higher degree of internal formalization than other types of dictionaries, they don't succeed in supplying adequate grammatical information to support language production. Most lexicographers prefer to include simple grammatical codes—such as those describing the transitivity of verbs, arguing that dictionary readers are not usually well-trained in linguistics. However, computer-oriented dictionaries allow us to present complex syntactic information in a more friendly way.

3.3 FunGramKB as a bilingualized electronic dictionary

The question on which type of dictionary is best for FL students has generated a lot of discussion among researchers, whose main proposals Koren (1997) summarizes as follows:

- 1a) Use both monolingual and bilingual dictionaries together (Piotrowski 1989).
- 2a) Use monolingual learners' dictionaries (Stein 1989).
- 3a) Use both sections of bilingual dictionaries (Walz 1990).
- 4a) Use *ad hoc* glossaries (Bensoussan 1983).

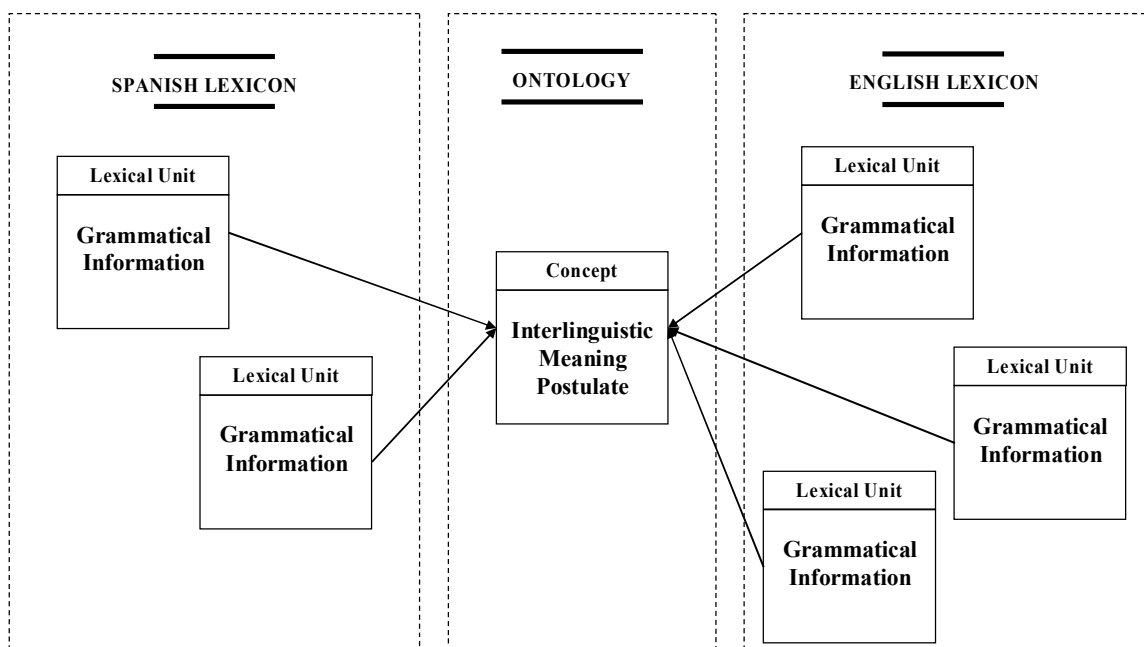
However, as Koren (1997) suggests, each one of these proposals raises some problems in practice:

- 1b) Looking up a word in several dictionaries is a tiresome task.
- 2b) Contrary to teachers' recommendations, most FL students prefer bilingual dictionaries to monolingual ones, since bilingual dictionaries are easier to use.
- 3b) This approach does not provide enough information for correct usage of words.
- 4b) FL teachers are not willing to prepare glossaries for unedited authentic materials; moreover, these glossaries would not be the type of materials that students would read in the future.

In fact, all these drawbacks of monolingual and bilingual dictionaries are rooted in the same problem: no dictionary can meet all the students' needs. This inconvenience can only be overcome with an interface capable of combining the strengths of both types of dictionaries in an integrated way. To this respect, FunGramKB can be used as

a “bilingualized”²³ electronic dictionary, which turns out to be more efficient for language learning than working with separate monolingual and bilingual dictionaries (Laufer and Hadar 1997; Laufer and Kimmel 1997). A bilingualized entry typically contains monolingual information about the L2 headword (e.g. grammatical specifications, definition, examples, etc) and L1 translation equivalents, thus avoiding the tendency to view vocabulary acquisition as word-pair learning. However, this model can be enhanced with a concept-oriented knowledge base capable of featuring the same type and amount of information for L1 and L2 headwords (figure 1).

Figure 1. Lexicon-Ontology mapping in FunGramKB



For example:

```

(21) <entry ConceptUserId="+SINK_00">
      <mono>
        <keyDC>
          <canForm>sink</canForm>
          <ptOfSpeech>verb</ptOfSpeech>
          <semReading>01</semReading>
        </keyDC>
        <monoDC>
          <monoSyn>
            <frame>
              <pattern>
                <slot phrase="NP" syn="S" sem="Theme" />
              </pattern>
              <alternation>Causativity</alternation>
            </frame>
          </monoSyn>
        </monoDC>
      </mono>
      <transfer>
        <keyDC>
          <canForm>hundir</canForm>

```

²³ A term coined by Laufer and Melamed (1994).

```

        <language>sp</language>
        <ptOfSpeech>verb</ptOfSpeech>
        <semReading>01</semReading>
    </keyDC>
</transfer>
</entry>

```

```

(22) <entry ConceptUserId="+SINK_00">
    <mono>
        <keyDC>
            <canForm>hundir</canForm>
            <ptOfSpeech>verb</ptOfSpeech>
            <semReading>01</semReading>
        </keyDC>
        <monoDC>
            <monoSyn>
                <synType>refl-g</synType>
                <frame>
                    <pattern>
                        <slot phrase="NP" syn="S" sem="Agent" />
                        <slot phrase="NP" syn="DO" sem="Theme" />
                    </pattern>
                </frame>
                <alternation>Inchoativity</alternation>
            </monoSyn>
        </monoDC>
    </mono>
    <transfer>
        <keyDC>
            <canForm>sink</canForm>
            <language>en</language>
            <ptOfSpeech>verb</ptOfSpeech>
            <semReading>01</semReading>
        </keyDC>
    </transfer>
</entry>

```

```

(23) <entry ConceptUserId="+SINK_00">
    <mono>
        <keyDC>
            <canForm>zozobrar</canForm>
            <ptOfSpeech>verb</ptOfSpeech>
            <semReading>01</semReading>
        </keyDC>
        <monoDC>
            <monoSyn>
                <synType>refl-n</synType>
                <frame>
                    <pattern>
                        <slot phrase="NP" syn="S" sem="Theme" />
                    </pattern>
                </frame>
            </monoSyn>
        </monoDC>
    </mono>
    <transfer>
        <keyDC>

```

```

    <canForm>sink</canForm>
    <language>en</language>
    <ptOfSpeech>verb</ptOfSpeech>
    <semReading>01</semReading>
  </keyDC>
</transfer>
</entry>

```

The architecture of this model presents some advantages over current lexicographical resources (monolingual, bilingual and bilingualized dictionaries). Firstly, with the aid of a natural language generator, meaning postulates in the form of interlinguistic cognitive representations can be transduced into natural language definitions; in this way there is no need to store identical definitions repeated in different languages. As FunGramKB can support a user-adaptive interface—while redundancy is maximally reduced in the knowledge base, users themselves will be allowed to select the language of definitions. Secondly, any of the two languages may be equally effective as L1 or L2, because the same amount of grammatical information is supplied. Thirdly, this model favours the integration of further lexica, resulting in a “multilingualized” dictionary. Finally, FunGramKB facilitates contrastive learning by automatically derive divergences between translation equivalents.²⁴

To illustrate the latter issue, let us consider the case of predicate frames. For example, an L1 lexical unit does not always own the same predicate frame as its L2 translation equivalent, since various types of divergences can occur—e.g. deviations in the phrase realizations of the Goal argument in (24-25):

```

(24) <entry ConceptUserId="+ENTER_00">
    <mono>
      <keyDC>
        <canForm>enter</canForm>
        <ptOfSpeech>verb</ptOfSpeech>
      </keyDC>
      <monoDC>
        <monoSyn>
          <frame>
            <pattern>
              <slot phrase="NP" syn="S" sem="Theme" />
              <slot phrase="NP" syn="A" sem="Goal" />
            </pattern>
          </frame>
        </monoSyn>
      </monoDC>
    </mono>
  </entry>

```

```

(25) <entry ConceptUserId="+ENTER_00">
    <mono>
      <keyDC>
        <canForm>entrar</canForm>
        <ptOfSpeech>verb</ptOfSpeech>
      </keyDC>
      <monoDC>
        <monoSyn>

```

²⁴ Dorr (1994) provides a formal description and classification of lexical-semantic divergences occurring when translating source structures into target ones.

```

<frame>
  <pattern>
    <slot phrase="NP" syn="S" sem="Theme" />
    <slot phrase="PP" syn="A" sem="Goal">
      <prep>en</prep>
    </slot>
  </pattern>
</frame>
</monoSyn>
</monoDC>
</mono>
</entry>

```

There can also be quantitative differences between predicate frames of two translation equivalents—that is, a mismatch can occur in the number of arguments between source and target predicate frames. FunGramKB conceptualist approach facilitates the description of both qualitative and quantitative divergences. More particularly, predicate frames are grounded on cross-linguistic cognitive constructs which are called “thematic frames”.²⁵ For example, *sink* [en], *hundir* and *zozobrar* [sp] are verbs which trigger the same thematic frame, since both of them place their participants in the same cognitive situation:

```

(26) <thFrame>
      <slot type="x" sem="Agent" />
      <slot type="x" pref="+SOLID_00" sem="Theme" />
      <slot type="x" pref="+LIQUID_00 ^ +MUD_00" sem="Location" />
      <slot type="x" sem="Origin" />
      <slot type="x" sem="Goal" />
      <slot type="f" pref="+SLOW_00" sem="Manner" />
</thFrame>

```

By virtue of their meaning postulate, these verbs are linked to the same concept (+SINK_00) and thus share the same thematic frame among other cognitive properties. However, these verbs differ in their predicate frames, since they show different profiled arguments (21-23).

In other words, both source and target lexical units are linked to the same thematic frame at the cognitive level (figure 1), but the instantiation of this thematic frame makes divergences occur in predicate frames at the lexical level. Descriptive capacity of electronic dictionaries is rather limited, even with grammatical features, and sometimes not powerful enough to account for the use of lexical units. In the case of verbs, this constraint can be overcome if you gain access to thematic frames underlying predicate frames and relate cognitive schemata to headwords’ grammatical behaviour.

4 Conclusion

In spite of the well-known advantages of the electronic medium, its potential for lexicography has not been significantly exploited yet, since most electronic dictionaries are still based on printed dictionaries; indeed, electronic dictionaries even try to reproduce the external image of their paper counterparts. Consequently, all shortcomings revealed in printed dictionaries are found in electronic ones. Main problems faced by modern electronic lexicography are summarized as follows:

²⁵ The difference between thematic frame and predicate frame is partly influenced by the distinction between argument roles and participant roles in the Construction Grammar (Goldberg 1995), where the first are related to the construction and the latter to the frame of a particular verb.

- Dictionaries cannot adapt themselves to users' profiles, so lexicographical resources should be more adaptive.
- Dictionaries disregard users' production needs, so lexicographical resources should be more complete.
- Dictionaries make use of primitive querying systems, so lexicographical resources should be more efficient.

As an attempt to remedy all these defects, our proposal is intended to reuse an NLP knowledge base as a human-oriented electronic dictionary. To this respect, the main strengths of developing a web service which can employ FunGramKB are the adaptability of lexicographical data to users' profiles, the completeness and accuracy of grammatical information and the provision of an intelligent search engine.

References

- Bensoussan, Marsha (1983): "Dictionaries and tests of EFL reading comprehension". *ELT Journal* 37 (4), 341-345.
- Bosque Muñoz, Ignacio (1999): "El nombre común". In: Bosque Muñoz, Ignacio and Demonte Barreto, Violeta (eds.) *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa Calpe, 3-75.
- Castillo Carballo, M. Auxiliadora and García Platero, Juan Manuel (2003): "La lexicografía didáctica". In: Medina Guerra, Antonia M. (ed.) *Lexicografía Española*. Barcelona: Ariel, 333-351.
- De Schryver, Gilles-Maurice (2003): "Lexicographers' dreams in the electronic-dictionary age". *International Journal of Lexicography* 16 (2), 143-199.
- Dorr, Bonnie J. (1994): "Machine translation divergences: a formal description and proposed solution". *Computational Linguistics* 20 (4), 597-633.
- Downing, Angela and Locke, Philip (1992): *A University Course in English Grammar*. London: Prentice Hall.
- EAGLES Lexicon Interest Group (1993): *EAGLES: Computational Lexicons Methodology Task*. EAGLES Document EAG-CLWG-METHOD/B.
[\[http://www.ilc.cnr.it/EAGLES96/method/method.html\]](http://www.ilc.cnr.it/EAGLES96/method/method.html)
- EAGLES Lexicon Interest Group (1996a): *EAGLES: Synopsis and Comparison of Morphosyntactic Phenomena Encoded in Lexicons and Corpora. A Common Proposal and Applications to European Languages*. EAGLES Document EAG-CLWG-MORPHSYN/R.
[\[http://www.ilc.cnr.it/EAGLES96/morphsyn/morphsyn.html\]](http://www.ilc.cnr.it/EAGLES96/morphsyn/morphsyn.html)
- EAGLES Lexicon Interest Group (1996b): *EAGLES: Preliminary Recommendations on Subcategorisation*. EAGLES Document EAG-CLWG-SYNLEX/P.
[\[http://www.ilc.cnr.it/EAGLES96/20synlex/synlex.html\]](http://www.ilc.cnr.it/EAGLES96/20synlex/synlex.html)
- EAGLES Lexicon Interest Group (1999): *EAGLES LE3-4244: Preliminary Recommendations on Lexical Semantic Encoding*. Final Report.
[\[http://www.ilc.cnr.it/EAGLES96/EAGLESLE.PDF\]](http://www.ilc.cnr.it/EAGLES96/EAGLESLE.PDF)
- Floridi, Luciano (1999): *Philosophy and Computing: An Introduction*. London-New York: Routledge.
- Gelpí Arroyo, Cristina (2003): "El estado actual de la lexicografía: los nuevos diccionarios". In: Medina Guerra, Antonia M. (ed.) *Lexicografía Española*. Barcelona: Ariel, 307-328.
- Goldberg, Adele E. (1995): *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Hernández, Humberto (2000): "Últimas tendencias en la lexicografía didáctica". In: Ruhstaller, Stefan and Prado Aragonés, Josefina (eds.) *Tendencias en la Investigación Lexicográfica del Español: El Diccionario como Objeto de Estudio*

- Lingüístico y Didáctico*. Huelva: Universidad de Huelva, 101-116.
- Koren, Shira (1997): "Quality versus convenience: comparison of modern dictionaries from the researcher's, teacher's and learner's points of view". *TESL-EJ* 2 (3). [\[http://tesl-ej.org/ej07/a2.html\]](http://tesl-ej.org/ej07/a2.html)
- Laufer, Batia and Hadar, Linor (1997): "Assessing the effectiveness of monolingual, bilingual, and 'bilingualized' dictionaries in the comprehension and production of new words". *The Modern Language Journal* 81 (2), 189-196.
- Laufer, Batia and Kimmel, Michal (1997): "Bilingualized dictionaries: how learners really use them". *System* 25 (3), 361-369.
- Laufer, Batia and Melamed, Linor (1994): "Monolingual, bilingual and 'bilingualised' dictionaries: which are more effective, for what and for whom?" In: Martin, W., Meijs, W., Moerland, M., ten Pas, E., van Sterkenburg, P. and Vossen, P. (eds.) *The 6th Euralex International Congress on Lexicography*. Amsterdam. 565-576.
- Lenci, Alessandro (2000): "Building an ontology for the lexicon: semantic types and word meaning". *Workshop on Ontology-Based Interpretation of Noun Phrases*, Kolding-Denmark.
- Lieske, Christian, McCormick, Susan and Thurmair, Gregor (2001): "The Open Lexicon Interchange Format (OLIF) comes of age". *Machine Translation Summit VIII: Machine Translation in the Information Age*. Santiago de Compostela. 211-216.
- Lyons, John (1968): *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- McCormick, Susan (2002): *The Structure and Content of the Body of an OLIF v.2.0/2.1*. The OLIF2 Consortium. [\[http://www.olif.net/documents/NewOLIFstruct&content.pdf\]](http://www.olif.net/documents/NewOLIFstruct&content.pdf)
- McCormick, Susan, Lieske, Christian and Culum, Alexander (2004): *OLIF v.2: A Flexible Language Data Standard*. The OLIF2 Consortium. [\[http://www.olif.net/documents/OLIF_Term_Journal.pdf\]](http://www.olif.net/documents/OLIF_Term_Journal.pdf)
- Moreno Fernández, Francisco (2000): "Diccionarios para el aprendizaje de lenguas extranjeras". In: Ruhstaller, Stefan and Prado Aragonés, Josefina (eds.) *Tendencias en la Investigación Lexicográfica del Español: El Diccionario como Objeto de Estudio Lingüístico y Didáctico*. Huelva: Universidad de Huelva, 151-170.
- Nesi, Hilary (1998): "Dictionaries on computer: how different markets have created different products". *Symposium on Language Learning and Computers*, Chemnitz University of Technology. [\[http://www.tu-chemnitz.de/phil/english/chairs/linguist/real/independent/llc/Conference1998/Papers/Nesi.htm\]](http://www.tu-chemnitz.de/phil/english/chairs/linguist/real/independent/llc/Conference1998/Papers/Nesi.htm)
- Perry, Brian C. (1997): "Electronic learners' dictionaries (ELDs): an overview of recent developments". *CALL Electronic Journal* 1 (2). [\[http://www.jaltcall.org/cjo/5_98/call_EJ/Perry.html\]](http://www.jaltcall.org/cjo/5_98/call_EJ/Perry.html)
- Piotrowski, Tadeusz (1989): "Monolingual and bilingual dictionaries: fundamental differences". In: Tickoo, Makhan L. (ed.) *Learners' Dictionaries: State of the Art*. SEAMEO Regional Language Centre, Singapore, 72-83.
- Porto Dapena, José Álvaro (2002): *Manual de Técnica Lexicográfica*. Madrid: Arco.
- Roberts, Roda P. (1997): "Using dictionaries efficiently". *38th Annual Conference of the American Translators Association*. Washington: American Translators Association, 283-296.
- Robertson, John S. and Turley, Jeffrey S. (2003): "A Peircean analysis of the American-Spanish clitic pronoun system". *Semiotica* 145 (1), 21-70.
- Stein, Gabriele (1989): "Recent development in EFL dictionaries". In: Tickoo, Makhan L. (ed.) *Learners' Dictionaries: State of the Art*. SEAMEO Regional Language Centre, Singapore, 10-41.
- Stein, Gabriele (1990): "From the bilingual to the monolingual dictionary". In: Magay, Tamás and Zigány, Judit (eds.) *BudaLEX '88*. Budapest: Akadémiai Kiadó, 401-407.
- Underwood, Nancy and Navarretta, Costanza (1997): "Towards a standard for the creation of lexica". Center for Sprogteknologi, Copenhagen.

Walz, J. (1990): "The dictionary as a secondary source in language learning". *The French Review* 64 (1), 79-94.