2. THE CONCEPTS OF ‘COLLOCATION’ AND ‘SELECTION RESTRICTION’ WITHIN FUNGRAMKB

FunGramKB is made up of two information levels (Periñán and Mairal, 2009; fc):

1. The lexical level = linguistic knowledge

2. The conceptual level = non-linguistic knowledge

1. The lexical level
   a. The lexicon stores information about lexical units, preserving the major linguistic assumptions of RRG - logical structures, macroroles, etc.
   b. The morphicon handles cases of inflectional morphology.

2. The conceptual level
   a. The ontology or the world model
   b. The cognicon: where procedural information is kept.
   c. The onomasticon: where information about instances of entities and events is stored.
2.1. COLLOCATIONS

- Firth (1957)
- Koike (2001):
- In FunGramKB, collocations are understood in a broad sense to refer to those combinations of lexemes that commonly and frequently co-occur in a language, including both grammatical and lexical collocations.

2.2. SELECTION RESTRICTIONS

- ‘Traditional’ selection restrictions are better known as ‘selectional preferences’ in FunGramKB.
- This view on selectional preferences is totally consistent with the approach taken by Coseriu (1967), McCawley (1968), Fillmore (1970), and Bosque (2004), to name just a few.
- Selection restrictions provide non-linguistic information, since the information expressed through features like ‘human’, ‘animal’, etc., has no relation whatsoever with our knowledge of languages like English, Spanish or Japanese, but with ‘the real world’ and our experiences there.

2.2. SELECTION RESTRICTIONS

- Unlike the restrictive treatment given by Generative Grammar, they are understood not as semantic requirements on the nature of the arguments a predicate subcategorizes for, but as conceptual constraints prototypically related to cognitive situations.
- They are not word-oriented, so their place in FunGramKB is the conceptual level, specifically, the ontology.
- Example: the concept ‘EAT’.

3. FUNGRAMKB SELECTIONAL PREFERENCES: THE DOMAIN OF POSSESSION

- According to Jackendoff (1992: 79), apud Faber & Mairal (1996: 264), POSSESSION is:
  an artificial relationship established between two entities, one of whom has the right or authority to use the other as he wishes and has the right or authority to control anyone else’s use of the other, and to impose sanctions for uses other than those he permits.
Thematic frames (TF) and meaning postulates (MP) provide the semantic properties used to characterize the basic and terminal concepts that populate the ontology.

Both TFs and MPs employ concepts to formally describe meaning.

They are language-independent conceptual schemata, not lexical representations.

3.1. Selectional Preferences in Basic Concepts

Selectional preferences of the concept +WEAR_00: +HUMAN_00, +PET_00, +GARMENT_00, +ORNAMENT_00, +BODY_AREA_00 and +ON_00.

They are situated in the TFs and MPs of the ontology because it is there that they can exert constraints typically related to the cognitive situation displayed by the events.
3.2. SELECTIONAL PREFERENCES IN TERMINAL CONCEPTS

- A terminal concept, preceded by symbol $, can only be encoded when there is a conceptual restriction on the meaning of a basic concept.
- Selectional preferences then allow us to codify the distinguishing parameters that differentiate them.
- Terminal concepts within POSSESSION: $ABOUND_00, $GRASP_00 and $SPORT_00.
3.3. SUBCONCEPTS

- There are cases in which the conceptual restriction or specification takes place exclusively in one or all of the participants of the TF of a basic or terminal concept, without varying the MPs.
- These are known as SUBCONCEPTS in FunGramKB and appear preceded by a minus symbol and in capital letters.

3.3. SUBCONCEPTS

- Subconcepts are not really visible in the ontology.
- In other words, they don’t ‘hang’ in the hierarchical organization of concepts because they are conceptual specifications of one of the participants of an already existing concept.

3.3. SUBCONCEPTS

- Within the domain of POSSESSION:
  a. -WIELD: a conceptual specification of the terminal concept $GRASP_00 and lexicalized as wield, carry, bear and empuñar.
  b. -MISPLACE: linked to the basic concept +LOSE_00 and lexicalized in Spanish as traspapelar (lit. ‘misplace a paper’).
  c. -SAVE: associated to the basic concept +STORE_00, which English and Spanish express as save and ahorrar.
  d. -TAKE: a specification of the basic concept +WEAR_00 and expressed in Spanish with the verb ‘calzar’ (‘wear shoes or boots’).
3.4. ELABORATION OF SELECTIONAL PREFERENCES

**English data:**
- MULTIWORDNET
- WORDREFERENCE
- COLLINS THESAURUS
- WOXICON
- LONGMAN
- CAMBRIDGE
- BBI, LTP, OCD
- The Corpus Concordance and Collocation Sampler from The Collins Wordbanks Online English corpus.

**Spanish data:**
- MARÍA MOLINER
- CASARES
- CLAVE
- DRAE
- REDES
- ADESSE
- CREA

a) **Look up** every single word belonging in POSSESSION in the English and Spanish resources mentioned above;

b) **Note down** all the lexical information given for their selection restrictions, collocations, words that typically occur as subjects, objects, etc.;
3.4. ELABORATION OF SELECTIONAL PREFERENCES

c) Look for abstract labels or ‘umbrella’ patterns that could work for every word linked to a particular concept and in every language we are working with, → abstracting away from specific words;

d) Find the appropriate concepts to codify them: +HUMAN_00, +GARMENT_00 ...

3.5. WHERE ARE COLLOCATIONS?

Collocations are word-oriented so they are stored in their appropriate lexica, depending on the language they are connected to.

EXAMPLES: atesorar and hoard, which lexicalize the basic concept +STORE_00 in the ontology.
4. CONCLUSIONS

- Advantages:
  A. By posing two information levels, that is, the ontology and the different lexicons, RRG semantic representations can be deeply enriched, including all types of information that go well beyond those aspects of meaning with an impact on syntax: selection preferences.

- This theoretical move is done at a very low cost, because the ontology is based on a hierarchical inference system, which means that information can be placed in and retrieved from all the different ontological properties: TFs, MPs, subconcepts, etc.
  - “redundancy is minimized while informativeness is maximized” (Periñán and Mairal, 2009)

C. Since ontological concepts are universal, in principle every single language could be implemented in FunGramKB.

Thank you for your attention!!!

5. References


5.1. Dictionaries, data bases and corpora.


COLLINS THESAURUS: Available online from http://dictionary.reverso.net/ (June-July 2009).


Rocio Jiménez Briones, Ph. D.
Universidad Autónoma de Madrid.
rocio.jimenez@uam.es
http://www.lexicom.es/drupal/

*** Financial support for this research has been provided by the Spanish Ministry of Science and Innovation, grants no. HUM2007-65755/FILO and FFI2008-05035-C02-01.