ABSTRACT
In Artificial Intelligence, the main problem in the successful development of natural language understanding systems lies on the lack of an extensive common-sense knowledge base. Common sense is mainly made up of semantic and procedural knowledge, which FunGramKB stores in the form of meaning postulates and cognitive macrostructures respectively. The objective of this paper is to describe the reasoner running on the ontology with the aim of minimizing redundancy as well as maximizing informativeness in the semantic knowledge repository of FunGramKB.

KEY WORDS
Natural language understanding, common sense, knowledge representation, meaning postulate, and reasoning.

1. Introduction
FunGramKB is a user-friendly environment for the semiautomatic construction of a multipurpose lexico-conceptual knowledge base for a natural language processing (NLP) system within the theoretical model of S.C. Dik’s Functional Grammar [1]. Efficient NLP systems require a knowledge base solidly based on a linguistic theory, so that the system can identify regularities in meaning extensions, capture syntactic and semantic generalizations from lexical clusters, and establish criteria to structure and interpret new data. However, FunGramKB is not a literal implementation of Dik’s lexical database, but we depart from the functional model in some important aspects with the aim of building a more robust knowledge base. FunGramKB is made up of five independent but interrelated components: lexicon, morphicon, onomasticon, cognicon and ontology. Lexical units are assigned syntactic, pragmatic and collocational information in the lexicon, but their meaning representations are conceived as conceptual properties in the ontology, so that every sense of a lexical unit is linked to a conceptual unit. The morphicon helps our system to handle cases of inflectional morphology. Names of entities, such as cities, products, etc., are stored in the onomasticon. Finally, the cognicon stores procedural knowledge by means of cognitive macrostructures, i.e. script-like schemata in which a sequence of stereotypical actions is structured on the basis of temporal continuity.

In this paper, we describe the reasoner running on the meaning postulates in the ontology with the aim of maximizing informativeness as well as minimizing redundancy.

2. Meaning Postulates in FunGramKB
Velardi et alii [2] distinguish two well-defined strategies when describing meaning in computational lexicography: the cognitive content in a lexical unit can be described by means of semantic features or primitives (conceptual meaning), or through associations with other lexical units in the lexicon (relational meaning). Strictly speaking, the latter doesn’t give a real definition of the lexical unit, but it describes its usage in the language via ‘meaning relations’ with other lexical units. It is certainly easier to state associations among lexical units in the way of meaning relations than describing the cognitive content of lexical units formally, but the inference power of conceptual meaning is much stronger. Surface semantics can be sufficient in some NLP systems, but the construction of a robust knowledge base guarantees its use in most NLP tasks, consolidating thus the concept of resource reuse.

When meaning postulates are built for an NLP knowledge base in order to represent an adult speaker’s linguistic competence, dictionaries must be our guide, since they are reliable repositories of information which has been judged to be relevant for lexical meaning by several generations of expert speakers [3]. Indeed, Ide and Véronis [4] recommend using several dictionaries as sources of lexical acquisition, because what a particular dictionary lacks is usually supplied by another dictionary. Machine-readable dictionaries used in the acquisition of our knowledge base are Collins COBUILD English Dictionary [5], Oxford Advanced Learner’s Dictionary [6] and Longman Web Dictionary [7].

In FunGramKB, a meaning postulate is a set of one or more logically connected predications, which are cognitive constructs carrying the generic features of the
concept. To illustrate, some predications in the meaning postulates of an entity (a), event (b) and quality (c) are presented:

a. BIRD
+(e_1; BE (x_1; BIRD)_{Theme} (x_2; VERTEBRATE)_{Referent})
*(e_2; HAVE (x_1)_{Theme} (x_3; m; FEATHER & \&; 2; LEG & \&; 2; WING)_{Referent})
*(e_3; FLY (x_1)_{Theme})

b. KISS
+(e_1; TOUCH (x_1; PERSON)_{Agent} (x_2)_{Theme} (f_1; 2; LIP)_{Instrument} (f_2; LOVE (x_1)_{Agent} (x_2)_{Theme}) \& (e_2; GREET (x_1)_{Agent} (x_2)_{Theme})_{Reason})

c. HUGE
+(e_1; BE (x_1)_{Theme} (x_2; HUGE)_{Attribute})
+(e_2; BE (x_1)_{Theme} (x_3; SIZE)_{Referent})
+(e_3; BE (x_2)_{Theme} (x_4; \&; BIG)_{Attribute})

For example, predications in (a) have the following natural language equivalents:

Birds are always vertebrates.
A typical bird has many feathers, two legs and two wings.
A typical bird flies.

Dik [1] proposes using lexical units from the own language when describing meaning postulates, since meaning definition is an internal issue of the language. However, this strategy contributes to lexical ambiguity in representation due to the polysemic nature of the defining lexical units. In addition, describing the meaning of lexical units in terms of other lexical units leads to some linguistic dependency [9]. Instead, FunGramKB employs conceptual units for the formal description of meaning postulates, resulting in an interlanguage representation of meaning.

3. Monotonic Reasoning and FunGramKB

An NLP application is actually a knowledge-based system, so it must be provided with a knowledge base and a reasoning engine. Researchers in Artificial Intelligence have proved that standard logic is not able to manage common-sense reasoning. Monotonic inheritance is not able to deal with intrinsic properties of middle-level concepts in the ontology, because this type of inheritance doesn’t admit exceptions to inherited default values. Therefore, non-monotonicity is a key issue in human reasoning, because it permits the withdrawal of conclusions which are true just for the typical members of a particular class.

Non-monotonic logics is an umbrella term for a family of formalisms based on default reasoning, where the system can override previous beliefs in the light of further information. Although non-monotonic logics has been widely developed in the last 25 years, resulting in formalisms such as circumscription [10], default logic [11] and autoepistemic logic [12], most of these models involve a high computational cost when being implemented in working applications which require knowledge representation. On the contrary, and despite its less-expressive power, the simplicity of defeasible logic [13] makes it one of the most efficient non-monotonic reasoning model for NLP [14, 15]. Defeasible reasoning allows the possibility of working with incomplete information, so a closed-world assumption cannot be applied.

There are usually three types of rules in a defeasible theory [13]: strict rules, defeasible rules and defeaters. Strict rules are law-like rules, which have no exceptions: e.g. whales are mammals, circles are round. On the other hand, defeasible rules can be defeated by contrary evidence: e.g. birds typically fly. Finally, defeaters are used to block some defeasible rules in order to prevent some conclusions. For example, a rule such as "if an animal is heavy, then it may not be able to fly" may override the conclusion drawn from the defeasible rule 'birds typically fly'. The superiority relation, in which a superior rule may override an inferior one, can be expressed by means of rules (e.g. r_2 > r_1, in case rules can be labelled) or a superiority operator.

In FunGramKB, each predications in a meaning postulate is preceded by a reasoning operator in order to state if the predication is strict (+) or defeasible (*). Our inference engine handles predications as rules, allowing monotonic reasoning with strict predications, and non-monotonic with defeasible predications. The superiority relation is not explicitly stated in the predicates, but the priority principle is applied at the different levels of the Microconceptual-Knowledge Spreading to resolve conflicts between predications. In other words, if two defeasible predications which meet during the spreading of the meaning postulate turn out to be contradictory, our system removes that predication being at the highest level of the spreading process. A more accurate account of this process is described in the following section.

4. Spreading Meaning Postulates

Lexical meaning is like an iceberg - only a small amount is visible from the surface, so a word is associated to much more semantic information which is really shown in its meaning postulate [16]. In FunGramKB, all this underlying cognitive information is revealed through a

---

1 Periñán-Pascual and Arcas-Túnez [8] describe the formal grammar of well-formed predications for meaning postulates in FunGramKB.
2 For the sake of clarity, the names of conceptual units have been oversimplified.
process called MicroKnowing (Microconceptual-Knowledge Spreading), which takes place in the ontological component of our system. This multi-level process is performed by means of two types of reasoning mechanisms: inheritance and inference. Our inheritance mechanism strictly involves the transfer of one or several predications from a superordinate concept to a subordinate one in the ontology. On the other hand, our inference mechanism is based on the structures shared between predications linked to conceptual units which do not take part in the same subsumption relation within the ontology.

The MicroKnowing process can be formally stated as follows:

$$\Omega_i \Rightarrow \Phi_0 \cup \Phi_1 \cup ... \Phi_n \Rightarrow \bigcup_{i=0}^{n} \Phi_i$$

$$\Phi_i \Rightarrow \Phi_i \cup \Phi_2 \cup ... \Phi_k \Rightarrow \bigcup_{j=1}^{k} \Phi_j$$

The MicroKnowing of a particular concept, which is called 'nuclear concept', is originated in the meaning postulate linked to that conceptual unit. Given that $\Phi$ represents one particular inherited or inferred predication at a particular spreading level, $\Phi_i$ is defined as the set of predications being inherited or inferred at spreading level $i$ providing that $i \geq 1$, since $\Phi_0$ represents the nuclear meaning postulate. On the other hand, $\Omega_i$ represents the extended meaning postulate which has just been spread at level $i$ providing that $i \geq 1$, since $\Omega_i$ is equivalent to $\Phi_0$. Assuming that $n$ is the total number of levels which are necessary to spread the meaning postulate completely, where $n \leq (2d - 1)$ being $d$ the length of the path from the nuclear concept to the father metaconcept, $\Omega_n$ represents the final output of a totally-extended meaning postulate.

To illustrate the MicroKnowing, let's take the concept TURKEY, whose meaning postulate is the following one:

$$+(e_i; \ BE (x_i; \ TURKEY)_{\text{Theme}} (x_2; \ POUlTRY)_{\text{Referent}})$$

$*(e_2; \ BE (x_1)_{\text{Theme}} (x_2; \ BIG)_{\text{Attribute}})$

$*(e_3; \ a \ HAVE (x_3)_{\text{Theme}} (x_4; \ FEATHER)_{\text{Referent}} (f_1; \ HEAD \ & \ NECK)_{\text{Location}})$

At the first spreading level, the inference mechanism is triggered, where the nuclear concept (i.e. TURKEY) infers a predication from the meaning postulate of another concept providing that the nuclear concept acts as the selection preference in any constituent of the meaning postulate of the target concept. For example, the following predication is inferred from the meaning postulate of WATTLE:

$$*(e_2; \ HAVE (x_3; \ TURKEY)_{\text{Theme}} (x_1; \ WATTLE)_{\text{Referent}})$$

Now the inferred predication is incorporated into the nuclear meaning postulate, resulting in a first version of the extended meaning postulate at the first level. Therefore:

$$\Omega_i \Rightarrow \Omega_i \cup \Phi_i$$

FunGramKB automatically readjusts indices for variables $e$ (predication), $x$ (argument) and $f$ (satellite) in inferred or inherited predications.

However, our inference model ignores those predications in which the genus of the target concept is the nuclear concept. FunGramKB's reasoner recognises these cases because the nuclear concept appears as a selection preference of the Referent argument in the first predication of the meaning postulate. This is the case of GOBBLER:

$$*(e_1; \ BE (x_1; \ GOBBLER)_{\text{Theme}} (x_2; \ TURKEY)_{\text{Referent}})$$

On the other hand, the inheritance mechanism is triggered at the second spreading level. Now the extended meaning postulate inherits those predications in the meaning postulate linked to the immediate superordinate of the nuclear concept. In our case, the father of TURKEY is POULTRY, whose meaning postulate is as follows:

$$+(e_1; \ BE (x_1; \ POULTRY)_{\text{Theme}} (x_2; \ BIG)_{\text{Referent}})$$

$*(e_2; \ BE (x_1)_{\text{Theme}} (x_2; \ DOMESTIC)_{\text{Attribute}})$

$*(e_3; \ KEEP (x_3; \ PERSON)_{\text{Theme}} (x_4)_{\text{Referent}} (f_1; \ FARM)_{\text{Location}})$

$*(e_4; \ OBTAIN (x_5; \ PERSON)_{\text{Theme}} (x_6; \ EGG, MEAT)_{\text{Theme}} (f_1; x_1)_{\text{Origin}})$

From this point on, both inference (I) and inheritance (H) mechanisms are cyclically applied to any genus in any meaning postulate within the conceptual path between the nuclear concept and its metaconcept. In figure 1, circles represent conceptual units, whose meaning postulates take an active part in the reasoning task:

*Figure 1. Inference and Inheritance in FunGramKB's Ontology.*
The following table presents some of the output in the MicroKnowing for the nuclear concept TURKEY, where levels 1 and 4 happen with inference, and levels 2, 3 and 5 with inheritance:

<table>
<thead>
<tr>
<th>Level</th>
<th>Meaning Postulate</th>
<th>Activator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Φ₀</td>
<td>+(e₁; BE (x₁; TURKEY)_theme (x₂; POULTRY)_referent)</td>
<td>TURKEY</td>
</tr>
<tr>
<td></td>
<td>*(e₂; (x₁)_theme (x₂; BIG)_attribute)</td>
<td>TURKEY</td>
</tr>
<tr>
<td></td>
<td>*(e₃; HAVE (x₁)_theme (x₄; FEATHER)_referent) (f₁; HEAD &amp; NECK)_location)</td>
<td>TURKEY</td>
</tr>
<tr>
<td>Φ₁</td>
<td>+(e₄; HAVE (x₁)_theme (x₅; WATTLE)_referent)</td>
<td>WATTLE</td>
</tr>
<tr>
<td></td>
<td>+(e₅; GOBBLE (x₁)_theme)</td>
<td>GOBBLE</td>
</tr>
<tr>
<td></td>
<td>+(e₆; OBTAIN (x₆; PERSON) agent (x₇; TURKEY)_theme (f₂; x₈) origin)</td>
<td>TURKEY_01</td>
</tr>
<tr>
<td>Φ₂</td>
<td>+(e₇; BE (x₁)_theme (x₉; BIRD)_referent)</td>
<td>POULTRY</td>
</tr>
<tr>
<td></td>
<td>*(e₈; BE (x₁)_theme (x₁₀; DOMESTIC)_attribute)</td>
<td>POULTRY</td>
</tr>
<tr>
<td></td>
<td>*(e₉; KEEP (x₁₀; PERSON) theme (x₁₁) referent (f₁; FARM) location)</td>
<td>POULTRY</td>
</tr>
<tr>
<td>Φ₃</td>
<td>+(e₁₀; BE (x₁)_theme (x₁₂; VERTEBRATE)_referent)</td>
<td>BIRD</td>
</tr>
<tr>
<td></td>
<td>*(e₁₁; HAVE (x₁)_theme (x₁₃; FEATHER)_referent)</td>
<td>BIRD</td>
</tr>
<tr>
<td></td>
<td>*(e₁₂; HAVE (x₁)_theme (x₁₄; LEG)_referent)</td>
<td>BIRD</td>
</tr>
<tr>
<td></td>
<td>*(e₁₃; HAVE (x₁)_theme (x₁₅; WING)_referent)</td>
<td>BIRD</td>
</tr>
<tr>
<td></td>
<td>*(e₁₄; LAY (x₁)agent (x₁₅; EGG)theme)</td>
<td>BIRD</td>
</tr>
<tr>
<td></td>
<td>*(e₁₅; FLY (x₁)theme (f₄; WING)instrument)</td>
<td>BIRD ∪ WING</td>
</tr>
<tr>
<td>Φ₄</td>
<td>*(e₁₆; CHIRP (x₁)theme)</td>
<td>CHIRP</td>
</tr>
<tr>
<td></td>
<td>*(e₁₇; HAVE (x₁)theme (x₁₈; CLAW)referent (f₅; TOE)location)</td>
<td>CLAW</td>
</tr>
<tr>
<td></td>
<td>*(e₁₈; PECK (x₁)agent (x₁₉; THING)theme)</td>
<td>PECK</td>
</tr>
<tr>
<td></td>
<td>*(e₁₉; HAVE (x₁)theme (x₂₀; BEAK)referent)</td>
<td>BEAK</td>
</tr>
<tr>
<td></td>
<td>*(e₂₀; PREEN (x₁)theme)</td>
<td>PREEN</td>
</tr>
<tr>
<td>Φ₅</td>
<td>+(e₂₁; BE (x₁)theme (x₂₁; ANIMAL)referent)</td>
<td>VERTEBRATE</td>
</tr>
<tr>
<td></td>
<td>+(e₂₂; CONTAIN (x₁)theme (x₂₀; SKELETON)referent)</td>
<td>VERTEBRATE</td>
</tr>
<tr>
<td></td>
<td>+(e₂₃; CONTAIN (x₁)theme (x₂₁; BACKBONE)referent)</td>
<td>VERTEBRATE</td>
</tr>
<tr>
<td></td>
<td>*(e₂₄; HAVE (x₁)theme (x₂₂; TAIL)referent (f₁; BACK)location)</td>
<td>VERTEBRATE</td>
</tr>
</tbody>
</table>

A key issue in defeasible reasoning is how to process the conflict that arises with competing rules. In FunGramKB, conflicts are presented between contradictory predications. When our system needs to check if predication φ is compatible with the predications within meaning postulate Ω, our reasoning engine uses both strict and defeasible predications. If φ is a strict predication, then it is automatically incorporated into Ω. On the contrary, if φ is a defeasible predication, then compatibility with predications in Ω must be validated; in other words, predication φ is compatible with predications in Ω if and only if Ω does not contain a predication ¬φ. In order to implement the resolution method of incompatible predications, it is necessary to apply the 'stepwise conceptual decomposition' (SCD) of all predications involved in the MicroKnowing.

The SCD is founded on the 'stepwise lexical decomposition' principle proposed by Dik [1]. This principle establishes a way of interrelating lexical entries where the definitens in a meaning postulate can be converted into the definiendum of another meaning postulate. The fact that the functional model is provided with a lexical decomposition process enables the construction of meaning postulates in a simple fashion, as well as minimizing information redundancy. In FunGramKB, the SCD is similar to Dik’s principle, apart from the fact that the main building blocks of our meaning representations are not lexical units but concepts. Therefore, the SCD is defined as the process in which conceptual units in a predication are substituted by their respective meaning postulates until a meaning representation composed of basic concepts is reached. When the SCD is applied in the MicroKnowing, the nuclear concept is the only one which does not undergo this SCD process; otherwise, a large number of predications would be repeated. The SCD of a predication φ which is incorporated into an extended meaning postulate Ω is essential in order to check consistency of φ with Ω, that is, if there is no ¬φ in Ω. The compatibility problem can be efficiently tackled just in case two predications are compared with the same conceptual depth, and more particularly at the root level. For example, the extended meaning postulate of TURKEY has predications [a] and [b] among others and we want to know if predication [c] can be inferred from AVIARY so that conclusion [d] can be reached:

[a] +(e₁; BE (x₁; TURKEY)_theme (x₂; BIRD)_referent)  
[b] *(e₂; BE (x₁; TURKEY)_theme (x₂; BIG)_attribute)  
[c] *(e₃; KEEP (x₉; PERSON)_theme (x₁; BIRD)_referent (f₁; AVIARY)_location) (e₁₅; BE (x₁)_theme (x₂₀; SMALL)_attribute)  
[d] *(e₁₆; KEEP (x₁₉; PERSON)_theme (x₁; TURKEY)_referent (f₁; AVIARY)_location)  

Predications [a], [b] and [c] are compatible at this level of decomposition, but after the SCD the system recognises that (b) is in conflict with (c) because of predication [e] in
the meaning postulate of SMALL, so predications \([c]\) is rejected:

\[
\begin{align*}
\text{SMALL} \\
+\langle e_1: \text{BE (x}_2)\text{Theme (x}_1: \text{SMALL})\text{Attribute}\rangle \\
+\langle e_2: \text{BE (x}_3)\text{Theme (x}_1: \text{SIZE})\text{Referent}\rangle \\
+\langle e_3: \text{BE (x}_4)\text{Theme (x}_1: \text{BIG})\text{Attribute}\rangle
\end{align*}
\]

The most relevant advantage of this reasoner for an NLP knowledge base is to avoid unnecessary duplication of information, as well as the possibility of updating the knowledge base in a consistent way. When the language engineer modifies an existing meaning postulate or builds a new one, just before being stored, FunGramKB automatically performs the MicroKnowing for that meaning postulate in order to check the compatibility of the newly-incorporated predications with other predications involved in the reasoning process. The language engineer is informed about all those inferred or inherited predications which are not true for the nuclear concept. In addition, FunGramKB displays the whole MicroKnowing process step by step, enabling us to verify inference and inheritance conditions in a transparent way.

5. Computational Implementation of Meaning Postulates

Language engineers enter meaning postulates through a dedicated editor in FunGramKB's ontology interface. The construction of our knowledge base is semi-automatic in the sense that human intervention is required but the engineer's intuition is guided and reviewed through a series of user-friendly tools for lexico-conceptual acquisition. For example, when creating or modifying meaning postulates, a syntactic-semantic checker is activated, so that consistent well-formed meaning postulates can be stored. When a meaning postulate is stored, a parser outputs an XML-formatted feature-value structure used as the input for the reasoning engine, so that inheritance and inference mechanisms can be applied. Both the syntactic-semantic validator of meaning postulates and the XML parser were written in C#. XML was chosen as the language for knowledge representation in FunGramKB because it helps the system with the establishment of premises for structured data transfer, the autonomous separation of knowledge from its representation formalism, and the atomization of the various components the meaning postulate is made up of, speeding up in this way the inference process and the direct access to particular conceptual units in the meaning postulate. To illustrate, figure 2 presents the XML representation of some predications in the meaning postulate of concept SILVER:

\[
\begin{align*}
+\langle e_1: +\text{BE}_00 (x}_1: +\text{SILVER}_00)\text{Theme (x}_1: x_1)\text{Means}\rangle \\
\langle e_2: +\text{TRAVEL}_00 (x}_4: +\text{ELECTRICITY}_00)\text{Theme (f}_1: x_1)\text{Means}\rangle \\
\langle S\rangle
\end{align*}
\]

\[
\begin{align*}
\langle LBCv: +\text{BE}_00 \rangle \\
\langle x_1: +\text{Theme}\rangle \\
\langle LBCea: +\text{SILVER}_00 \rangle \\
\langle x_2: +\text{Referent}\rangle \\
\langle LBCea: +\text{METAL}_00 \rangle \\
\langle x_3: +\text{Attribute}\rangle \\
\langle LBCea: +\text{EXPENSIVE}_00 \rangle \\
\langle x_4: +\text{Means}\rangle \\
\langle f_1: x_1 \rangle
\end{align*}
\]

Figure 2. An XML Meaning Postulate.

6. Conclusion

FunGramKB is a workbench for the semi-automatic construction of a multipurpose knowledge base for NLP systems, mainly those requiring natural language understanding. In this paper we have presented the multi-level process of spreading those knowledge representations linked to ontological concepts. In particular, we have shown that the cyclical application of the inheritance and inference mechanisms on our meaning postulates allow FunGramKB to minimize redundancy as well as keeping our knowledge base as informative as possible.

Since lexical meaning is not enough to build human common-sense, FunGramKB is provided with general procedural knowledge through the cognicon. We are currently working on the MacroKnowing (Macroconceptual-Knowing Spreading), i.e. the process of integrating meaning postulates from the ontology with the cognitive macrostructures in the cognicon. This combination of semantic and procedural knowledge, so distinctive of human reasoning, is hardly found in NLP systems to date.

7. Acknowledgements

FunGramKB is part of the research carried out in the project "Implementación de una base de conocimiento
para el razonamiento del sentido común" funded by Universidad Católica San Antonio de Murcia, code number PMAFI-PI-11/1C015.

References:


