A computational account of illocutionary meaning

Rocío Jiménez-Briones
Universidad Autónoma de Madrid / Spain

Abstract – This paper looks at how illocutionary meaning could be accommodated in FunGramKB, a Natural Language Processing environment designed as a multipurpose lexico-conceptual knowledge base for natural language understanding applications. To this purpose, this study concentrates on the Grammaticon, which is the module that stores constructional schemata or machine-tractable representations of linguistic constructions. In particular, the aim of this paper is to discuss how illocutionary constructions such as Can You Forgive Me (XPREP)? have been translated into the metalanguage employed in FunGramKB, namely Conceptual Representation Language (COREL).

The formalization of illocutionary constructions presented here builds on previous constructionist approaches, especially on those developed within the usage-based constructionist model known as the Lexical Constructional Model (Ruiz de Mendoza 2013). To illustrate our analysis, we shall focus on the speech act of CONDOLING, which is computationally handled through two related constructional domains, each of which subsumes several illocutionary configurations under one COREL schema.

Keywords – illocutionary meaning, NLP, FunGramKB, LCM, Condoling

1. INTRODUCTION

FunGramKB is a Natural Language Processing (NLP) environment which has been designed as a multipurpose lexico-conceptual knowledge base to be implemented in natural language understanding applications (Periñán Pascual 2013; Periñán Pascual and Arcas Túnez 2014, among others). It is made up of a number of different modules which account for linguistic information (the lexical and grammatical levels) and non-linguistic knowledge (i.e. conceptual information in the form of an Ontology, an Onomasticon and a Cognicon).

In this context, the scope of this work is limited to the grammatical level or Grammaticon, which stores constructional schemata, i.e. machine-tractable representations of linguistic constructions. The Grammaticon comprises four Constructicons, which are the computational implementation of the four descriptive layers of the Lexical Constructional Model (LCM): argumental, implicational, illocutionary and discursive. Implicational constructions (e.g. Can You X?; Panther and Thornburg 1998; Stefanowitsch 2003; Pérez 2013) are housed in the Level-3 Constructicon of FunGramKB, which is thus the focus of our attention here. Thus, the aim of this paper is to discuss how, in the Level-3 Constructicon, illocutionary constructions have been translated into the metalanguage employed in FunGramKB, namely COREL. Since, as any metalanguage, COREL imposes its own semantics and syntax (Periñán Pascual and Mairal Usón 2010), this article details the challenges faced in order to provide a computational account of illocutionary meaning in FunGramKB in an NLP environment.

1 Financial support for this research has been provided by the Spanish Ministry of Economy and Competitiveness, grants no. FFI2013- 43593-P and no. FFI2014-53788-C3-1-P.
The structure of this paper is as follows. Section 2 provides a brief description of FunGramKB and its architecture, paying special attention to the grammatical level or Grammaticon, which is the component that stores constructional schemata. The four constructional levels of the LCM which inspire the Grammaticon are dealt with in Section 3. In particular, the constructions that are the focus of our attention, that is, illocutionary configurations, are defined and exemplified here. Section 4 constitutes the core of the study. Section 4.1 illustrates the formalization in the Level-3 Construction of the speech act of CONDOLING, which we approach as consisting of two constructional domains: several illocutionary expressions that are grouped under different, yet related, COREL schemas. Section 4.2 discusses the challenges faced when dealing with configurations of this kind within a computational environment like FunGramKB. Finally, section 5 offers some concluding remarks.

2. FunGramKB: A BRIEF INTRODUCTION

FunGramKB is defined as “a user-friendly online environment for the semiautomatic construction of a multipurpose lexico-conceptual knowledge base for NLP systems, and more particularly for natural language understanding” (Periñán Pascual and Arcas Túnez 2010: 2667). Grounded in theoretical linguistics, specifically Role and Reference Grammar (RRG; Van Valin 2005) and the Lexical Constructional Model (LCM; Ruiz de Mendoza and Mairal Usón 2008, 2011; Mairal Usón and Ruiz de Mendoza 2009; Ruiz de Mendoza 2013), FunGramKB comprises a number of different modules which account for linguistic information (the lexical and grammatical levels) and non-linguistic knowledge (i.e. conceptual information in the form of an Ontology, an Onomasticon and a Cognicon). Figure 1 shows the architecture of this knowledge base:

![Figure 1: The architecture of FunGramKB (http://www.fungramkb.com/)](http://www.fungramkb.com/)

For our purposes here, we are only concerned with the Grammaticon of English, that is, the component which stores constructional schemata in order to help RRG to construct the syntax-to-semantics linking algorithm (Periñán Pascual 2013: 209). Such schemas are defined as machine-tractable representations of linguistic constructions, which, as Goldberg (2006: 5) puts forward, are form-function pairings that exist at all levels of grammatical analysis. As illustrated in Figure 2, the Grammaticon of English is composed of four Constructicons, each of which is the computational implementation of the four descriptive layers of the LCM detailed in Section 3.
3. ILLOCUTIONARY MEANING WITHIN THE LCM

The LCM, which is defined by Ruiz de Mendoza (2013: 232) as “a comprehensive model of meaning construction through language in context”, puts forward four levels or layers of analysis (see Ruiz de Mendoza and Galera 2014):

- Level 1, at which lexical structure is integrated into argument-structure constructions like the ditransitive, resultative, caused-motion, etc. (Goldberg 1995).
- Level 2 deals with implicational structure. Oft-quoted constructions within this level are *Who’s Been XVP-ing Y?* or *What’s X doing Y?* (cf. Kay and Fillmore 1999).
- Level 3 is concerned with illocutionary structure constructions like *Can You X?, You Shall Have X, Let’s X* (cf. Panther and Thornburg 1998).
- Level 4 addresses discourse structure constructions (e.g. *X Let Alone Y or Just Because X Doesn’t Mean Y*; cf. Fillmore et al. 1988).

Besides displaying a multilayered architecture in which constructions of varying size and complexity are accounted for, each of these descriptive levels is based on a cognitive model type (cf. Lakoff 1987). Illocutionary meaning, for example, is based on high-level situational cognitive models, which, in line with Ruiz de Mendoza and Baicchi (2007), can be identified by means of illocutionary scenarios (cf. Panther and Thornburg 1998), that is, generalizations over everyday situations where people offer, forgive, apologize, etc. For instance, the declarative sentence *It is hot in here* may stand as a request in the context of a request scenario based on the cultural convention according to which, if people manifest that they are negatively affected by a specific situation, other people are expected to help. Thus, illocutionary constructions are regarded within the LCM as “(sets of) grammatical resources that are capable of (jointly) activating relevant parts of an illocutionary scenario in connection to a context of situation (which may activate other parts of the scenario in a complementary fashion). The direct consequence is the production of indirect speech acts with different degrees of explicitness” (Ruiz de Mendoza and Baicchi 2007: 108). Consequently, Level-3 configurations are based on generic or high-level models like “requesting”, “ordering”, “boasting”, etc. They consist of fixed and variable elements (cf. Ruiz de Mendoza and Mairal Usón 2008; Mairal Usón and Ruiz de Mendoza 2009; Pérez 2013). For example, in a construction such as *Can You X?* (e.g. *Can you pass me the salt?*), “Can you” would be the fixed element and “X” (i.e. “pass me the salt”) would be the variable element.

By way of illustration, let us take a look at the semantic structure of ORDERS, whose generic structure generalizes over multiple everyday cases of social interaction in which we try to get something done by other people. In line with Del Campo and Ruiz de Mendoza (2012: 19), the generic structure of ORDERS could be described as follows:

A person (A) has authority over someone else (B). A wants B to do something. Through an utterance, A makes B aware of his/her wish. B is under the obligation to act as commanded. B is expected to act as commanded.
This generic structure can be realized by means of specific constructions such as those in (1)-(4):2

1. I want you to leave me alone! (COCA, 2011)
2. Shut up! (COCA, 2012)
3. You will clean the house for me. (Google Books Corpus, 2008)
4. Jones, I order you to clean your boots. (Google Books Corpus, 2007)

There are also other, less direct ways of ordering. For example, interrogative sentences, whose openness clashes with the imposition conveyed by orders. Consider the examples in (5) and (6):

5. Can You X? (e.g. Can you shut up for a minute, Sidra? (Google Books Corpus, 2012))
6. Why Don’t You X? (e.g. Why don’t you just be quiet for a while? (Google Books Corpus, 2006))

The structure Can You X? is a conventionalized form of request. However, its request meaning in (5) has been overridden through inference, thus resulting in an order. Similarly, the pattern Why Don’t You XVP? in (6) conventionally conveys a suggestion. This meaning can, nevertheless, be overridden through inference in a context in which the speaker is evidently irritated. In such a case, the speaker is not likely to be making a suggestion and, thus, the addressee has to select a different interpretation. Here, the order interpretation of the construction presupposes that the addressee is behaving improperly and not acting as the speaker wants him to.

After this brief description of illocutionary constructions in the LCM, we return now to how these configurations are approached within the computational environment of FunGramKB or its Level-3 Constructicon.

4. ILOCUTIONARY MEANING IN FUNGRAMKB: THE LEVEL-3 CONSTRUCTICON

As explained in Sections 2 and 3, the English Grammaticon of FunGramKB is the module in which computationally implemented representations of constructions or constructional schemata are stored. Since, as following the LCM, there exist four layers of analysis to account for meaning (i.e. the argument, implicational, illocutionary and discourse levels), the Grammaticon consists of four Constructicons, each of which is in charge of implementing each type of construction. Figure 3 provides the interface of the English Level-3 Constructicon.

---

2 In line with the usage-based approach to linguistic analysis (cf. Langacker 1987), the examples in this paper have been extracted from systematic searches within the Corpus of Contemporary American English (Davies 2008-) and the Google Books Corpus (Davies 2011-). The former contains over 520 million words from a large variety of texts, both spoken and written. The latter allows searches of more than 200 billion words in the American and British English datasets, as well as the One Million Books and Fiction datasets.
The first type of information included in the Level-3 Constructicon, at the left-hand side of Figure 3, is the list of constructional domains identified for English. In the description box, a short definition of the constructional domain, in this case THANKING, is provided. As for the box labeled “Realizations”, it includes a list of all the constructions that syntactically realize the generic structure THANKING (e.g. *I am grateful for NP, Let me thank you for NP, I appreciate NP*). Finally, at the bottom of the editor, we can find the schema by means of which the semantics of each constructional domain is codified through COREL, that is, the formal language that FunGramKB uses to record the information in all its modules. In other words, the COREL schema accommodates the definition of the construction in the metalanguage that the machine can understand. In the case at hand, the COREL schema for THANKING reads as follows: i) the hearer (x1) does something (x2) good (x3) for the speaker (x4); ii) the speaker thanks the hearer for what the speaker has done. A further explanation on the nuances of COREL and on how the different types of constructional domains were arrived at is supplied in 4.1.

4.1. Codifying speech acts through COREL: the case of CONDOLING

Drawing on Del Campo (2012), the following twelve speech acts were taken as the point of departure to account for illocutionary meaning in FunGramKB:

(7) advising, apologizing, boasting, condoling, congratulating, offering, ordering, pardoning, promising, requesting, thanking, threatening

In most cases, it was possible to define each construction on its own. In other cases, due to the nature of COREL, which will be detailed later on, we were forced to group various constructions under the same COREL schema. Each grouping is what has been called “constructional domain”. To date, thirty-four constructional domains have been identified and formalized in the Level-3 Constructicon, following the two methodological principles in (8):

(8) a. The identification of key distinctive semantic features within each speech act, supported by a relevant number of constructions;
b. The possibility of codifying these distinctive features through COREL.

A word is needed here in relation to COREL. COREL, which stands for *Conceptual Representation Language*, is the language of representation employed to formally describe the knowledge stored in each of the modules that make up FunGramKB. This implies that not only are the concepts stored in the Ontology, the entities in the Onomasticon and the scripts in the Cognicon defined using this metalanguage, but also the constructions kept in the four Constructions. In
other words, if we want the machine to understand what the resultative construction or the speech act of ORDERING are, we need to define them by means of COREL.

As any other type of language, COREL has its own semantics and syntax (cf. Periñán Pascual and Mairal Usón 2010). Semantically speaking, COREL employs a number of conceptual units, variables, reasoning operators, aspectual operators, logical operators, etc., to construct meaning. As for its syntax, it follows a particular system of notation and formalization that needs to be obeyed to produce well-formed structures (see Jiménez-Briones and Luzondo-Oyón 2011 for further details).

Although the advantages of employing a metalanguage are unquestionable, it goes without saying that the expressive power of a formal language is quite limited and, in some cases, does not allow to express the same semantic nuances as a natural language does. However, we believe that, in this study, a quite productive compromise has been achieved between the demands of each system. Therefore, each type of illocutionary domain or dimension identified (i.e. the thirty-four constructional domains) consists of several illocutionary constructions grouped under different, yet related, COREL schemata.

To illustrate this laborious process of codifying illocutionary constructions through COREL, let us consider the act of CONDOLING, whose semantic structure could be characterized as:

It is manifest to a person (A) that another person (B) is involved in a negative state of affairs. A is unable to change the state of affairs to B’s benefit.

According to Del Campo (2012), the following illocutionary constructions capture this meaning:

<table>
<thead>
<tr>
<th>Constructions</th>
<th>Semantic structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) I XVP My Condolences (XPREP); My Condolences (XPREP); I XVP My Sympathy (XPREP); My Sympathy (XPREP)</td>
<td>A makes his feeling of sympathy manifest to B</td>
</tr>
<tr>
<td>(2) I Am Sorry (XP)</td>
<td>A feels sympathy about B’s misfortune and makes it manifest to B</td>
</tr>
<tr>
<td>(3) Accept My Condolences; Let Me XVP My Condolences</td>
<td>A makes his feeling of sympathy manifest to B and B may accept it</td>
</tr>
<tr>
<td>(4) Accept My Sympathy; Let Me XVP My Sympathy</td>
<td>A feels sympathy about B’s misfortune and makes it manifest to B. B may accept A’s sympathy</td>
</tr>
</tbody>
</table>

Table 1: Illocutionary constructions for CONDOLING

In line with the LCM, this author distinguishes: (a) constructions that make explicit A’s sympathy to B (those at (1) in Table 1); (b) constructions in which A also feels sympathy about B’s misfortune (e.g. the *I Am Sorry (XP)* construction); (c) configurations in which B may accept or not A’s sympathy (those at (3) in Table 1); and, finally, (d) those constructions that express all the features already mentioned (e.g. (4) in Table 1). However, when dealing with this type of information within FunGramKB, only two key distinctive features, well supported by a relevant number of constructions (cf. the first methodological principle in (8a)), have been taken into account: (i) the fact that A explicitly expresses his/her feeling of sympathy to B and (ii) whether B accepts it or not. Accordingly, only two constructional domains for CONDOLING have been formalized in COREL since, in our view, these two domains express the necessary and sufficient information that the machine will need to process natural language. Their COREL schemata are displayed in Tables 2 and 3:
The Condoling-type-1 domain shown in Table 2, realized by the illocutionary expressions previously seen in (1) and (2) in Table 1, codifies the information that we believe every speaker has in his/her mind when generalizing about an everyday situation or illocutionary scenario such as condoling (cf. Panther and Thornburg 1998). Hence, the first proposition or “e1” expresses that the speaker knows that the hearer is sad because a relative or a friend has died. The second proposition (e4) codifies the fact that the speaker cannot change this sad situation. Finally, the third proposition (e5) captures the idea that the speaker tells the hearer s/he feels pity for this misfortune.

As for the Condoling-type-2 domain in Table 3, the same three propositions already commented on form part of its COREL schema. However, a fourth one needs to be included to formalize the hearer’s acceptance of the speaker’s pity. It is worth noting that this lack of obligation on the speaker’s side to accept the speaker’s sympathy is formalized by means of the reasoning operator for defeasible predications “*”, whereas the rest of predications are preceded by the strict operator “+” (cf. Periñán Pascual and Mairal Usón 2010). The difference between them is of paramount importance here: whereas the defeasible operator expresses features that can be overridden in the light of contradictory information, predications preceded by the strict operator contain features that cannot be refuted under any circumstance. Thus, marking the fourth proposition or “e7” with the star symbol allows us to express in a machine-tractable way the fact that the hearer may accept or not the speaker’s pity: *(e7: +AGREE_00 (x3)Theme (x6)Referent (x1)Goal).

4.2. Codifying speech acts through COREL: some challenges

As already described in 4.1, COREL imposes its own semantics and syntax, which has made us face a number of challenges when providing a computational account of illocutionary constructions in FunGramKB. Some of them are detailed below so that they can help linguists or knowledge engineers working in the Level-3 Constructicon supply more accurate formalizations of constructional domains.

To begin with, it must be pointed out that COREL does not allow the expression of comparative meaning. This semantic feature is especially relevant in some illocutionary constructions for the act of BOASTING. For example, the domain Boasting–type2 is defined as “the speaker says that he feels proud because he is better than others at doing something”. Undoubtedly, the distinctive feature of this illocution is the fact that the speaker thinks that s/he is better than the hearer. Our proposal to solve this situation consists in including the quantification operator “m”, which stands for “many”), before the quality +GOOD_00, as can be seen in the COREL schema in (9). We admit, however, that the semantic content is not the same.
The second issue we had to tackle in almost all the thirty-four codified constructional domains is the optionality of complements in the illocutionary expressions. For instance, the realization *I XVP My Condolences (XPREP) of Condoling-type 1* can omit the PP-complement or not, as (10) illustrates:

(10) a. I’m calling to express my condolences on the death of your brother (COCA, 2005)

b. I’m calling to express my condolences.

Since optionality is not considered explicitly in COREL, two realizations have been included in the appropriate box: one with the PP-complement and another one without it. Figure 4 depicts optionality.

![Figure 4: Optionality of complements](image)

The third challenge faced, and which still remains unsolved, is related to the lack of syntagmatic labels for imperative clauses, embedded gerund clauses, embedded *to*-infinitival clauses and *that*-clauses. Even though the inventory of grammatical categories employed in FunGramKB is finished for English (NP, AP, PP, VP, etc.), the set of syntagmatic categories still needs to be designed within the framework of RRG, since this is the theoretical model upon which ARTEMIS is being built.³ Note, however, that such labels are indeed urgent in order to codify illocutionary expressions belonging to constructional domains such as Requesting-type-2 (11) or Promising-type-1 (12):

(11) *Do/Would You Mind XVP?* (e.g. Do you mind taking a picture of us? (COCA, 2003))

(12) *I Assure You XP* (e.g. I assure you that I will do a good job (COCA, 2003))

5. CONCLUSION

This paper presents some preliminary steps towards the computational treatment of illocutionary constructions within FunGramKB. To the best of our knowledge, no such approach has been undertaken by other computational systems for NLP. We believe that this study has been possible due to the fruitful collaboration between theoretical linguistics and NLP systems, in the form of the LCM and FunGramKB, respectively. Thus, the LCM has provided us with a detailed, thorough description and analysis of the different illocutionary constructions, whereas the notation and formalization employed by FunGramKB has given us the opportunity to reuse this linguistic analysis to enhance the knowledge base.

Such a move, however, demands sacrifices on both sides and compromises need to be reached between the fine-grained descriptions in linguistics and the real needs in NLP. By way of illustration, according to the LCM, the desirable situation would have been to codify each construction (e.g. *Can I Thank You For VP, I Am Sorry, Can You XVP*) through a separate COREL schema, that is, to treat each construction on its own right (Goldberg 2006). For FunGramKB, however, the codification of the twelve speech acts (see (7)) through only twelve COREL schemata would have probably been quite accurate.

In this paper we have grouped various illocutionary constructions under the same COREL schema, thus resulting in what we have labeled “constructional domains/dimensions” (e.g. advising-type-1, advising-type-2, etc.). In our view, the thirty-four constructional domains identified and codified so far are rich enough to express the necessary and

³ ARTEMIS (*Automatically Representing Text Meaning via an Interlingua-based System*) is a proof-of-concept prototype which generates the semantic representation of a text automatically under the format of a Conceptual Logical Structure (CLS). Periñán Pascual and Arcas Túnez (2014) detail the implementation of this parser using RRG and FunGramKB.
sufficient information that the machine will need to process natural language, without betraying the linguistic analysis that supports them.

REFERENCES


**Corresponding author**
Departamento de Filología Inglesa. 203-Módulo VI-bis
Facultad de Filosofía y Letras. Ciudad Universitaria de Cantoblanco
C/ Francisco Tomás y Valiente, 1
28049 Madrid
e-mail: rocio.jimenez@uam.es

received: March 2016
accepted: November 2016