

# PRAGMATEXT: Annotating the C-ORAL-ROM Corpus with Pragmatic Knowledge<sup>1</sup>

---

Ana González-Ledesma<sup>2</sup>

## Abstract

This paper outlines the first phase of the PRAGMATEXT project. The aim of PRAGMATEXT is to introduce pragmatic knowledge into the transcriptions of C-ORAL-ROM, a spontaneous spoken corpus of Spanish. The paper is divided in four sections. The first section presents the most relevant features of the C-ORAL-ROM corpus. The second describes the pragmatic-discursive annotation model. The phenomena tagged are: emotional discourse, argumentative operations, modalization operations, evidentiality, phraseological units with metaphoric meaning and speech acts in interrogative clauses. The third section, resolves the three challenges related to the implementation of such annotation model to the XML language: (1) pragmatic-discursive operations are expressed at different grammatical levels (lexicon, prosody, syntax, etc.); (2) a linguistic unit can have as attributes different types of pragmatic information; (3) the pragmatic knowledge is not expressed by a closed word class. The fourth section discuss future work and mentions some uses of a corpus tagged with pragmatic knowledge such as in the field of man-machine conversational systems and teaching of Spanish as a foreign language.

## 0. Introduction

Pragmatics is the science of the use of language in its context. In the second half of the 20<sup>th</sup> century, linguists and philosophers started to study the importance of context in the interpretation of statements. There have been two areas of progress in this field: the study of spoken language and the conception of the verbal communication as inferential process. A speaker only codifies the most relevant part of what they actually want to express. At the same time, hearer does not decode but rather interpret the statements expressed by the speaker. As a result, some linguists believe that it is impossible to define the meaning of a statement *a priori*, since this meaning is the result of a negotiation process carried out by the parties during the interaction. The dependency of the meaning to the context is main reason why it is so difficult to objectify, systematize and implement pragmatic knowledge.

The first step is to translate information from implicit to explicit. Many research groups have started to introduce pragmatic knowledge in its natural languages processing systems. To understand language, it is more important to recognize the speech acts of a verbal interaction than the grammatical categories of their constituent words. This is true particularly for machines or L2 students who don't know how to use the language. For this reason the most important initiatives in this area have been developed in order to improve machine-human interaction and second language learning. Part of the initiatives focused on annotating corpus with

---

<sup>1</sup>This research has been supported by the Spanish Ministry of Education and Science under the grant TIN2004-07588-C03-02 and a research scholarship supported by *Comunidad Autónoma de Madrid*.

<sup>2</sup>Laboratorio de Lingüística Informática, Universidad Autónoma de Madrid  
*e-mail*: ana.ledesma@maria.llif.uam.es

pragmatic information. The phenomena being tagged are: co-reference relations, discourse structure, speech acts, emotional language and discourse markers. The tagset design normally depends on the system necessities. These efforts present three limitations for the current study: the corpora used are domain-specific; relevant pragmatic phenomena in statement interpretation are not represented (e.g. discursive modalization); these researches have not been generally developed for Spanish spoken language.

This paper presents PRAGMATEXT, a pragmatic annotation model designed for C-ORAL-ROM, a Spanish spoken language corpus of 300000 words representing a wide range of communicative situations. The paper is divided in four sections. The first section presents the most relevant features of the C-ORAL-ROM corpus. The second describes the pragmatic-discursive annotation model. The phenomena tagged are: emotional discourse, argumentative operations, modalization operations, evidentiality, phraseological units with metaphoric meaning and speech acts in interrogative clauses. In developing the model, *discourse particles* have been the object of our study. The third section resolves the three challenges related to the implementation of such annotation model to the XML language: (1) pragmatic-discursive operations are expressed at different grammatical levels (lexicon, prosody, syntax, etc.); (2) a linguistic unit can have as attributes different types of pragmatic information; (3) the pragmatic knowledge is not expressed by a closed word class. The fourth section discusses the uses of a corpus tagged with pragmatic knowledge in the field of man-machine conversational systems and teaching of Spanish as a foreign language.

## **1. C-ORAL-ROM, a spontaneous spoken corpus**

The corpus used for the present study is developed within the framework of the European project C-Oral-Rom. The main goal of this project was to build four corpora in four romance languages (Italian, Spanish, French and Portuguese) with similar design features: the same number of words, the same types of communicative situations and the same transcription format.

C-ORAL-ROM is made up of 300000 words covering a wide range of communicative topics, genres and communicative situations: from colloquial interactions to interactions highly influenced by the written modality, such as the mass media or formal contexts. The corpus design is represented in tables 1.1 and 1.2 below.

**Table 1:** Distribution of informal part of C-ORAL-ROM

**Table 2:** Distribution of formal part of C-ORAL-ROM

This variety of modalities and communicative genres provides a great chance to validate this annotation model. The Spanish C-ORAL-ROM corpus is annotated in XML with categorical and semantic information based on eventive semantics.

## **2. PRAGMATEXT, a pragmatic annotation model**

All phenomena described here have a cognitive origin, hence they are universals. A cognitive phenomenon influences social interaction and language. Therefore, the verbal communication process can be analyzed from three closely related levels: cognitive social, pragmatic and linguistic.

At a cognitive social level, the theoretical framework described is based on the idea that human perception of reality is subjective: perception is the sum of what we perceive through our senses and our mental frameworks. Therefore language is a way of codification of perception and transmission of knowledge. However, at a pragmatic level, the transmission of knowledge could be limited by personal interests and different perspectives, even between members of the same community with a common socio-cultural background. Anyhow, human beings not only do an effort to argument their points of view, but also to reach an agreement about the perception of reality. After all, people aim to be understood in a way that their perception of events matches *reality*. Sharing information is the only chance to achieve this: we consider *reality* a constant negotiation process to define a common standard. When information is exchanged, the nature of the information affects the truth conditions of a statement. For example, it is important, sometime vital, to specify how we know something is true: we need to know the source of information to assess the truth of a statement.

At the linguistic level, each language offers its own grammatical resources to underline a cognitive phenomenon. In the Spanish language, these are expressed at different grammatical levels; e.g., reformulation is expressed at a prosodic (in C-ORAL-ROM transcribed as ['] and [///]) or lexical level (the discourse markers *o sea, es decir*). For this reason, tagsets must be conceived from semantics to grammatical categories. If we analyze, from a pragmatic perspective, a particular grammatical category, for example, the interjection, we run the risk of falling into unnecessary typologies and ambiguity problems. This is especially observed in categories such as interjections or adverbs which have always been problematic to traditional grammar. Hence, this first part of PRAGMATEXT project is focused on the words that modify the whole sentence: interjection, conjunctions, discourse markers, sentence-adverbs, etc.

The three levels of under which a communication process are analyzed are summarized in the table 1.3 below

<b>COGNITION</b>	<b>PRAGMATICS</b>		<b>LINGUISTICS</b>	
<b>Cognitive and social strategies</b>	<b>Pragmatic phenomena</b>	<b>Pragmatic Categories</b>	<b>Grammatical category</b>	<b>Linguistic Forms</b>
How do we do reason?	Argumentative relations	Addition	Sentence adverbs Conjunctions Locutions Prepositional Phrases	Además ( <i>In addition</i> ) Pero ( <i>but</i> ) si ( <i>if</i> ) o sea ( <i>that is</i> ) etc.
		Adversativity		
		Hypothesis		
		Reformulation		
		etc.		
How do we adapt our point of view to the other hearer?	Discursive Modalization	Hedges	Sentence-adverbs Interjections Verbs Locutions Adjectives Prepositional Phrases Sentences	Bueno ( <i>well</i> )
		Boosters		Por supuesto ( <i>of course</i> )
		Interactives		Sabes? ( <i>you know?</i> )
Which type of knowledge source?	Evidentiality	In person	Prepositional Phrases Sentence Adverbs Sentences Etc.	Por lo visto ( <i>apparently</i> ) Según X ( <i>According to</i> ) En general ( <i>Generally</i> ) Mejor pájaro en mano, que ciento volando ( <i>a bird in the hand is worth two in the bush</i> )
		Heard		
		Other senses		
		Intellectual Inference		
		Other person		
		Written Knowledge		
		Oral Knowledge		
What emotions do we feel?	Emotional Discourse	Evaluative	Interjection Sentence Adverbs Exclamative sentences	Afortunadamente, ( <i>Fortunately</i> ) Desgraciadamente ( <i>Fortunately</i> ) Guay, qué bien, genial (great, cool) ¡Madre mía! Goodness me!
		Emotion		
How do we conventionalize language?	Phraseological Units	Locutions	Phrases	Tomar nota
		Collocations		a un tiro de
		Sayings Proverbs	Sentences	Piedra quien a hierro mata, a hierro muere
How do we conceptualize the reality?	Metaphor	Target domain	Body Space Time Discourse	En primer lugar (first) (Space = discourse)

		Source domain		
Which are our intentions?	Speech acts	Ask	Interrogative Utterances	Puedo abrir la ventana? <i>May open the window?</i>
		Request		
		Order		
		Etc.		

**Table 3:** The cognitive, pragmatic and linguistic levels

### 3. PRAGMATEXT TAGSET

The phenomena described here are:

1. Emotional discourse
2. Discourse relations
3. Modalization
4. Evidentiality
5. Phraseological Units
6. Metaphor
7. Speech Acts in interrogative clauses

Each phenomenon described is analyzed according to the methodology adopted to design of the tagset: description of Pragmatic phenomenon; classification of pragmatic categories; inventory of lexical linguistic forms in which a phenomenon is codified and formalization in XML language.

#### 3.1 Emotional discourse

The analysis of emotions has been addressed from different knowledge areas and with different theoretical frameworks. However, research related to emotions from a scientific perspective together with its systematization has been pursued only very recently.

In the linguistic domain, pragmatics has focused on the study of emotions through interjections. The most interesting contributions highlight the scope of these particles over the interpretation of the statements:

- (1) *Franco ha muerto ¡bien!* → Franco has died, great!
- (2) *Franco ha muerto ¡que pena!* → Franco has died, *what a pity!*
- (3) *¡ah! Franco ha muerto* → *Oh!* Franco has died

In the above examples, statements are interpreted in very different ways. The same fact (Franco's death) causes different reactions in hearer: happiness in (1), sorrow in (2) and surprise in (3).

There is no closed list of these emotions nor a complete description of the different uses of each interjection in Spanish, for a systematization of the meaning of these particles and a typology of the emotions they codify. On the contrary, the consulted bibliography insists on the fact that the meaning depends on the context in the end.

The PRAGMATEXT project established two types of interjections: a first group which presents a positive or negative evaluation of a particular situation; and a second group which expresses other emotions such as surprise or wish. When the interjection meaning is difficult to define, its context of use is described in a special tag.

Apart from the interjection, there are also other categories used to express emotions, such as sentence adverbs, exclamation sentences, and prepositional phrases.

This is a brief list of linguistic forms from the inventory produced: *hala, hostia, joder, de puta madre, genial, madre mía de mi vida y de mi corazón, Jesús bendito, hala, afortunadamente, desgraciadamente, superbien, fatal, de pena, etc.*

In XML language, this linguistic forms are tagged in the following way:

(1) <PI ED=evaluative/interjective type=positive/negative/other context=description>

An with an example:

(2) <PI ED=emotion type=other context=surprise>¡dios mío!</PI>  
<PI ED=emotion type=other context=surprise>¡oh my god!</PI>

### 3.2 Discourse relations

This section describes how the human being argues his perception of the world. The study of discursive relations has been often approached from the analysis of discourse markers. However, according to the theoretical principles mentioned in section 2, the model operates from the abstract plane to the concrete one. A typology of discourse relations has been designed and then codified in linguistic forms. The model does not use the discourse marker category due to the ambiguity of the tag, as they codify not only discourse relations but also other discursive phenomena such as modalization.

In the model, discourse relations are: reformulation, condition, change the focus, manner, contrast, result, purpose.

Conjunctions are the most important category used to express discourse relations, although we can find others as prepositional phrases, adverbs, noun phrases, etc. For example: *en primer lugar, aunque, si, pero, y, además, incluso, finalmente, porque, o sea, es decir, para, con el fin de, a pesar de, como consecuencia de, de tal modo que, primero, en cuanto a x, a este respecto, etc.*

The formalization in XML looks as following:

(1) <IP DR={reason/purpose/hypothesis/reformulation/etc.}>

An example of a linguistic forms with argumentative meaning :

(2) <IP DR=change\_focus>en cuanto a</IP>  
<IP DR=change\_focus>as regards</IP>

### 3.3 Modalization: hedges, booster and interactive

This group is made up of phenomena related to the social interaction, where two or more people share information and expect to be understood, and, at the same time, to reinforce their emotional bonds.

Through modalization, human beings express agreement o disagreement with the truth of the statements, of each other. Two types of modalization are

distinguished: hedges and boosters. Hedges are used for two reasons: when unsure of the truth of a statement or when the interlocutor is in disagreement and there is no intention to create conflict. Boosters are employed to reinforce engagement with the truth of a statements expressed or heard, but also to strengthen solidarity ties with the interlocutor.

Finally, the latest group of this section comprises the interactive forms. These linguistic forms operate to open or keep the communicative channel, offer options, guarantee that the message has been heard and ask the hearer's agreement.

Following, some example of each type is showed:

1. Hedges: *bueno, o sea, quizás, etc.*
2. Booster: *claro, evidentemente, por supuesto, seguramente, indudablemente, clarísimamente, etc.*
3. Interactives: *ah, eh? no? sabes? vale? ok? venga hola oye adiós buenos días, ...*

In XML format:

(1) <IP MOD=hedge/booster/interactive>

Example:

(2) <IP MOD=interactive>vale?<IP>  
<IP MOD=interactive>ok?<IP>

### 3.4 Evidentiality

The degree of trust about the truth conditions of a statement depends, to a great extent, on the origin of knowledge. This fact is marked in languages through what has been called evidentiality by linguists, and it is defined as the verbal codification of the knowledge source.

Unlike other languages, the evidentiality in Spanish is not expressed through a closed system of grammatical marks. Therefore, it is a difficult phenomenon to recognize. At present, although we have access to a great deal of information, sometime we are unable to assess if that information is reliable or not.. Evidentiality tagging may be a profitable way to face this problem; hence a typology must be established for it. The bibliography about this subject offers us some models; however, useful typology is the one where semi-automatic tagging is possible. It is critical to find a clear correspondence between the phenomena, its categories and the linguistic forms through which these phenomena are codified.

The model uses the following classification of knowledge:

- (1) experimented through senses
- (2) obtained through intellectual inference
- (3) 3<sup>rd</sup> party knowledge
- (4) written knowledge
- (5) oral or popular knowledge

The formalization in XML looks as follows:

(1) <IP Evidentiality={1/2/3/4/5}></IP>

Example:

(2) <IP Evidentiality=5>al parecer</IP>  
<IP Evidentiality=5>apparently</IP>

### 3.5 Linguistic convention: phraseological units and Metaphor

A second language student realizes that it is not enough to learn the grammar and vocabulary of a language, but also it is necessary to know how this community usually acts in a specific situation, to select the appropriate linguistic register or style of language, in short, the most common way to say something. Therefore, metaphors and phraseological units play a very important role in how the language becomes conventional.

The model distinguishes three types of phraseological units: collocations, locutions and proverbs. Both collocations and locutions are group of words which usually appear together but the first ones have a compositional meaning; the second ones not. On the other hand, proverbs are to be considered as a sentence, not a group of words.

Cognitive Linguistics highlighted that metaphorical language is a way to express our knowledge of *reality*. Through the metaphor, abstract concepts are expressed as concrete concepts, which are also culture dependant.

This phenomenon together with the phraseological units are the most typical part of a language. Because of this, the model uses tags for the metaphoric expressions included in the phraseological units. Tagging phraseological units, it must decide:

1. if it is a collocation, a locution or a proverb
2. if the expression contains a metaphor
3. the semantic domains present in the metaphor

The formalization in XML looks as follows:

(1) <IP PU={collocation/locution/proverb} MET=true Source={body, space, nature, etc.} Target={time, prudence, wisdom...}</IP>

Example:

(2) <IP PU=proverb MET=true Source=nature Target=prudence>más vale pájaro en mano que ciento volando</IP>  
<IP PU=proverb MET=true Source=nature Target=prudence>a bird in the hand is worth two in the bush</IP>

### 3.6 Interrogative speech acts

A speech act is a communicative unit which expresses the speaker's intention. There is no bi-univocal relation between the linguistic form of statements and its communicative intention. For this reason, many linguists think the automatic recognition of a speech act is impossible, because in this case the context, above all, power relationships. Until now, most initiatives have been developed for specific domain. At this stage, PRAGMATEXT is focused only on interrogative utterance to be studied in all domains of the CORAL ROM corpus.



The formalization in XML looks as follows:

(1) <IP SA={order, ask, rhetoric, request, ...}></IP>

For example:

(2) <IP SA=request>¿Puedo abrir la ventana?</IP>  
<IP SA=request> May I open the window, please?</IP>

#### 4. PRAGMATEX in XML

To formalize an annotation model in a Schema or DTD in XML language, there is a need to define what information will be introduced in the attribute labels and in the elements label. The application of the annotation model to the Schema is based on the principles below.

The preferred option has been to map the annotation model into the Schema, so that the Schema is linguistically justified. It is impossible to include pragmatic information at the element level due to the lack of hierarchy between the phenomena to be tagged: each phenomenon belongs to a different domain and there is not necessarily a dependency between all of them. For example, in a linguistic form such as: *a causa de (because of)*, we can find two types of pragmatic information: it can be viewed as a phraseological unit, but it also contains information about the discourse relation, more specifically, about reason. If this information would be incorporated as elements instead of as attributes, it could not be possible to systematize the relationship between phenomenon in a consistent way.

As an additional example, a discourse marker can be a phraseological unit but not all the phraseological units are discourse marker. For example, *correr el riesgo (to run the risk)*.

Finally a word could be encoding more than one pragmatic phenomenon simultaneously. For example, the word “por lo visto” (apparently) contain two types of pragmatic knowledge: Therefore, the pragmatic meaning of this particle in (1) is in XML format:

(1) <IP DR= MOD=hedge EVI=3>por lo visto</IP>  
<IP DR= MOD=hedge EVI=3>apparently</IP>

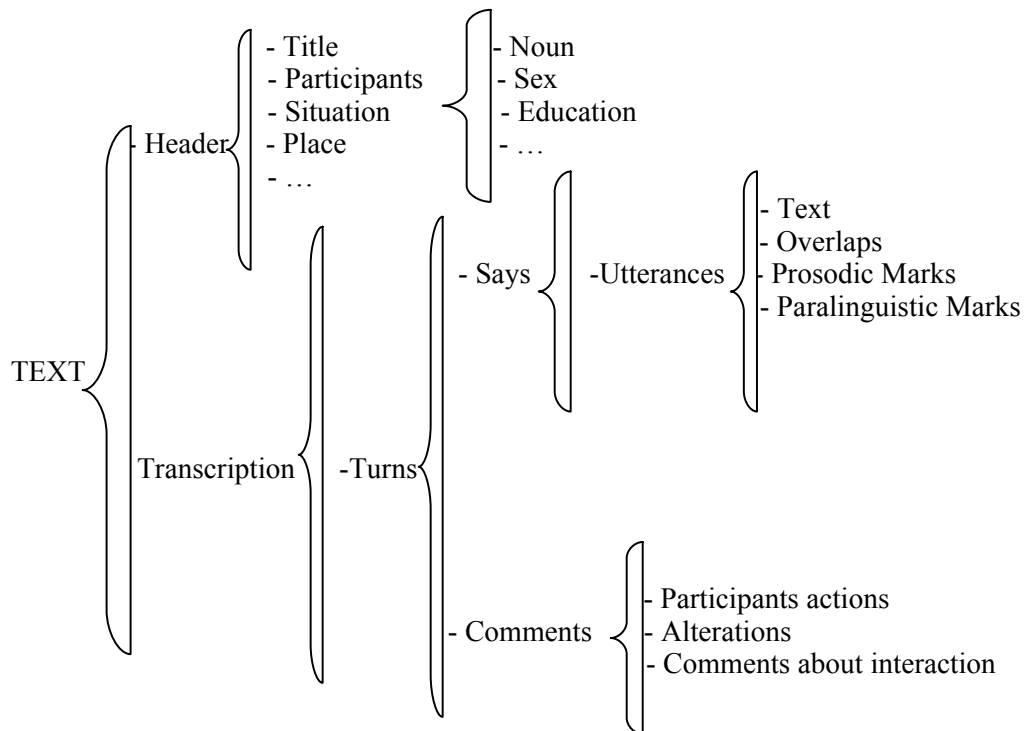
In this way, the relation between epistemic modality and evidentiality is reflected.

Clearly the only option is to map the pragmatic knowledge into the Schema for the XML language at attribute level.

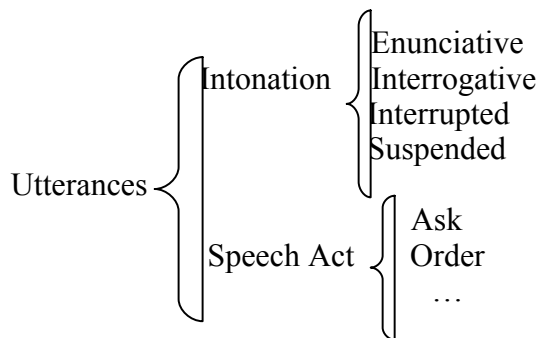
The element level of the Schema is utilized to define the text structure and the transcription format tags. Within the CORAL ROM corpus, the text information is coded in two parts: header and transcription. In the header, all contextual information about the communicative interaction is represented without hierarchy: personal data about participants, place, date, numbers of words, length of the conversation, communicative genre, etc. On the other hand, the transcription element is organized under a hierarchy: it is structured in turns, which are made of comments and interventions. Interventions are further differentiated in utterances, which contain as attributed information related to intonations (interrogative, interrupted, non-finished

and enuntiative) and speech act. Finally utterance can be text, transcription marks (overlaps, paralinguistic signs, etc) and/or pragmatic information.

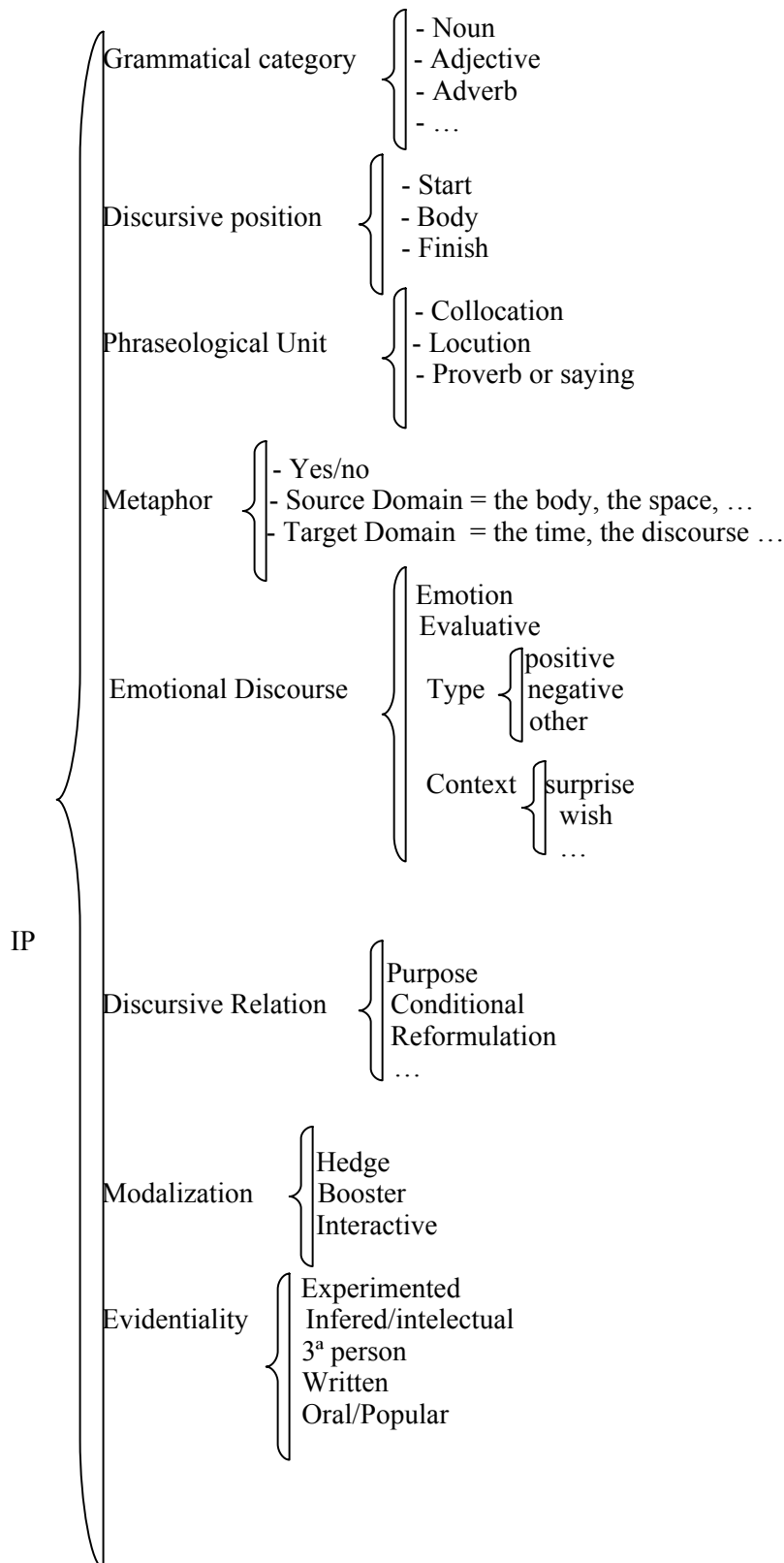
As follows, a schema of the text elements in XML language is presented:



Next, it shows the utterances attributes:



At the attribute level, all pragmatic phenomena are represented. In PRAGMATEXT, pragmatic information must be specified as follows:



The PRAGMATEXT scheme in XML language is detailed in **Appendix A**.

#### 4.1. An example in XML format

Appendix B shows an example of a telephonic conversation tagged in format. In this section, only a fragment of the whole interaction has been include. In it, the speaker makes a phonecall to his friend's, whose mother explain that his son is currently out.

(1) \*PAD: sí ?  
\*MIG: andeh / hola / está / Rubén ?  
\*PAD: andeh / no / no está //  
\*MIG: ¡ah! / no está ?  
\*PAD: no //  
\*MIG: y / sabes cuándo va a llegar ?  
\*PAD: pues / no tengo ni idea //

(2) \*PAD: hello ?  
\*MIG: andeh / hello / Rubén is there ?  
\*PAD: andeh / no / he isn't here //  
\*MIG: ¡ah! / isn't he ?  
\*PAD: no //  
\*MIG: and / do you know when he come back?  
\*PAD: so / I have any idea //

If we analyze this fragment using our annotation model, the following in bold information should be tagged:

(1) \*PAD: sí ?  
\*MIG: **andeh / hola / ¿está / Rubén ?**  
\*PAD: **andeh / no / no está //**  
\*MIG: **¡ah! / no está ?**  
\*PAD: no //  
\*MIG: **y / sabes cuándo va a llegar ?**  
\*PAD: **pues / no tengo ni idea //**

And the pragmatic information of the linguistic forms look as follows:

(2) sí? → Utterance=Interrogative Speech\_Act=asked Modalization=Interactive  
(3) hola → Modalization=Interactive  
(4) andeh → Discourse\_Relation=Support  
(5) ¿está Rubén? → Utterance=Interrogative Speech\_Act=Ask  
(6) ¡ah! → Modalization=Interactive  
(7) ¿no está? → Utterance=Interrogative Speech\_Act=Verify  
(8) y → Discourse\_Relation=Addition  
(9) sabes → Modalization=Interactive  
(10) pues → Discourse\_Relation=addition  
(11) no tengo ni idea → Modalization=hedge

Finally, this is the pragmatic analysis of turn “¡ah! ¿no está?” in XML format :

```
<Turn>
  <Says Speaker="MIG">
    <Utterance Prosody="Interrogative" Speech_Act="Verify">
      <PI GC="Interjection" DP="start" PU="no" MET="no" DR="no" ED="no"
        MOD="Interaction" EVI="no">¡ah!</PI>
      <Tone_Unit /> no está ?
    </Utterance>
  </Says>
</Turn>
```

In **Appendix C**, the whole conversation tagged in XML is included.

## 5. Conclusions and future work

In this paper, a pragmatic annotation model and its implementation in XML language is presented. Currently the spontaneous spoken corpus CORALROM has been tagged. This is a corpus with a wide range of communicative situations and a large number of linguistic registers. The pragmatic phenomena tagged are emotions, discourse relations, modalization, evidentiality, phraseological units, metaphor and speech acts in interrogative utterances. In the XML PRAGMATEXT implementation, the structure of the text is tagged at the element level, whereas the pragmatic information is introduced at the attribute level, so that more than one pragmatic phenomenon can be simultaneously presented for a single word. Additionally the sentence modifiers, such as conjunctions, interjections, sentence adverbs, etc., have been studied. A telephonic conversation has been manually tagged and it will be used to validate the performance of the XML annotation model. This example is included in the Appendix C.

The second part of the project involves semiautomatic tagging of the corpus, which is currently underway. The tagging strategies to recognize pragmatic information are based on context rules including prosodic and categorical information.

The final part of the project is focused on two of its possible applications: the improvement of man-machine conversational systems and teaching of Spanish as a foreign language.

## References

- ALBELDA, M (2002) La intensificación en el español coloquial, <http://www.tdx.cesca.es/TDX-0701105-125232/>
- Briz, A. (1995) La atenuación en la conversación coloquial. Una categoría pragmática, en Cortés, L. (ed.), El español coloquial. Actas del I Simposio sobre análisis del discurso oral. Almería: Servicio de Publicaciones, págs.: 103–122; ampliado en Briz. 1998, cap. 4 y 6.
- Corpas, L (1994) Manual de Fraseología española, Gredos, Madrid.
- Cresti, E. Et al. “The C-ORAL-ROM project. New methods for spoken language archives in a multilingual romance corpus”. Proceedings of LREC 2002. Las Palmas de Gran Canaria.
- Devillers, L., Vasilescu, I., and Lamel, L. (2002). Annotation and detection of emotion in a task oriented human-human dialog corpus, *International Standards for Language Engineering, Edinburgh*.

- Lakoff, G and Johnson (1998) M. *Metaphors we live by*, Chicago: Chicago University Press
- Marcos, M. (2002) El territorio de la evidencialidad en español *Actas V Congreso de Lingüística General*, 1858-1866.
- Maynard, C. and Leicher S. (2006) *Pragmatic Annotation of an Academic Spoken Corpus for Pedagogical Purposes* In E. Fitzpatrick (Ed.), *Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse*. Amsterdam: Rodopi B.V.
- Escandell, V (1993) *Introducción a la pragmática*, Madrid, Anthropos
- Sperber, D y Wilson, D (1994) *La relevancia .Comunicación y procesos cognitivos*, Madrid, Visor.
- Swift, M. D. Dzikovska M. O., Tetreault J. R., and Allen, J. F. (2004) Semi-automatic syntactic and semantic corpus annotation with a deep parser. In *LREC-2004*.
- Tetreault J., Swift M., Prithviraj P., Dzikovska M., and Allen J (2004) Discourse annotation in the Monroe corpus. In *ACL workshop on Discourse Annotation*

## Appendix A

### SCHEMA PRAGMATEXT

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified"
attributeFormDefault="unqualified">
  <xs:element name="TEXT">
    <xs:annotation>
      <xs:documentation>Comment describing your root element</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element name="Header" type="HEADER"/>
        <xs:element name="Transcription" type="TRANSCRIPTION"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <!--CABECERA-->
  <xs:complexType name="HEADER">
    <xs:sequence>
      <xs:element name="Title" type="xs:string"/>
      <xs:element name="File" type="xs:string"/>
      <xs:element name="Participants" type="PARTICIPANTS"/>
      <xs:element name="Date" type="xs:date"/>
      <xs:element name="Place" type="xs:string"/>
      <xs:element name="Situation" type="xs:string"/>
      <xs:element name="Topic" type="xs:string"/>
      <xs:element name="Source" type="xs:NMTOKEN" fixed="C-ORAL-ROM"/>
      <xs:element name="Class" type="CLASS"/>
      <xs:element name="Length" type="xs:string"/>
      <xs:element name="Words" type="xs:positiveInteger"/>
      <xs:element name="Acoustic_quality">
        <xs:simpleType>
          <xs:restriction base="xs:string">
            <xs:enumeration value="A"/>
            <xs:enumeration value="B"/>
            <xs:enumeration value="C"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:element>
      <xs:element name="Transcribers" type="TRANSCRIBERS"/>
      <xs:element name="Revisors" type="REVISORS"/>
      <xs:element name="Comments">
        <xs:complexType>
          <xs:complexContent>
            <xs:extension base="xs:anyType"/>
          </xs:complexContent>
        </xs:complexType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="PARTICIPANTS">
    <xs:sequence>
      <xs:element name="Participant" type="PARTICIPANT" maxOccurs="10"/>
    </xs:sequence>
  </xs:complexType>
  <xs:complexType name="PARTICIPANT">
    <xs:sequence>
      <xs:element name="ID">
        <xs:simpleType>
          <xs:restriction base="xs:string">
            <xs:pattern value="[A-Z]{3}"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:element>
      <xs:element name="Name" type="xs:NMTOKEN"/>
      <xs:element name="Sex">
        <xs:simpleType>
          <xs:restriction base="xs:string">
            <xs:enumeration value="woman"/>
            <xs:enumeration value="man"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
</xs:schema>
```

```

    </xs:simpleType>
  </xs:element>
  <xs:element name="Age">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="A"/>
        <xs:enumeration value="B"/>
        <xs:enumeration value="C"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:element>
  <xs:element name="Education">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="1"/>
        <xs:enumeration value="2"/>
        <xs:enumeration value="3"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:element>
  <xs:element name="Occupation" type="xs:string"/>
  <xs:element name="Role" type="xs:string"/>
  <xs:element name="Origin" type="ORIGIN"/>
  <!--aquí podríamos poner x o interviewed y interviewer, porque en realidad no aparece nada más-->
  <!--referencia al identificador del hablante, la segunda va a estar en el texto, en la DTD de C-ORAL-ROM
aparece como speaker en el texto y en la cabecera, en este mismo punto, aparece como shortname-->
  </xs:sequence>
</xs:complexType>
<xs:complexType name="ORIGIN">
  <xs:sequence>
    <xs:element name="Place" type="xs:NMTOKEN"/>
    <xs:element name="Comment" type="xs:string" minOccurs="0"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="CLASS">
  <xs:sequence>
    <xs:element name="Type1">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="Informal"/>
          <xs:enumeration value="Formal"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="Type2">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="Family private"/>
          <xs:enumeration value="Public"/>
          <xs:enumeration value="Formal in natural context"/>
          <xs:enumeration value="Media"/>
          <xs:enumeration value="Telephone"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="Type3">
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="Conversation"/>
          <xs:enumeration value="Monologue"/>
          <xs:enumeration value="Dialogue"/>
          <xs:enumeration value="Political speech"/>
          <xs:enumeration value="Political debate"/>
          <xs:enumeration value="Preaching"/>
          <xs:enumeration value="Teaching"/>
          <xs:enumeration value="Professional explanation"/>
          <xs:enumeration value="Conference"/>
          <xs:enumeration value="Law"/>
          <xs:enumeration value="News"/>
          <xs:enumeration value="Sport"/>
          <xs:enumeration value="Interviews"/>
          <xs:enumeration value="Talk shows"/>
          <xs:enumeration value="Scientific press"/>
          <xs:enumeration value="Meteo"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
  </xs:sequence>
</xs:complexType>

```



```

        <xs:enumeration value="Reportage"/>
        <xs:enumeration value="Private conversation"/>
        <xs:enumeration value="Phone call service"/>
        <xs:enumeration value="Business"/>
    </xs:restriction>
</xs:simpleType>
</xs:element>
<xs:element name="Type4" minOccurs="0">
    <xs:simpleType>
        <xs:restriction base="xs:string">
            <xs:enumeration value="Culture"/>
            <xs:enumeration value="Science"/>
            <xs:enumeration value="Political debate"/>
            <xs:enumeration value="Discussion thematic"/>
            <xs:enumeration value="Machine interaction"/>
            <xs:enumeration value="Man interaction"/>
        </xs:restriction>
    </xs:simpleType>
</xs:element>
<xs:element name="Type5" minOccurs="0">
    <xs:simpleType>
        <xs:restriction base="xs:string">
            <xs:enumeration value="Tourism"/>
            <xs:enumeration value="Health"/>
            <xs:enumeration value="Meteo"/>
            <xs:enumeration value="Traffic"/>
            <xs:enumeration value="Restaurants"/>
            <xs:enumeration value="Train"/>
        </xs:restriction>
    </xs:simpleType>
</xs:element>
</xs:sequence>
</xs:complexType>
<xs:complexType name="TRANSCRIBERS">
    <xs:sequence>
        <xs:element name="Transcriber" type="xs:NMTOKEN"/>
    </xs:sequence>
</xs:complexType>
<xs:complexType name="REVISORS">
    <xs:sequence>
        <xs:element name="Revisor" type="xs:NMTOKEN"/>
        <xs:element name="Prosody" type="xs:NMTOKEN" minOccurs="0"/>
    </xs:sequence>
</xs:complexType>
<!-- TRANSCRIPTION -->
<xs:complexType name="TRANSCRIPTION">
    <xs:sequence>
        <xs:element name="Turn" type="TURN" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
</xs:complexType>
<xs:complexType name="TURN">
    <xs:sequence>
        <xs:element name="Says" type="SAYS"/>
        <xs:element name="Comments" type="COMMENTS" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
</xs:complexType>
<xs:complexType name="COMMENTS">
    <xs:sequence>
        <xs:element name="Comment">
            <xs:complexType>
                <xs:complexContent>
                    <xs:extension base="xs:anyType">
                        <xs:attribute name="Type">
                            <xs:simpleType>
                                <xs:restriction base="xs:string">
                                    <xs:pattern value="alt|act|com"/>
                                </xs:restriction>
                            </xs:simpleType>
                        </xs:attribute>
                        <xs:attribute name="Position" type="xs:positiveInteger"/>
                        <xs:attribute name="Text" type="xs:string"/>
                    </xs:extension>
                </xs:complexContent>
            </xs:complexType>
        </xs:element>
    </xs:sequence>
</xs:complexType>

```

```

    </xs:element>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="SAYS">
  <xs:sequence>
    <xs:element name="Utterance" type="UTTERANCE" maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="Speaker">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:pattern value="[A-Z]{3}"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
</xs:complexType>
<xs:element name="Open_Overlap">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="xs:anyType"/>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:element name="Close_Overlap">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="xs:anyType"/>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:complexType name="UTTERANCE" mixed="true">
  <xs:sequence minOccurs="0" maxOccurs="unbounded">
    <xs:element name="Backwards" minOccurs="0">
      <xs:complexType>
        <xs:complexContent>
          <xs:extension base="xs:anyType"/>
        </xs:complexContent>
      </xs:complexType>
    </xs:element>
    <xs:element ref="Open_Overlap" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="Close_Overlap" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="Unintelligible" minOccurs="0" maxOccurs="unbounded">
      <xs:complexType>
        <xs:complexContent>
          <xs:extension base="xs:anyType"/>
        </xs:complexContent>
      </xs:complexType>
    </xs:element>
    <xs:element name="Partial_Reformulation" minOccurs="0" maxOccurs="unbounded">
      <xs:complexType>
        <xs:complexContent>
          <xs:extension base="xs:anyType"/>
        </xs:complexContent>
      </xs:complexType>
    </xs:element>
    <xs:element name="Total_Reformulation" minOccurs="0" maxOccurs="unbounded">
      <xs:complexType>
        <xs:complexContent>
          <xs:extension base="xs:anyType"/>
        </xs:complexContent>
      </xs:complexType>
    </xs:element>
    <xs:element name="Non_Linguistic" minOccurs="0" maxOccurs="unbounded">
      <xs:complexType>
        <xs:complexContent>
          <xs:extension base="xs:anyType"/>
        </xs:complexContent>
      </xs:complexType>
    </xs:element>
    <xs:element name="Fragment" minOccurs="0">
      <xs:complexType>
        <xs:complexContent>
          <xs:extension base="xs:anyType"/>
        </xs:complexContent>
      </xs:complexType>
    </xs:element>
  </xs:sequence>
</xs:complexType>

```

```

</xs:element>
<xs:element name="Continues" minOccurs="0">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="xs:anyType"/>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:element name="Pause" minOccurs="0">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="xs:anyType"/>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:element name="Tone_Unit" minOccurs="0" maxOccurs="unbounded">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="xs:anyType"/>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:element name="PRAGMATIC_INFORMATION" type="PI" minOccurs="0"
maxOccurs="unbounded"/>
</xs:sequence>
<xs:attribute name="Prosody" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="Enunciation"/>
      <xs:enumeration value="Interrogative"/>
      <xs:enumeration value="Suspension"/>
      <xs:enumeration value="Interruption"/>
      <xs:enumeration value="Non finished"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="Speech_Act" use="optional">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="Rethorical"/>
      <xs:enumeration value="Request"/>
      <xs:enumeration value="Ask"/>
      <xs:enumeration value="Order"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
</xs:complexType>
<xs:complexType name="PI" mixed="true">
  <xs:sequence minOccurs="0" maxOccurs="unbounded">
    <xs:element ref="Open_Overlap" minOccurs="0"/>
    <xs:element ref="Close_Overlap" minOccurs="0"/>
  </xs:sequence>
  <xs:attribute name="GC" use="required">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="Adv"/>
        <xs:enumeration value="Adj"/>
        <xs:enumeration value="N"/>
        <xs:enumeration value="V"/>
        <xs:enumeration value="Prep"/>
        <xs:enumeration value="Interj"/>
        <xs:enumeration value="Det"/>
        <xs:enumeration value="Q"/>
        <xs:enumeration value="O"/>
        <xs:enumeration value="SV"/>
        <xs:enumeration value="SN"/>
        <xs:enumeration value="SP"/>
        <xs:enumeration value="SAdv"/>
        <xs:enumeration value="SAdj"/>
        <xs:enumeration value="SQ"/>
        <xs:enumeration value="C"/>
        <xs:enumeration value="DM"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>

```

```

</xs:attribute>
<xs:attribute name="DISCURSIVE_POSITION" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="utterance_beginning"/>
      <xs:enumeration value="utterance_inside"/>
      <xs:enumeration value="utterance_end"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>

<xs:attribute name="FRASEOLOGICAL_UNIT" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="col"/>
      <xs:enumeration value="loc"/>
      <xs:enumeration value="enun"/>
      <xs:enumeration value="no"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="MET" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:boolean"/>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="source_domain" type="xs:string" use="optional"/>
<xs:attribute name="target_domain" type="xs:string" use="optional"/>

<xs:attribute name="DISCOURSE_RELATION" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="Order"/>
      <xs:enumeration value="Cause"/>
      <xs:enumeration value="Concessive"/>
      <xs:enumeration value="Consequence"/>
      <xs:enumeration value="Adition"/>
      <xs:enumeration value="Purpose"/>
      <xs:enumeration value="Condition"/>
      <xs:enumeration value="Change_Focus"/>
      <xs:enumeration value="Support"/>
      <xs:enumeration value="Inference"/>
      <xs:enumeration value="no"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="EMOTIONAL_DISCOURSE" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="Evaluative"/>
      <xs:enumeration value="Emotion"/>
      <xs:enumeration value="no"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="Type" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="Positive"/>
      <xs:enumeration value="Negative"/>
      <xs:enumeration value="Other"/>
      <xs:enumeration value="no"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="MODALITY" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="Hedge"/>
      <xs:enumeration value="no"/>
      <xs:enumeration value="Booster"/>
      <xs:enumeration value="Interactive"/>
    </xs:restriction>
  </xs:simpleType>

```

```
</xs:attribute>
<xs:attribute name="EVIDENCIALITY" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="1"/>
      <xs:enumeration value="2"/>
      <xs:enumeration value="3"/>
      <xs:enumeration value="4"/>
      <xs:enumeration value="5"/>
      <xs:enumeration value="no"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
</xs:complexType>
</xs:schema>
```

## Appendix B

### Telephonic Conversation in C-ORAL-ROM format

<TimeStamp>C81CB06F-8054-4B18-8CAE-D193BC2F2DD2</TimeStamp>[  
@Title: La bici del amigo de Miguelito  
@File: etelef05  
@Participants: PAD, Ignacio, (man, C, 2, taxi-driver, participant, Madrid)  
MIG, Miguel (man, A, 3, student, participant, Madrid)  
@Date: 05/09/2002  
@Place: Madrid  
@Situation: chat between Ruben's father and Ruben's friend  
@Topic: Miguel phone Rubén, but he isn't and talk to his father  
@Source: C-ORAL-ROM  
@Class: informal, familiar/private, dialogue  
@Length: 0' 26''  
@Words: 89  
@Acoustic\_quality: A  
@Transcriber: Joaquín  
@Revisor: Guillermo  
@Comments:  
]

\*PAD: sí ?  
\*MIG: andeh / hola / está / Rubén ?  
\*PAD: andeh / no / no está //  
\*MIG: ¡ah! / no está ?  
\*PAD: no //  
\*MIG: y / sabes cuándo va a llegar ?  
\*PAD: pues / no tengo ni idea //  
\*MIG: se ha ido con la <bici> ?  
\*PAD: [<] <creo> que sobre las dos / y media / vendrá //  
\*MIG: ¡ah! // bueno / <pues nada> //  
\*PAD: [<] <de todos> modos / se / debe haber llevado el móvil //  
\*MIG: que se ha llevado el móvil ?  
\*PAD: sí // si le quieres llamar ...  
\*MIG: vale //  
\*PAD: vale ?  
\*MIG: pues le dices / <que de todas formas> +  
\*PAD: [<] <dime quién eres> //  
\*MIG: andeh / le dices que ha llamado Miguel / vale ?  
\*PAD: ¡ah! / vale //  
\*MIG: <venga> //  
\*PAD: [<] <hasta luego> //  
%alt: (1) talogo  
\*MIG: hasta luego //  
%alt:

(1)

talogo

## APPENDIX B

### Telephonic Conversation in PRAGMATEXT format

```

<?xml version="1.0" encoding="UTF-8" ?>

<TEXT xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="C:\Documents and Settings\ana
gonzález\Escritorio\birmingham\PRAGMATEXT.xsd">

  <Header>

    <Title>La bici del amigo de Miguelito</Title>
    <File>etelef05</File>
    <Participants>
      <Participant>
        <ID>PAD</ID>
        <Name>Ignacio</Name>
        <Sex>man</Sex>
        <Age>C</Age>
        <Education>1</Education>
        <Occupation>taxi-driver</Occupation>
        <Role>participant</Role>
        <Origin>
          <Place>Madrid</Place>
        </Origin>
      </Participant>
      <Participant>
        <ID>MIG</ID>
        <Name>Miguel</Name>
        <Sex>man</Sex>
        <Age>A</Age>
        <Education>3</Education>
        <Occupation>student</Occupation>
        <Role>participant</Role>
        <Origin>
          <Place>Madrid</Place>
        </Origin>
      </Participant>
    </Participants>
    <Date>1001-05-09</Date>
    <Place>Madrid</Place>
    <Situation>between Ruben's father and Ruben's friend</Situation>
    <Topic>phone Rubén, but he isn't and talk to his father</Topic>
    <Source>C-ORAL-ROM</Source>
    <Class>
      <Type1>Informal</Type1>
      <Type2>Family private</Type2>
      <Type3>Dialogue</Type3>
    </Class>
    <Length>16S</Length>
    <Words>89</Words>
    <Acoustic_quality>A</Acoustic_quality>
    <Transcribers><Transcriber>Joaquín</Transcriber></Transcribers>
    <Revisors><Revisor>Guillermo</Revisor> </Revisors>
    <Comments />
  </Header>

  <Transcription>
  <Turn>

```

```

: <Says Speaker="PAD">
: <Utterance Prosody="Interrogative" Speech_Act="Ask">
<IP GC="V" DP="start" PU="no" MET="false" DR="no" ED="no" Type="no" MOD="Interaction"
  EVI="no">sí ?</IP>
  </Utterance>
</Says>
</Turn>
: <Turn>
: <Says Speaker="MIG">
: <Utterance Prosody="Interrogative" Speech_Act="Ask">
<IP GC="Interj" DP="start" PU="no" MET="false" DR="Support" ED="no" MOD="no"
  EVI="no">eh</IP>
<Tone_Unit />
<IP GC="SP" DP="start" PU="no" MET="false" DR="Order" ED="no" MOD="interaction"
  EVI="no">hola</IP>
<Tone_Unit />
está
<Tone_Unit />
Rubén ?
</Utterance>
</Says>
</Turn>
: <Turn>
: <Says Speaker="PAD">
: <Utterance Prosody="Enunciation">
<IP GC="Interj" DP="start" PU="no" MET="false" DR="Support" ED="no" Type="no" MOD="no"
  EVI="no">eh</IP>
<Tone_Unit />
no
<Tone_Unit />
no está
</Utterance>
</Says>
</Turn>
: <Turn>
: <Says Speaker="MIG">
: <Utterance Prosody="Interrogative" Speech_Act="Ask">
<IP GC="Adj" DP="start" PU="no" MET="false" DR="no" MOD="Interaction" EVI="no">¡ah!</IP>
<Tone_Unit />
no está ?
</Utterance>
</Says>
</Turn>
: <Turn>
: <Says Speaker="PAD">
<Utterance Prosody="Enunciation">no</Utterance>
</Says>
</Turn>
: <Turn>
: <Says Speaker="MIG">
: <Utterance Prosody="Interrogative" Speech_Act="Ask">
y
<Tone_Unit />
sabes cuándo va a llegar ?
</Utterance>
</Says>
</Turn>
: <Turn>
: <Says Speaker="PAD">
: <Utterance Prosody="Enunciation">
<IP GC="C" DP="start" PU="no" MET="false" DR="Consequence" ED="no" Type="no" MOD="no"
  EVI="no">pues</IP>
<Tone_Unit />

```



**no tengo ni idea**  
 </Utterance>  
 </Says>  
 </Turn>  
 = <Turn>  
 = <Says Speaker="MIG">  
 = <Utterance Prosody="Interrogative" Speech\_Act="Ask">  
**se ha ido con la**  
 <Open\_Overlap />  
**bici**  
 <Close\_Overlap />  
 ?  
 </Utterance>  
 </Says>  
 </Turn>  
 = <Turn>  
 = <Says Speaker="PAD">  
 = <Utterance Prosody="Enunciation">  
 <Backwards />  
 <Open\_Overlap />  
**creo**  
 <Close\_Overlap />  
**que sobre las dos**  
 <Tone\_Unit />  
**y media**  
 <Tone\_Unit />  
**vendrá**  
 </Utterance>  
 </Says>  
 </Turn>  
 = <Turn>  
 = <Says Speaker="MIG">  
 = <Utterance Prosody="Enunciation">  
 <IP GC="Adj" DP="start" PU="no" MET="false" DR="no" MOD="Interaction" EVI="no">¡ah!</IP>  
 </Utterance>  
 = <Utterance Prosody="Interrogative">  
 <IP GC="Adj" DP="start" PU="no" MET="true" source\_domain="Human\_Quality"  
 target\_domain="Discourse" DR="Change\_Focus" ED="no" Type="no" MOD="Hedge"  
 EVI="no">bueno</IP>  
 <Tone\_Unit />  
 <Open\_Overlap />  
 <IP GC="C" DP="start" PU="no" MET="false" DR="Consequence" ED="no" Type="no" MOD="no"  
 EVI="no">pues</IP>  
 <IP GC="Adv" DP="end" PU="no" MET="false" DR="no" ED="no" Type="no" MOD="Hedge"  
 EVI="no">nada</IP>  
 <Close\_Overlap />  
 </Utterance>  
 </Says>  
 </Turn>  
 = <Turn>  
 = <Says Speaker="PAD">  
 = <Utterance Prosody="Enunciation">  
 <Backwards />  
 <Open\_Overlap />  
 = <IP GC="SP" DP="start" PU="loc" MET="false" DR="Concessive" ED="no" MOD="no" EVI="no">  
**de todos**  
 <Close\_Overlap />  
**modos**  
 </IP>  
 <Tone\_Unit />  
**se**  
 <Tone\_Unit />  
**debe haber llevado el móvil**

```

</Utterance>
</Says>
</Turn>
<Turn>
<Says Speaker="MIG">
<Utterance Prosody="Interrogative" Speech_Act="Ask">
<IP GC="C" DP="start" PU="no" MET="false" DR="no" ED="no" Type="no" MOD="Interaction"
EVI="no">que</IP>
se ha llevado el móvil ?
</Utterance>
</Says>
</Turn>
<Turn>
<Says Speaker="PAD">
<Utterance Prosody="Enunciation">
<IP GC="Adv" DP="start" PU="no" MET="false" DR="no" ED="no" MOD="Interaction"
EVI="no">sí</IP>
</Utterance>
<Utterance Prosody="Interruption">
<IP GC="C" DP="start" PU="no" MET="false" DR="Condition" ED="no" MOD="no"
EVI="no">si</IP>
le quieres llamar
</Utterance>
</Says>
</Turn>
<Turn>
<Says Speaker="MIG">
<Utterance Prosody="Enunciation">
<IP GC="V" DP="utterance_end" PU="no" MET="true" DR="no" ED="no" MOD="Booster"
EVI="no">vale</IP>
</Utterance>
</Says>
</Turn>
<Turn>
<Says Speaker="PAD">
<Utterance Prosody="Enunciation" Speech_Act="Ask">
<IP GC="V" DP="start" PU="no" MET="false" DR="no" ED="no" MOD="no" EVI="no">vale ?</IP>
</Utterance>
</Says>
</Turn>
<Turn>
<Says Speaker="MIG">
<Utterance Prosody="Interruption">
<IP GC="C" DP="start" PU="no" MET="false" DR="Consequence" ED="no" MOD="no"
EVI="no">pues</IP>
le dices
<Tone_Unit />
<Open_Overlap />
que
<IP GC="SP" DP="start" PU="loc" MET="false" DR="Concessive" ED="no" Type="no" MOD="no"
EVI="no">de todas formas</IP>
<Close_Overlap />
</Utterance>
</Says>
</Turn>
<Turn>
<Says Speaker="PAD">
<Utterance Prosody="Enunciation">
<Backwards />
<Open_Overlap />
dime quién eres
<Close_Overlap />
</Utterance>

```

GRAMMATICAL CATEGORY

```
</Says>
</Turn>
=<Turn>
=<Says Speaker="MIG">
=<Utterance Prosody="Enunciation" Speech_Act="Ask">
<IP GC="Interj" DP="start" PU="no" MET="false" DR="Support" ED="no" MOD="no"
  EVI="no">eh</IP>
<Tone_Unit />
le dices que ha llamado Miguel
<Tone_Unit />
P GC="V" DP="start" PU="no" MET="false" DR="no" ED="no" MOD="interactive" EVI="no">vale
?</IP>
</Utterance>
</Says>
</Turn>
=<Turn>
=<Says Speaker="PAD">
=<Utterance Prosody="Enunciation">
<IP GC="Adj" DP="start" PU="no" MET="false" DR="no" MOD="Interaction" EVI="no">¡ah!</IP>
<Tone_Unit />
<IP GC="V" DP="end" PU="no" MET="true" DR="no" ED="no" MOD="Booster"
  EVI="no">vale</IP>
</Utterance>
</Says>
</Turn>
=<Turn>
=<Says Speaker="MIG">
=<Utterance Prosody="Enunciation">
<Open_Overlap />
<IP GC="V" DP="end" PU="no" MET="true" ED="no" Type="no" MOD="interactive"
  EVI="no">venga</IP>
<Close_Overlap />
</Utterance>
</Says>
</Turn>
=<Turn>
=<Says Speaker="PAD">
=<Utterance Prosody="Enunciation">
<Backwards />
<Open_Overlap />
<IP GC="SP" DP="end" PU="loc" MET="false" DR="no" ED="no" MOD="interactive"
  EVI="no">hasta luego</IP> <Close_Overlap />
</Utterance>
</Says>
=<Comments>
<Comment Type="alt" Position="1" Text="talogo" />
</Comments>
</Turn>
=<Turn>
=<Says Speaker="MIG">
=<Utterance Prosody="Enunciation">
<IP GC="SP" DP="end" PU="loc" MET="false" DR="no" ED="no" MOD="interactive"
  EVI="no">hasta luego</IP>
</Utterance>
</Says>
=<Comments>
<Comment Type="alt" Position="1" Text="talogo" />
</Comments>
</Turn>
</Transcription>
</TEXT>
```