

Exploiting a Spoken Corpus in Language Teaching/Learning: An Advanced Web-based Tool¹

Leonardo Campillos,² Paula Gozalo,³ José María Guirao⁴
and Antonio Moreno Sandoval⁵

Abstract

We present a web-based tool for teaching real and spontaneous Spanish language to intermediate and advanced students. Language samples for grammar, communicative and lexical contents have been extracted from the Spanish subcorpus from C-ORAL-ROM. In addition, fragments of files from the Spanish subcorpus have been selected, whose sound and transcription are retrieved according to their features: difficulty levels, grammar, communicative functions, vocabulary, speed of speech, register or diction clarity.

1. Introduction

The aim of the project we have carried out is twofold: to adapt a spoken corpus for Language Teaching / Learning purposes, and to develop a web-based interface for the users to consult it easily. Firstly, a brief introduction to the C-ORAL-ROM spoken corpus will be presented. Secondly, we will explain the methodology and stages followed to adapt the corpus, and finally the development and the functionality of the web-based interface.

2. The C-ORAL-ROM Spoken Corpus

The European Commission C-ORAL-ROM project (IST 2000-26228) developed a spontaneous spoken corpus for four Romance languages: Italian, French, Spanish and Portuguese (Cresti and Moneglia, eds. 2005). The subcorpus for each language contains approximately 300.000 words and a wide range of registers and communicative contexts are represented: informal conversations and dialogues, media broadcasts, political speech, preaching, teaching, conferences, business and law discourse, weather news, talk shows, telephone conversations... Each sound file includes the following data (Cresti *et al.*, 2004):

- metadata about the participants (sex, age, profession, educational level, geographical origin), the recording circumstances (date, place, context, topic, length, acoustic quality, number of words), the copyright and the linguists who worked on it.

¹ Acknowledgments: this research has been partially supported by the Spanish Ministry of Education and Science under the grant TIN2004-07588-C03-02

² Laboratorio de Lingüística Informática, Univ. Autónoma de Madrid (Spain)
e-mail: leonardo.campillos@estudiante.uam.es

³ Laboratorio de Lingüística Informática, Univ. Autónoma de Madrid (Spain)
e-mail: paula.gozalo@uam.es

⁴ Computer Science Dept., Univ. de Granada (Spain)
e-mail: jmguirao@ugr.es

⁵ Laboratorio de Lingüística Informática, Univ. Autónoma de Madrid (Spain)
e-mail: antonio.msandoval@uam.es

- the original sound and the orthographic transcription (CHAT format) are synchronized
- morphosyntactical information for every transcribed word, for lemmas and tokens

C-ORAL-ROM was designed as a general spoken language corpus for descriptive linguistic studies, speech system training or language teaching. After an evaluation made by a group of 277 experts (Sandoval and Urresti, 2005) some improvements were proposed: to develop a search tool to retrieve documents according to their grammar and communicative contents or their vocabulary; to classify the files according to the *Common European Framework for Languages*, or to add the phonological transcription.

3. The Use of Spoken Corpora for Language Teaching / Learning

Several studies have already been done about the advantages and disadvantages of using corpora in Language Teaching / Learning (Biber, 2000; Sinclair, 2004; Adolphs, 2004; Hunston, 2004; Gabrielatos, 2005). As far as Spanish is concerned, there exists fewer studies (Gimeno Sanz, 2002; San Mateo, 2003; Nicolás, 2002). Obviously, a corpus represents a reduced model of a language, so it may not contain usual communicative situations or important structures of current spoken Spanish. Besides, it can include errors of expression and grammar or offensive words, so it would be advisable for the teacher to pre-edit the corpus. In any case, it is advisable for the teacher to get familiarized with the corpus (principles of design, grammar, communicative and lexical contents...) and the software tool.

Nevertheless, there are lots of advantages, mainly the possibility to work with authentic spoken data directly taken from the use of the language, not from intuition or an academic norm. Although any corpus has limitations, presenting every selected word in its original production context helps to understand it. Furthermore, not only does an oral corpus keep the main characteristics of the oral language (ellipsis, hesitation phenomena, retractings, overlappings, etc), but it also presents structures rarely shown in manuals such as discourse markers (reference of the file and enunciation context from C-ORAL-ROM in brackets):

- (1) *INT: estamos andespera [/] a la espera / de [/] de que hayan abonado / los gastos de instalación //
*GAR: **¡ah!** / **bueno** //
[*INT: we andwait [/] are waiting / for [/] for them to pay / the installation expenses //
*GAR: **ah** / **all right**] (epubdl13-54, epubdl13-55)
- (2) *SUS: **vamos a ver** / andeh estos son los resus / eh ? (enatco01-28)
[*SUS: **let's see** / andeh these are the rhesus / ok ?]
- (3) *MON: **venga** / cómo se solucionaría esto ? (epubdl16-139)
[*MON: **come on** / how would it be solved ?]

Using the C-ORAL-ROM speech corpus benefits from the fact that it contains real conversations and dialogues, so that typical phenomena from these genres are considered: structures which break the dialogic turns (4), conversation openings (5), etc.:

- (4) *ANS: <parece> +
*LUI: / [<] <sí / andperdo> [/] **perdona** / Luis María // antes de conocer tu opinión // vamos a recoger / las [/] las opiniones que al contestador están dejando los palestinos e israelíes / (from emedts02-74 to emedts02-77)
[*ANS: <it looks like> +
*LUI: [<] <yes / andexcu> [/] **excuse me** / Luis María // before knowing your opinion // we are going to collect / the [/] the opinions that Palestinian and Israeli people are giving on the answering machine]
- (5) *SEV: **bueno** / yo no sé si / el que no se bajen los tipos / es bueno o no para nuestra economía
[*SEV: **well** / I don't know if / the fact that the rates are not reduced / is good or not for our economy]

(emedin02-18)

Also, usual affirmative or negative answers from the colloquial oral Spanish are included:

- (6) *GAB: bueno / esto es lo que decía Santo Tomás / no ?
*JUA: **eso es** // (emedsp06-41)
[*GAB: well this is what Saint Thomas used to say isn't it?
*JUA: **that's it**]
- (7) *VIT: [<] <¡ah! / no os ha salido muy caro> / entonces //
*BEA: no // **qué va** // (efamd135-31)
[*VIT: [<] <ah / so it didn't work out very expensive> //
*BEA: no // **of course not**]

Another advantage is that other pragmatic phenomena such as courtesy markers are also included in our selection. In example 8, the modal *poder* ('can') is used to request information in a polite way beyond its semantic meaning ('be able to do something'):

- (8) *PED: nos **puede** explicar esto ? (emedsp06-114)
[*PED: **could you** explain this to us ?]

Likewise, example 9 shows another structure with the modal *poder*, this time used when somebody makes an invitation to someone in an informal way:

- (9) *PRI: y y luego / dice ya sabes que **puedes** venir cuando quieras // (efammn08-306)
[*PRI: and and then / she says you know you **can** come whenever you want //]

One of the most attractive features is that some expressive idioms or phraseological units from colloquial Spanish are presented in context, so they are understood more easily:

- (10) *LAU: está gorda / está fea / está **arrugada como una pasa** / tiene el pelo de bruja (efamd102-79)
[*LAU: she is fat / she is ugly / **her face is completely wrinkled** / she has a witch face]
- (11) *ABE: voy a tomarme un café / y ya nos vamos a **dar una vuelta** por ahí // (efamd131-16)
[*ABE: I'm going to take a coffee / and **we'll go out for a walk**]

Finally, collocations show usual patterns of concurrence which are fundamental not only to express things very usual in daily life (example 12), but also to get a proficiency level by using structures in a formal register (example 13):

- (12) *MAR: bastantes deseosos de [/] de **darnos un baño** / de dormir / en fin / (efammn09-32)
[*MAR: very eager to [/] to **take a bath** / to sleep / anyway]
- (13) *VIC: quería / estar seguro de que el club actuaba bien / que **no cometíamos ningún error** /
[*VIC: I wanted / to be sure that the club was acting well / that we **didn't make any mistake**]

(emedsp03-19)

4. Adapting the Spanish C-ORAL-ROM to the Teaching / Learning

The C-ORAL-ROM corpus has two main inconveniences for it to be used directly in Second Language Teaching: some of the conversations are difficult to use in Teaching / Learning contexts and some computer tools included in the DVD published by Benjamins are not highly user-friendly. Thus, an adaptation was necessary to be used directly by teachers in the classroom or by students in self-learning contexts. Therefore, for our project to improve these inadequacies we followed a methodology focused on four action lines:

- 1) Design of a Teaching / Learning of Spanish as a foreign language syllabus.
- 2) Selection and extraction from the Spanish subcorpus of C-ORAL-ROM of those samples most adequate for Teaching / Learning Spanish.
- 3) Selection of files from C-ORAL-ROM for Teaching / Learning Spanish and classification according to the reference levels established in the *Common European Framework of Reference for Languages*.
- 4) Development of a web-interface to consult the selected subcorpus from C-ORAL-ROM.

We should remark that we didn't use the corpus directly in the teaching process, what some linguists call a *Data-driven Learning* (Johns, 1991; Bernardini, 2004).

4.1 Design of a Teaching / Learning of Spanish as a Foreign Language Syllabus

After documentation work based on manuals for teaching Spanish as a Foreign Language, the selection of structures to learn Spanish was carried out at three levels: grammatical, communicative and lexical contents. A descriptive analysis and communicative teaching approach was adopted, with contributions from Pragmatics (specially for categories such as "Oral communication skills" or "Discourse markers").

4.1.1 Grammar Structures

Normative grammars such as Gómez Torrego (1998) were not sufficient to design the grammar syllabus, being it necessary to look up grammars for foreign students (Matte Bon: 2004; Palencia and Aragonés: 2003; Benítez and Gelabert: 2000; Cerrolaza: 2005). Our syllabus was augmented with specific learning difficulties of Spanish as a foreign language such as the omission of the article or the uses of verb *tener* ('to have'); furthermore, we created specific categories focused on the Subjunctive, *Ser* and *estar* ('to be') or the uses of the *se* pronoun (e.g. *se* for passive, *se* personal pronoun...). Table 1 lists the selection for the Grammar.

<p><u>I – Word level</u></p> <ol style="list-style-type: none"> 1. Articles 2. Nouns 3. Determiners and pronouns <ol style="list-style-type: none"> 3.1. Demonstratives 3.2. Possessives 3.3. Indefinites 3.4. Numerals 3.5. Interrogatives and exclamatives 3.6. Personal pronouns 3.7. Relative pronouns y relative adverbs 4. Adjectives 5. Verbs and tenses. <ol style="list-style-type: none"> 5.1. Indicative tenses <ol style="list-style-type: none"> 5.1.1. Present, future, preterite ('indefinido'), imperfect, past perfect ('pluscuamperfecto') 5.1.2. Conditional 5.2. Subjunctive tenses 	<ol style="list-style-type: none"> 5.3. Imperative 5.4. Infinitive, gerund and past participle 5.5. Verbal periphrasis 5.6. Passive 5.7. Ser / estar 6. Markers of space and time 7. Adverbs and adverbial idioms 8. Prepositions and prepositional idioms <p><u>II – Enunciation and clause level</u></p> <ol style="list-style-type: none"> 1. Coordinate clauses 2. Subordinate clauses <ol style="list-style-type: none"> 2.1. Noun clauses 2.2. Relative clauses 2.3. Adverb clauses 3. Discourse markers 4. Uses of the <i>se</i> pronoun <p><u>III - Orthography</u></p> <p>Orthography, accent and acronyms</p>
--	---

Table 1: Grammar structures selected for Teaching / Learning Spanish

4.1.2 Communicative Notions and Functions

The selection and organisation of communicative contents was based on grammars for foreign students such as Matte Bon (2004), and indexes of communicative notions and functions, such as those developed for the threshold level or B1 (van Ek, 1975) and the waystage level or A2 (van Ek, Alexander and Fitzpatrick, 1977), or other specific classifications for Spanish (Gelabert *et al.* 1996). Besides, we considered the speech acts early proposed by J. L. Austin or J. Searle. Communicative functions were put together in five groups: “Social customs”, “Physical and emotional states”, “Attitudes and knowledge”, “Influence”, and “Communication”, as table 2 shows.

<u>Notions</u>	<u>Communicative functions</u>
Being and existing	ATTITUDES AND KNOWLEDGE
Quantity	Expressing possibility/impossibility and ability/inability
Time	Certainty and probability
Location and spatial relations	Expressing knowledge, memory and oblivion
Relations among events or processes	Opinion
– Conditional relations	Agreement and disagreement
– Concessive relations	Obligation and necessity
– Causal relations	
– Consequence relations	INFLUENCE
– Final relations	Advice, warnings and recommendations
Expressing manner, means and instrument	Suggestions and proposals
Expressing comparison	Requests
Expressing ownership	Complaints and reclamations
Expressing intensity and exclaiming	Arguments, threats and insults
Impersonality	Encourage to act
<u>Communicative functions</u>	Promises, commitments and oaths
SOCIAL CUSTOMS	Instructions, orders, interdictions. Allow and request permission
Greetings and goodbye. Introducing people	
Invitations. Dates and appointments	
Giving thanks	
Apologizing	
Social functions: toast, condolences, congratulations, compliment and courtesy	
PHYSICAL AND EMOTIONAL STATES	COMMUNICATION
Likes and preferences	Oral communication skills
Feelings and emotions	Organising information
Wishes	Controlling the language
Physical states	Reported speech

Table 2: Communicative notions and functions selected for Teaching / Learning Spanish

4.1.3. Lexical Categories

Lists of vocabulary classified in semantic fields (Sánchez Lobato and Aguirre, 1992) have been a model for our selection, which is shown in table 3.

1. Personal identifications 2. Profession and job 3. Religion, politics and ideology 4. Family 5. Character, values and emotional states 6. Home and house 7. Free time and leisure time 8. Nature 9. Weather 10. Art and culture 11. Communication media 12. Sports, shows and entertainments 13. Travelling 14. Residence and accommodation	15. Transports 16. Body and health 17. Education 18. Food and gastronomy 19. Commerce, business and industry 20. Public administration and civil institutions 21. Units of measurement 22. Society and State 23. Describing objects, places, situations 24. Computing, Internet and other technologies 25. Cultural heritage 26. Phraseology and idioms 27. Collocations
--	--

Table 3: Categories of vocabulary selected for Teaching / Learning Spanish

4.2 Selection and Extraction of Samples from the Spanish C-ORAL-ROM

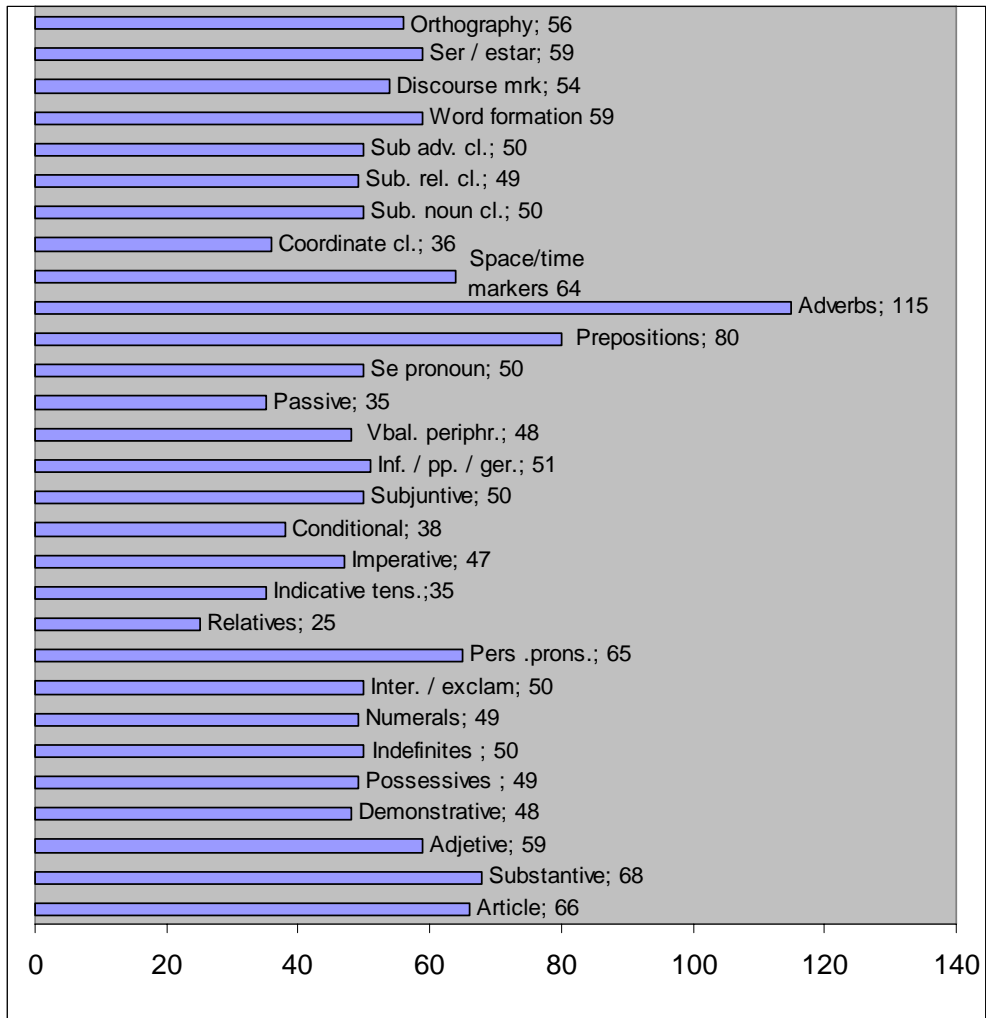
Two methodologies were followed to collect the language samples:

- extraction and search of a list of structures for each category with a concordance program. On the one hand, finding any structure is fast and straightforward; however, the discourse is fragmented, the context for the whole comprehension is often lost.
- reading and listening the whole documents: it involves more time and attention, but the global comprehension is improved and structures not taken into account at first are found.

Samples were selected following a descriptive analysis with these criteria:

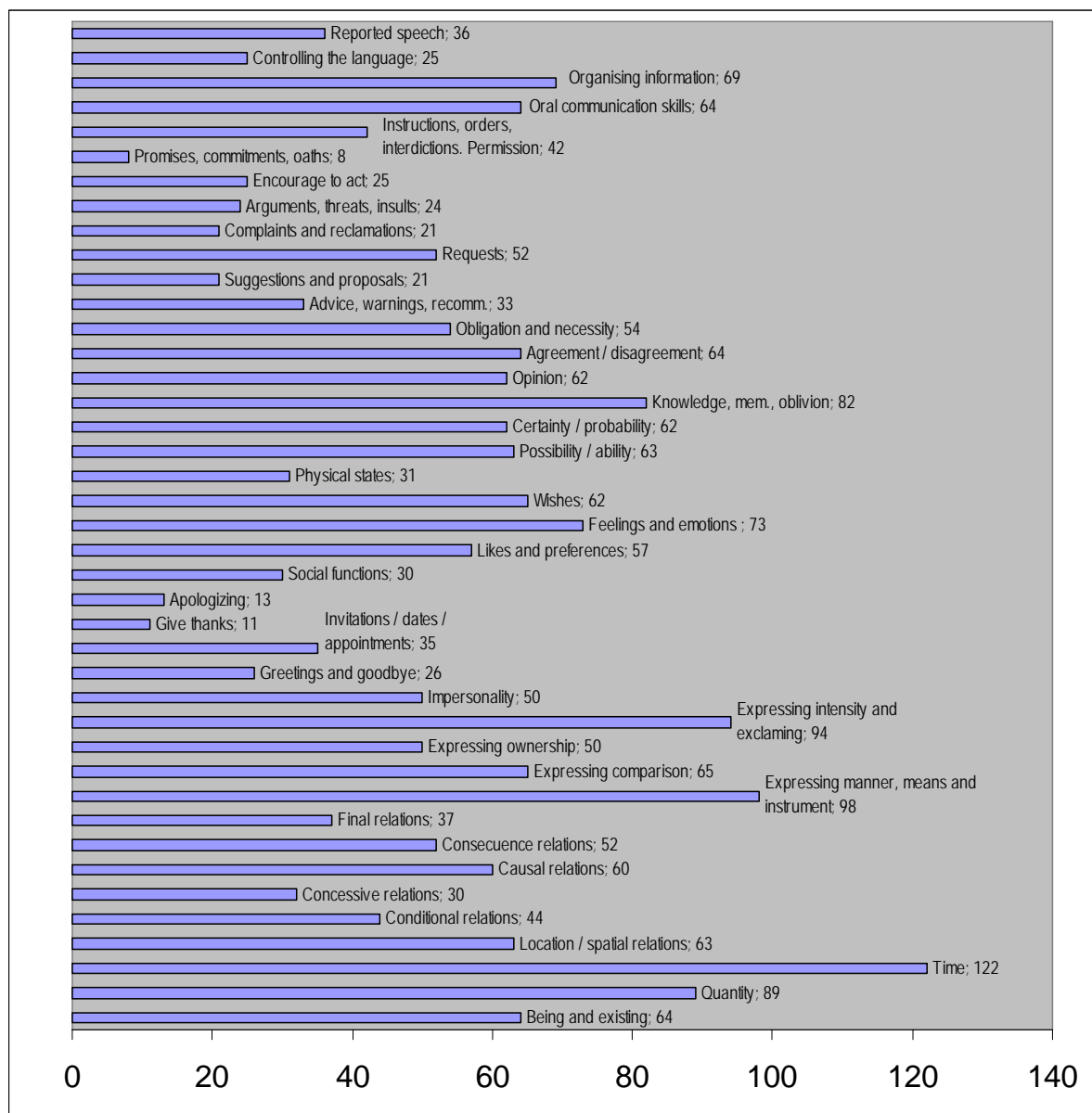
- Linguistic criteria:
 - Grammar, communicative and lexical contents, as explained before.
 - Linguistic variety: samples from a standard peninsular Spanish were preferred, but some samples from American or Southern Spanish were also selected if the accent did not affect the comprehension of the searched structure.
- Extralinguistic criteria:
 - Number of samples: at first we considered a selection of 50 samples for each category, but it was surpassed for categories such as “Expressing time” or “Ser / estar”
 - Acoustic quality: some structures were not collected if they were not heard properly, due to a lack of quality of the recording or due to speech overlappings.

The number of samples for each category of grammar structures is presented below.



Graph 1: Comparative graph with the number of samples selected for each grammar category

Regarding the communicative contents, the following graph shows the number of samples for each category:



Graph 2: Comparative graph with the number of samples selected for each communicative category

4.3 Selection of the Files from C-ORAL-ROM and Classification in Levels

Apart from collecting the language samples, a selection of files from the Spanish corpus of C-ORAL-ROM has been made for Teaching / Learning purposes. In a first stage, 92 texts (along with their aligned sound files) were selected out of the 183 in the original C-ORAL-ROM, which represents approximately 50% of the Spanish subcorpus. Nevertheless, as the duration of the majority of these files is more than five minutes, in a second stage we fragmented them to make it easier its use. We obtained a total of 199 fragments, each one having a total unity of sense.

These fragments have been classified according to the reference levels established in the *Common European Framework for Languages*, after taking in consideration the characteristics of each document:

- Grammar structures, communicative functions, and lexical topics explained in §4.1
- Speed of speech

- Register
- Diction clarity
- Linguistic variety
- Characteristics that influence the auditory comprehension (Anderson and Lynch, 1988; from Gozalo Gómez: 2000):

- the organization of the information: comprehension is easier when listening to well structured contents, but more difficult if interruptions or reformulations appear in the discourse.
- the specificity of the subject: topics or contents from very specialized thematic domains may be appropriate for high levels or for Teaching / Learning for specific purposes.
- The explicitness of the information: redundant terms or expressions are easier to listen to, whereas referential expressions (deixis phenomena such as anaphors or cataphors) are more difficult.

Out of a total of 199 fragments, 3 were classified in level C2; 55, in level C1; 9, in level B1; and 132, in level B2. As the graph shows, the level B2 predominates in our selection, followed by the level C1.

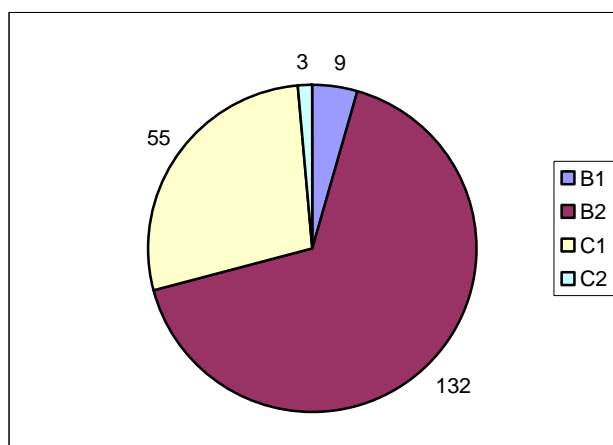


Figure 1: Number of fragments of files from the Spanish C-ORAL-ROM selected for Teaching / Learning and difficulty levels according to the *Common European Framework of Reference*

4.4 Development of the Web-Based Interface

A web-based interface has been developed so that teachers or students are able to retrieve samples and fragments of documents from C-ORAL-ROM appropriate for Teaching / Learning Spanish. Figure 3 shows the initial (front) page.



Figure 2: Initial (front) page

After the initial page, a general index (fig. 4) is presented for users to consult the selection from the Spanish C-ORAL-ROM in two ways:

- 1) The language samples selected from the Spanish C-ORAL-ROM for each category of Grammar structures, Communicative notions and functions, and Lexical topics
- 2) The fragments of files selected from the Spanish C-ORAL-ROM (“Textos”)

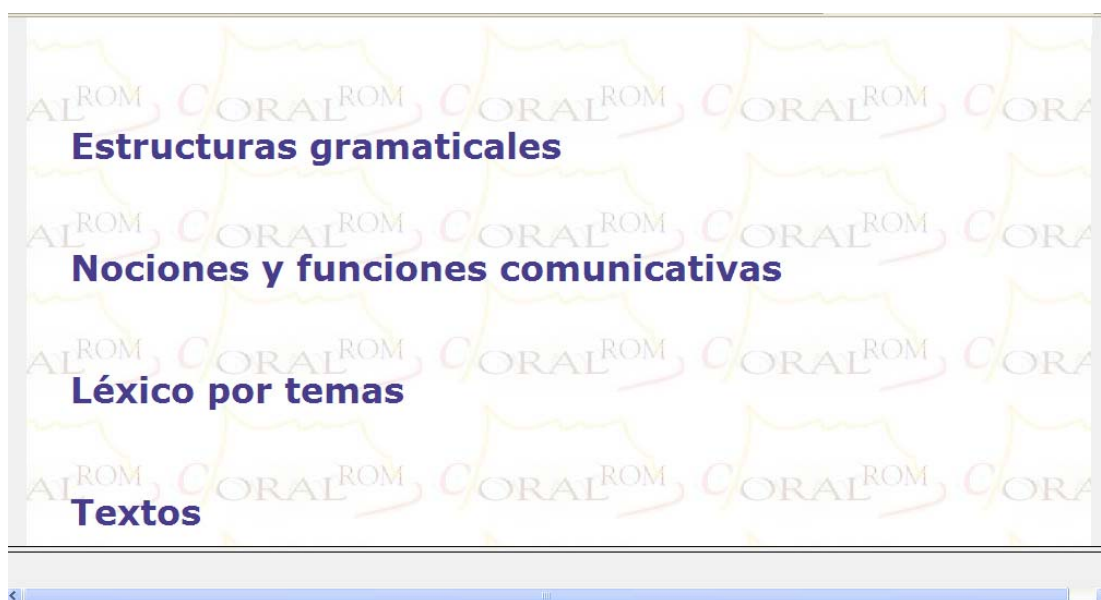


Figure 3: Index of subsections: “Grammar structures”, “Communicative Notions and Functions”, “Lexical topics” and “Texts”

Figure 4 shows the browsing scheme for the web interface.

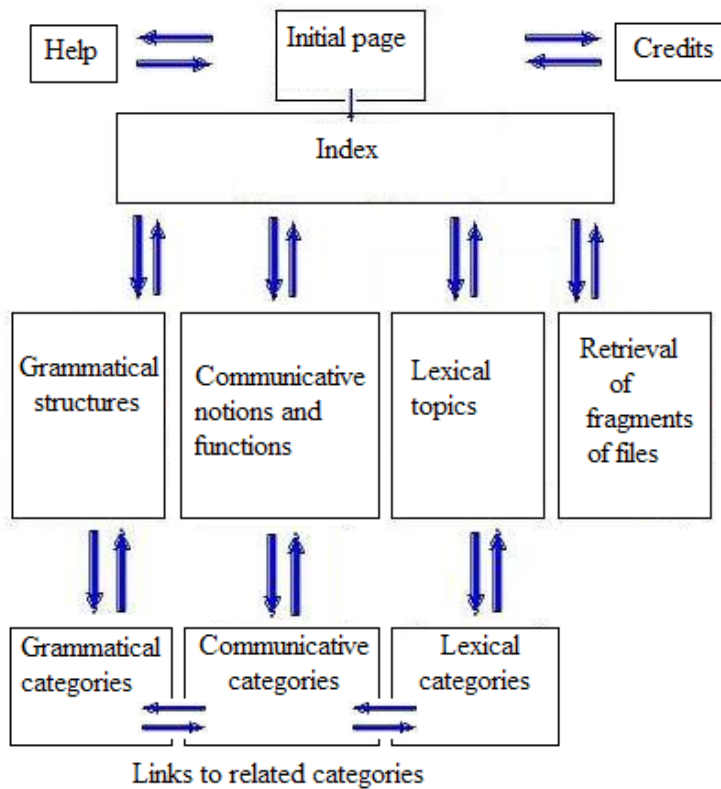


Figure 4: Browsing scheme for the web interface

4.4.1 Consulting the Language Samples

After the general index of subsections, a menu is presented with all the subcategories corresponding for the Grammar structures (fig. 6), Communicative notions and functions (fig. 7) and Lexical categories (fig. 8).

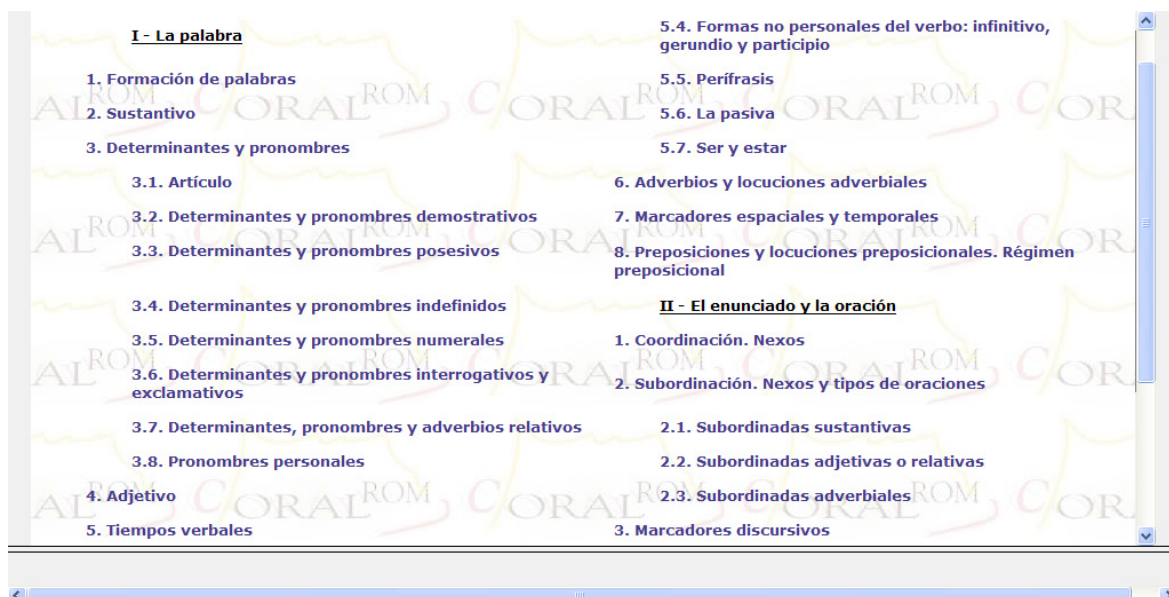


Figure 5: Index of categories of the grammar contents

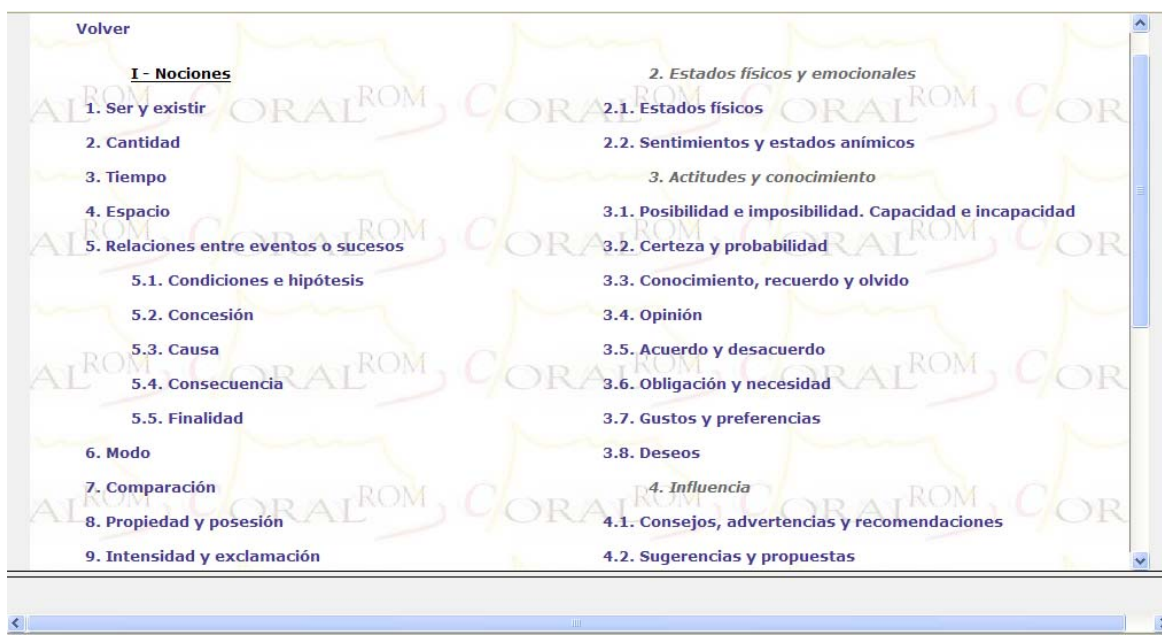


Figure 6: Index of categories of the communicative notions and functions

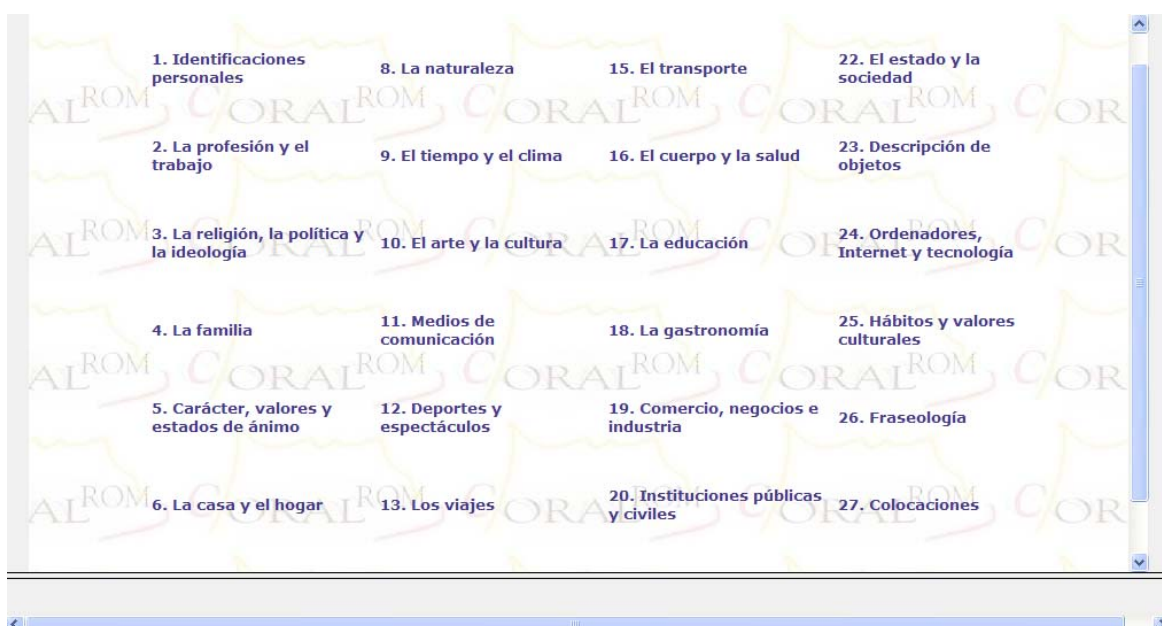


Figure 7: Index of lexical categories

A menu presenting all the sections for the selected category is opened by clicking on any subcategory. The user can also return to the previous category by clicking on the link “Volver” (circled in red in figure 9).

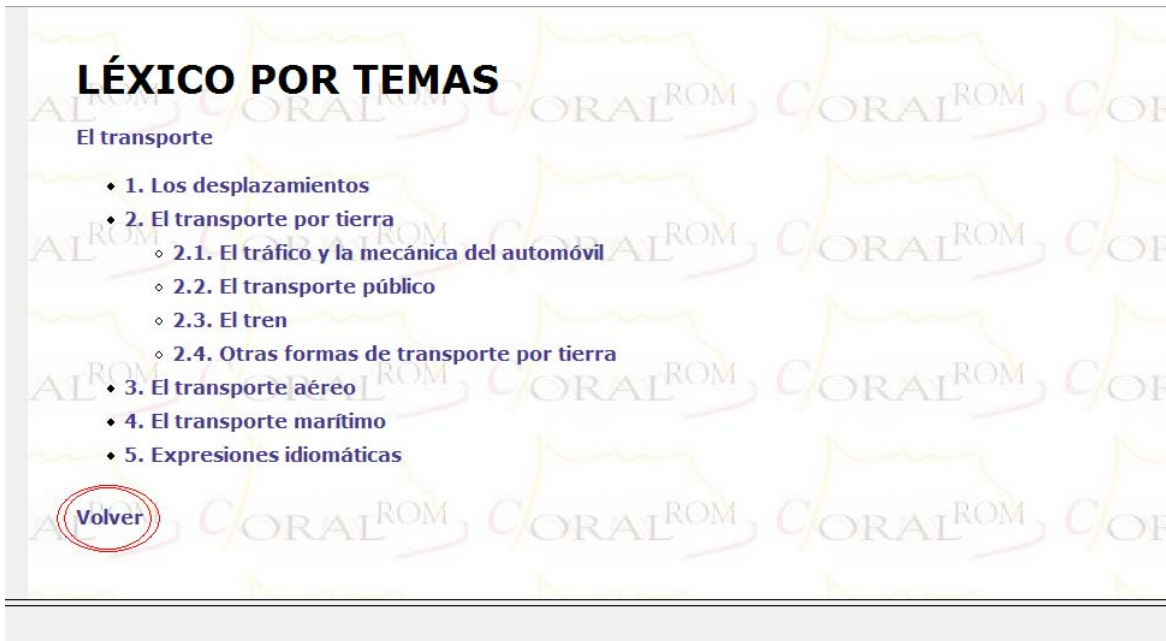


Figure 8: Index of sections for the lexical category “Transports”. The link to the previous index (“Lexical topics”) is circled in red

By clicking on the corresponding link for any section in the subcategory menu, the user directly accesses the searched structures. Grammar explanations are exemplified with samples from real oral language, and the consulted structures are shown in bold in their original context. The user can listen to the samples by clicking on the play button to the left of the icon (fig. 10).



Figure 9: Grammar explanations and language samples of the structures in their real oral context

Some sections are related to others in different categories; for example, lexical category “Transport” is related with “Travelling”. For the user to consult the related category easily, a link is presented with the name of the category and a right arrow (circled in red in figure 10). Besides, a link permits the user to return to the main index (“Índice”, circled in green).

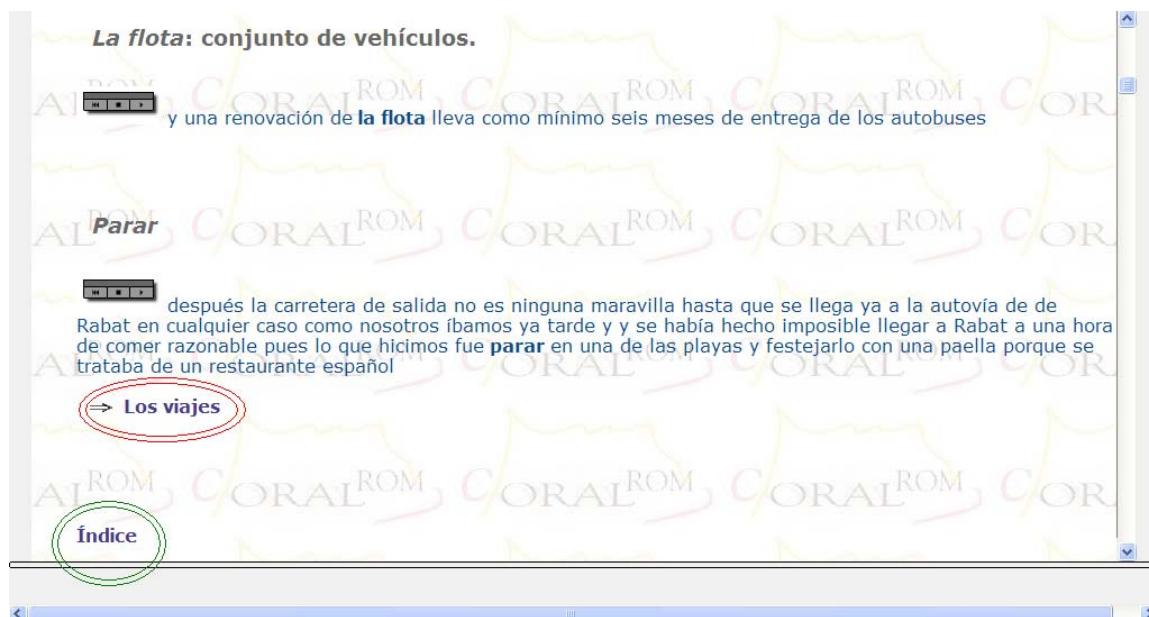


Figure 10: Link to a related category (circled in red) and to the main index of the category (circled in green)

4.4.2 Retrieval of the Fragments of Documents from C-ORAL-ROM

The sound and the transcription of the fragments of files are retrieved according to their features:

- grammatical structures, communicative notions and functions, lexical categories as explained in §4.1
- speed of speech: five levels are distinguished: slow, middle-slow, middle, middle-fast and fast.
- diction clarity: depending on the quantity of pronunciation errors present in the files, five levels are proposed: bad, bad-normal, normal, good, very good.
- register: four levels have been proposed:
 - Colloquial: informal speech, including the familiar register
 - Formal: typical in the communication media or the academic language
 - Professional: specialized jargon, such as scientific vocabulary
 - Standard: neutral discourse level, neither too learned nor too relaxed.

5. Conclusions

As outlined above, an effort has been made not only to take advantage of corpus in Applied Linguistics, but of the new computing possibilities to complement Language Teaching / Learning methods. The future lines of work would be to develop activities and exercises based on samples from C-ORAL-ROM to practice oral comprehension skills, and to increase the number of language samples with new recordings and transcriptions. A special focus should be done to develop new corpus to be used in Linguistic applications for Specific Purposes.

References

- Benítez, P.; Gelabert, M. J. (2000) *Breve gramática: español lengua extranjera*. (Barcelona: Difusión)
- Bernardini, S. (2004), Corpora in the classroom. An overview and some reflections on future developments, in J. M. Sinclair, *How to use Corpora in Language Teaching*. Studies in Corpus Linguistics, 12 (Philadelphia: John Benjamins)
- Biber, D., Conrad, S., Reppen, R. (1998) *Corpus linguistics. Investigating language structure and use*. (Cambridge: Cambridge University Press)
- Campillos Llanos, L.; Gozalo Gómez, P.; Moreno Sandoval, A. (2006) El corpus C-ORAL-ROM en la enseñanza de ELE. *Actas del XVII congreso internacional ASELE para la enseñanza del español como lengua extranjera*. Universidad de La Rioja (not published yet)
- Cerrolaza Gili, O. (2005) *Diccionario práctico de gramática*. (Madrid: Edelsa)
- Cresti, E.; Bacelar do Nascimento, F.; Moreno Sandoval, A.; Veronis, J.; Martin, P., Choukri, K. (2004) The C-ORAL-ROM CORPUS: A Multilingual Resource of Spontaneous Speech for Romance Languages, Available on-line from http://lablita.dit.unifi.it/coralrom/papers/coralrom_lrec2004.pdf (accessed: March 22th, 2007)
- Cresti, E.; Moneglia, M. (2005) *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Studies in Corpus Linguistics, 13. (Amsterdam/Philadelphia: J. Benjamins). The book including a DVD can be acquired in http://www.benjamins.com/cgi-bin/t_bookview.cgi?bookid=SCL%2015. More information on the C-ORAL-ROM project in <http://lablita.dit.unifi.it/coralrom/> (accessed: March 22nd, 2007)
- Ek, J A van; Alexander, L G; Fitzpatrick, M A (1977) *Waystage English* (European Council: Pergamon Press)
- Gabrielatos, C. (2005) Corpus and Language Teaching: Just a fling or wedding bells?, *Teaching Language as a Second or Foreign Language (TESL-EJ)*, vol. 8, n. 4, A-1. Available on-line from <http://www.tesl-ej.org/ej32/a1.html> (accessed: March 22nd, 2007)
- Gelabert, M^a. J.; Herrera, M.; Martinell, E.; Martinell, F. (1996) *Repertorio de funciones comunicativas del español: Niveles umbral, intermedio y avanzado* (Madrid: SGEL)
- Gimeno Sanz, A. M. (coord.)(2002) *Tecnologías de la información y de las comunicaciones en la enseñanza de ELE - Actas del XII congreso de ASELE, 5-8 septiembre de 2001*. (Valencia: Publicaciones de la Universidad Politécnica de Valencia)
- Gómez Torrego, L (1998) *Gramática didáctica del español* (Madrid: SM)
- Gozalo Gómez, P. (2000) El papel de la comprensión auditiva en la enseñanza aprendizaje de E/LE, *Frecuencia L*, n. 15 (Madrid: Edinumen)
- Hunston, S. (2002), *Corpora in Applied Linguistics*. Cambridge: University Press
- Johns, T. (1991) Should you be persuaded – Two samples of data-driven learning materials, in T. Johns and P. King (eds.) *Classroom concordancing*, ELR Journal, 4, 1-16
- Matte Bon, F. (2004) *Gramática comunicativa del español. Tomo I: de la lengua a la idea; Gramática comunicativa del español. Tomo II: de la idea a la lengua* (Madrid: Difusión)
- Moreno Sandoval, A., Urresti, J. (2005) El proyecto C-ORAL-ROM y su aplicación a la enseñanza del español, *ORALIA*, vol. 8, 81-104 (Madrid: Editorial Arco/Libros)
- Nicolás, C. (2002) Una propuesta de utilización de corpus orales en la enseñanza de segundas lenguas, Congreso CLiP 2002, *Lamusa Digital*, n. 3 (monographic), (Universidad de Castilla-La Mancha). Available on-line from www.uclm.es/lamusa/index.asp?accion=misceandid=52andlengua=es (accessed: March 22th, 2007)

- Palencia, R.; Aragonés, L. (2003) *Gramática de uso del español*. (Madrid: SM).
- San Mateo, A. (2003) Los corpus lingüísticos y la enseñanza de ELE, *Frecuencia ELE*, 22. (Madrid: Edinumen)
- Sánchez Lobato, J.; Aguirre, B (1992) *Léxico fundamental del español (situaciones, temas y nociones): Glosario multilingüe*. (Madrid: SGEL)
- Sinclair, J. M. (2004) *How to use Corpora in Language Teaching*. Studies in Corpus Linguistics, 12. (Philadelphia: John Benjamins)
- van Ek, J. A. (1975) *The Threshold Level*. Strasbourg: Council of Europe. (Repr. 1980. Oxford: Pergamon Press.)
- van Ek, J. A.; Alexander, L. G.; Fitzpatrick, M. A. (1977) *Waystage English*. (Oxford: Pergamon Press)