The C-ORAL-ROM Project. New methods for spoken language archives in a multilingual romance corpus

Emanuela Cresti, Massimo Moneglia*, Fernanda Bacelar do Nascimento*, Antonio Moreno Sandoval*, Jean Veronis*, Philippe Martin*, Kalid Choukri, Valerie Mapelli*, Daniele Falavigna*, Antonio Cid*, Claude Blum*

* Dipartimento di Italianistica, Università di Firenze
Piazza Savonarola 1,50125 Firenze Italy
elicresti@unifi.it
*Centro de Linguistica da Universidade de Lisboa
Complexo Interdisciplinar, Av Gama Pinto, 2, 1649-003 Lisboa Portugal
fbacelar.nascimento@clul.ul.pt
*Laboratorio de Lingüística Informática Departemento de Linguistica, Universidad Autonoma de Madrid
Carretera de Colmenar Viejo Km 15 Cantoblanco 28049 Madrid Spain
Sandoval@maria.lllf.uam.es
*Description Linguistique Informatizée sur Corpus, Université de Provence
29, Avenue Robert Schuman13621 AIX EN PROVENCE - Cedex 1 France
Jean.Veronis@up.univ-mrs.fr
*European Language Distribution Agency
European Language Association Agency (ELDA)
55-57, Rue Brillant-Savarin 75013 Paris France
choukri@elda.fr
*Istituto Trentino di Cultura, Trento
Istituto Trentino di Cultura (Centro per la ricerca scientifica e tecnologica)
38050Povo, Trento, Italy
falavi@itc.it
*Pitch Instruments France
24, Rue Las Cases 75005 France
philippe.martin@fnac.net
*Instituto Cervantes, Oficina del Español en la Sociedad de la Información
Livreros, 23 28801 Alcalà de Henares - Madrid Spain
diracad@cervantes.es
*Editions Honoré CHAMPION
7, Quai Malaquais 75006 PARIS France
piquart@club-internet.fr

Abstract

C-ORAL-ROM is a multilingual corpus of spontaneous speech of around 1.200.000 words representing the four main Romance languages: French, Italian, Portuguese and Spanish.. The resource will be delivered in standard textual format, aligned to the audio source in a multimedia edition. C-ORAL-ROM aims to ensure at the same time a sufficient representation of spontaneous speech variation in each language resource and the comparability among the four resources with respect to a definite set of variation parameters. The multimedia conception of C-ORAL-ROM allows simultaneously alignment and full appreciation of the acoustic information through the speech software WINPITCHCORPUS. The storage of spoken language resources is based on the identification of *utterances* in the four corpora through perceptively relevant prosodic properties. In C-ORAL-ROM all the textual information is tagged simultaneously with respect to prosodic parsing and utterance limits. Each prosodic unit corresponding to an utterance is easily and directly aligned to its acoustic counterpart, thus ensuring a natural text - sound correspondence and the definition of a data base of possible speech act in the four romance languages.

1. Introduction

The main goal of the C-ORAL-ROM Project is to provide a comparable set of corpora of spontaneous speech for the main Romance Languages, namely French, Italian, Portuguese and Spanish (roughly 300,000 words for each language). The project has been funded under the IST program of the EU and is being carried out by a European consortium co-ordinated by the University of Florence[1]. The resource has been set up during 2001 with a large reuse of corpora of spontaneous speech collected in previous academic studies (See. Cresti, 2000; Bacelar do Nascimento, 2001; Lavacchi & Nicolas; 2000; Blanche-Benveniste, in Press)

The C-ORAL-ROM Corpora will be delivered in the same textual format following present EU standard (EAGLE) in a multimedia edition on DVDs, integrated

[1] C-ORAL-ROM (IST 200026228). Official web site: http://lablita.dit.unifi.it/coralrom

with tools, assuring both concordances of the text and detailed analysis of the acoustic signal. The Corpus edition will be associated with comparative linguistic studies, models and standard linguistic measures of spontaneous spoken language variability. Edition and distribution for academic studies will be performed by Champion, while ELDA will distribute the LR to speech industry for HLT purpose.

The paper focus on two features of the project that constitute the main novelty of the LR:

- sampling criteria adopted to ensure comparability and spontaneous speech representation;
- the multimedia designing of the C-ORAL-ROM spoken resource.

2. Representation of spontaneous speech and comparability in a multilingual LR

2.1. The representation issue

The Spontaneous Spoken Language areas have become consolidated only in quite recent times (See. Biber, 1988; Blanche-Benveniste, 1990; Cresti, 2000; Givon, 1979; Miller & Weinert, 1999). Spontaneous speech is characterised by:

(a) variable sound quality;
(b) face-to-face dialogue in large variety of communicative structures
(c) mental programming simultaneous with vocal execution (un-scripted)
(d) contextually undetermined linguistic behaviour (unpredictable behaviour)

The setting up of Spontaneous Speech databases is a complex task. Spoken resources set up in controlled environments (such as telephone information, health dialogues, map tasking) constitute at present the majority of the databases used for the validation of language engineering. Their acoustic/phonetic quality is excellent, but they deal with highly predictable semantic domains. Should one wish to represent Spontaneous Speech in a LR, the constitution criteria must ensure the widest possible variation in speech contexts, and a low control on the speech event, that is exactly the opposite of what dedicated resources do.

There are many reasons for this necessity. Variability is the main property of spontaneous spoken texts. As a matter of fact almost the complete set of linguistic levels of language description varies their quantitative weight a lot, when considered with respect to different pragmatic domains. See the following arguments:

Frequency lexicon level. The representation of a sufficient number of contexts covering, as far as possible, relevant types of speech events in the universe, is the only possible strategy to identify significant frequency lexicons. High frequency lexicon defined with respect to general corpora may be under-represented in specific pragmatic domains which on the contrary, by definition, maximise the probability of occurrence of low frequency lexical items. That is the real interest for the rigid definition of a semantic domain in the setting up of comparable corpora of dedicated resources.

Syntactic level. It has been noted that in general corpora (Biber et al., 1999) nouns are more frequent than verbs, but also that the relative frequency of nouns is much lower in informal conversations with respect to formal contexts (1/1 vs 1/3). Adjectives, on the contrary are much more frequent in formal speech.

In the domain of *corpus based grammars*, the induction of the main syntactic properties is strongly correlated to text variation parameters. For example in English, both main types of dependent clauses (relative and complement clauses) vary their relative frequency according to socio-linguistic parameters. Generally speaking, in syntactic structures controlled by a noun, the frequency of both *that-clauses* and *to-clauses* is higher in formal language, while, in *verb-controlled structures*, *that-clauses* are much more frequent in conversation (Biber, 2000). Similar conclusions can be drawn with respect to relative clauses. Relative constructions are much more frequent in formal speech, while the restrictive function is the more frequent, among relative clause functions, in the all corpus variation (Biber et al., 1999). In other words, the pragmatic domain of corpora collection strongly influences the probability of occurrence of syntactic properties of spontaneous speech in the core area of grammar.

In between syntactic and lexical properties. It is essential to the grammatical description of spoken language to note that the majority of complement clauses which depend on a verb, depend on a *putandi* verb in spontaneous conversation. However, such important data is also relative to variation parameters. For example, a complement clause depends quite frequently on a *dicendi* verb in broadcasting and media contexts (Biber et al., 1999).

Prosodic level. In the *map tasking* coding scheme (Anderson et al., 1991), the set of possible dialogue acts, whose investigation is relevant to the link between prosodic and discourse structures, corresponds to roughly 16 possible moves in the map task (Stirling et al., 2001). On the contrary, current trends in corpora which document a huge variety of socio-linguistic and pragmatic domains, show that the set of possible speech acts includes as many as 80 categories which are distributed all over the corpus variation (Firenzuoli in preparation). Of course the inductive data on the link between prosody and speech acts have a severe limitation in map tasking and need to be documented in general corpora.

The study of prosody needs natural speech variations for many reasons. For instance, quite surprisingly we noticed that thematic prosodic structures (*topic*/*prefix intonation* see. 't Hart et al., 1990), largely characterised formal texts, while the so called *comma intonation* (*appendix/suffix* 't Hart et al., 1990) strongly correlates to everyday dialogues (Tizzanini in press).

Middle length of utterances (MLU). The demarcation of the *utterances*, is an essential data for the interpretation of natural speech and it turns out that such tagging level allows the verification of important basic speech measurements (Biber et al., 1999). In recent works (see Tizzanini, in press; Rossi, 1999; Cresti, 2000, Moneglia, in press; Firenzuoli, 2000) has been verified, that MLU of texts marked by a strong degree of spontaneity (family conversations, country wakes, conversations among work colleagues and conversations among university students) systematically differs from MLU of formal texts (university lectures and radio interviews).

Fig. 1 shows that the MLU is almost constant all through the contextual variation with the significant exception of formal contexts, where we find a *iato*[2].

The systematic correlation between type of contexts and MLU allow a strong a quantitative prevision on the internal structure of the texts defining the probability of the possible length of the utterance in each domain.

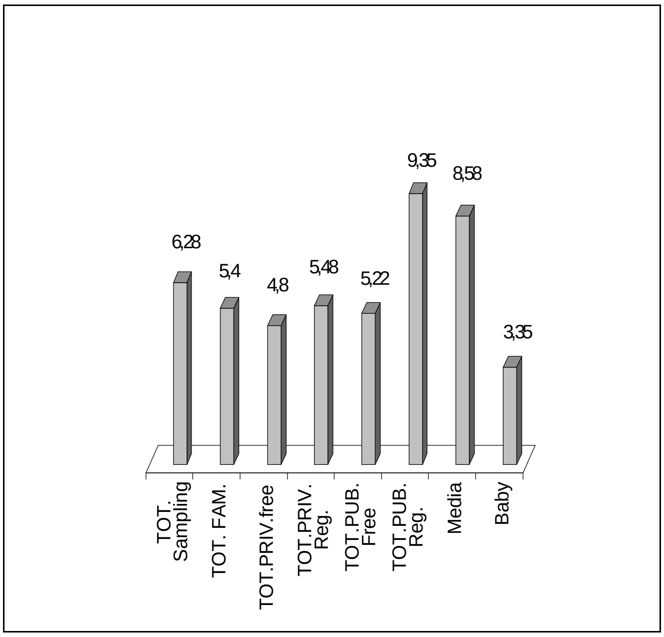


Figure 1. Middle Length Average in text typologies

The representation of spontaneous speech must therefore necessarily represent spoken text variation.

2.2. The comparability issue

The central problem which a multilingual corpus of Spontaneous Speech must solve is the question of comparability between different language resources in the domain of Spontaneous Speech.

Comparability in large Written Language Corpora was tested in two forms:

- Parallel corpora (for ex. CRATER and EUROROM)
- Corpora of the same type or of the same specialised field in several languages[3].

Clearly, with respect to the task of collecting Multilingual Spontaneous Spoken Language Corpora, only the second alternative is, in principle, available. As a matter of fact, it is impossible to realise parallel corpora without losing the spontaneity characteristic (Character c). In the domain of speech, parallel corpora are possible only in reading and in acting performances.

Comparability is quite easy to pursue with respect to resources based on the selection of a specific semantic domains (telephone information, health information, map tasking etc.) “people in the same controlled situation doing the same things”. However such resources are acquired in a restricted series of situations and are submitted to elicitation parameters (limited contexts) and therefore lack the main character of spontaneous speech (character d).

If we assume that the representation of spontaneous speech must necessarily represent spoken text variation, in a multilingual resource the more variability is represented in each language resource, the more the language resource is difficult to compare with the other resources and comparability is a function of the application of variation parameters.

2.2.1. C-ORAL-ROM sampling

The definition of significant variation parameters is, therefore, a basic step towards the development of a comparable LR of spontaneous speech.

A long tradition of socio-linguistic studies (see Bilger, 1997; Labov, 1966; Biber, 1998; Berruto, 1987; Gadet, 1996) has frequently dealt with the significance of "socio-situational parameters": 1) Socio-linguistic (age, education, occupation, sex); 2) semiological (monologue, dialogue, conversation); 3) sociological (family, public); 4) transmission (face-to-face, transmitted); 4) gender. In practice[4].

C-ORAL-ROM sampling of the four romance languages resources is based on the following set of variation parameters that constitute the semiological and sociological structure of the corpus:

(a) Dialogical structure (monologues, dialogues, conversations);
(b) Social domain of use (family; private life, public life, media productions.)
(c) Genders variation
(d) Formal vs. informal distinction
(e) Speaker parameters (Age, Sex, Education, and Occupation).

In C-ORAL-ROM, which has a quite limited dimension, such parameters are not uniformly verified through the all variation. That should be of course much better. In particular the use in the sampling strategy of the *formal* / *informal* partition, which is absent in the Dutch corpus, allow to restrict the number of parameters under investigation reducing the set of possible variations, with low damage for representation purpose. In particular text gender variation is the main criterion applied in the formal part, while social contexts of use and dialogue structure variation are the variation parameters systematically

[2] From Cresti, 2000. Legend: TOT.Sampling: *total data of sampling*; TOT.FAM.: *family typology*; TOT.PRIV.free: *private fee typology*; TOT.PRIV.reg.: *private regulate typology*; TOT.PUB.free*: public free typology*; TOT.PUB.reg*: public regulate typology*; Media: *media typology*; Baby: *baby talk typology*.

[3] The prototype example is the relation between the Brown Corpus (early 60’s, Brown University USA) and LOB Corpus (Lancaster/Oslo/Bergen, 1970) which realise together a comparable sampling of American English and British English.

[4] *The Spoken Dutch Corpus* (also under constitution at present) is a concrete example of the use of such parameters in corpus design (documenting the Netherlands and the Flanders). We were not aware of the corpus design of the Dutch Corpus when the C-ORAL-ROM project was prepared (1999), but when sampling was decided (January 2001), its structure at http//lands/let.kun.nl/cgn/edesign.htm, confirmed the overall criteria.

adopted for the informal part, where on the contrary genders variation is not strictly defined as a parameter.

C-ORAL-ROM does not represent dia-topical phonetic variations. In a multilingual collection dia-topical limits for each language must be established. Corpora are collected in Continental Portugal, Central Castilia Spain, Southern France, Western Tuscany, and are intended to represent some possible standard, rather than all the varieties of pronunciation, which need collections of interlinguistic corpora with a wide dia-topical variation[5]. Therefore, each corpus does not represent phonetic variation, but rather is expected to demonstrate a sufficient variation across language uses for at least studying *communicative acts*, *lexicon*, *syntax* and *prosody*.

The main choices adopted in C-ORAL-ROM for the representation of speech variability in four 300.000 word corpora are the following:

- splitting formal speech (50%) and informal speech (50%), variation ensuring a sufficient representation of dialogical Informal Speech (which is the resource with higher added value);
- selecting distinct criteria for sampling the *formal* and *informal* part of the corpus.
- defining a text weight (from 1500 to 3000 words for each text) that ensures both the possible appreciation of macro-textual properties and sufficient representation of the universe in each 300,000 word corpus.
- representing a variety of possible recording situations within the range of perception and intelligibility of the human ear [6].
- recording as part of the meta-data: a) Speaker characteristics; (gender, age, geographical region education and occupation); b) acoustic quality of the text.

The comparable Romance Spoken Corpus is identified by means of common Sampling criteria, and the same proportion each type in the four corpora: the following are the tables for the formal and informal part of each romance corpus in the C-ORAL-ROM resource.

Private /Family Context 113.000 words		Public context 37.000 words	
Monologue 33.000 w	Dialogue[7] 80.000w	Monologue 6.000	Dialogue 31.000

Table 1: Informal Corpora[8]

Formal in Natural Context 65.000 w	Formal in Media Context 60.000 w	Telephon 25.000 w
Political speech	News	Private Dialogues
Political debate	Meteo	Phone to call services
Preaching	Interviews	
Teaching	Reportage	
Professional explanations	Scientific Press	
Conferences	Sport	
Business	Talk show Political	
Law	Talk show Thematic Discussion	
	Talk show Culture	
	Talk show Science	

Table 2: Formal Corpora[9]

As a consequence of those choices, each corpus in the multilingual resource cannot be said to be comparable to the others with respect to specific semantic domains, but rather, with respect to the possible occurrence of spoken language structure/s at both syntactic and prosodic levels in a variety of possible significant contexts

2.2.2. Textual format

The four Romance Corpora have been transcribed or converted into standard textual (Gibbon et al., 1997).The format definition of spoken texts involves: 1) dialogue representation; 2) text co-ordinates; 3) prosodic tagging . The C-ORAL-ROM textual format is defined as an implementation of the CHAT architecture (Mac Whinney, 1994). Texts are divided into:

a) Heading, containing a definite set of meta-textual information
b) Text lines in orthographic transcription divided as follows:
c) vertically, in dialogic turns (introduced by a speaker label)
d) horizontally, by prosodic parsing and utterance limit, representing terminal and non terminal prosodic breaks of the speech continuum.
e) Dependent tiers for context information and possible morpho-syntactic tagging.

The C-ORAL-ROM textual Corpus will turns tagged with respect to: a) utterances corresponding to speech acts (Austin, 1962; Cresti, 1994 and 2000); b) prosodic parsing

[5] This limitation is quite severe for Italian, where local varieties may strongly diverge from the standard (De Mauro et al., 1993)

[6] The sound files of the acoustic database are set on a quality scale (recording, volume, voice overlapping and noise) and are comparable with respect to it. The quality scale extends from the highest level of clarity of the voice signal to low levels of acoustic quality. The quality is gauged spectrographically.

[7] At least 23.000 conversations with more then two participants

[8] 10 long sample 4.500w; at least 64 short sample, 1500w; 7.500w collections of very short dialogues in public context

[9] 2 or 3 sample for each gender of 3000 words average with only one small sample for News and Meteo.

of each utterance ('t Hart et al., 1990); c) words vs. word fragments distinction; d) overlapping.

3. Multimedia

The definition of the text to speech interface in C-ORAL-ROM is based on the idea that the access to acoustic information in a multimedia corpus (*alignment*) must go hand in hand with the representation of prosody. Such a method can be proposed as a possible standard for storing oral language in multimedia and multi-modal language resources. C-ORAL-ROM will ensure simultaneously:

a) tagging with respect to prosodic parsing & action values of the all textual information
b) acoustic analysis with special functions for F0 detection on low quality signal.
c) utterance based text - speech alignment

3.1. Acoustic format

C-ORAL-ROM comes from the reuse of previously established resources recorded with various analogue or digital equipment and from new recordings. The following are the requirements for the acoustic format:

Format: mono or stereo .wav files (Windows PCM), Sampling frequency: 22050Hz, 16 bit

Recording and storing process for old Analogue recording: directly derived in wav files (20.050 hz 16 bit) from the original analogue tapes through a standard sound card (Sound Blaster live or compatible) with a professional sound editor.

Recording and storing process for new recording:

a) *dialogues*: stereo DAT or minidisk recording (44.100Hz) with two unidirectional Micro-phones, converted into mono or stereo .wav files (Windows PCM, 22050Hz, 16 bit) via SPDIF port of a standard sound card (Sound Blaster live or compatible) with a professional sound editor

b) *conversations with more than two participants*: mono DAT or minidisk recording with cardioid or omni-directional microphone converted into mono .wav files (Windows PCM, 22050Hz, 16 bit) via SPDIF port of a standard sound card (Sound Blaster live or compatible) with a professional sound editor.

3.2. WinPitchCorpus

In synthesis the function of the Align Programme in C-ORAL-ROM is to orient the sound signal exploitation allowing, not only the transit from text to sound, but also, from text to sound analysis.

Text-speech alignment and acoustic analysis are ensured through the speech software WinPitchCorpus implemented in the C-ORAL-ROM Project. WinPitchCorpus (see http://www.winpitch.com) is a general purpose speech analysis tool working under Windows 2000/XP with many functions devoted to the alignment and annotation of large corpora. In particular Text-speech aligner tool, is based on a user adjustable speech slow-down process, in order to easily select text by mouse clicking as slowed speech is perceived, and automatically building of an aligned text database (up to 8 layers of text annotation and alignment). It incorporates a mouse driven file segmentation tools, with precise time adjustment on on-screen speech spectrogram and prosodic parameters display. This allows a fast and precise segmentation of both long prosodic units (utterances) and small speech units such as syllables or phones. Among its numerous features:

a) Recording, and playback of long signals (memory limited) at standard sampling rates (8,000 Hz, 11,025 Hz, 16,000 Hz, 22050 Hz, 32,000 Hz, 44,000 Hz and 64,000 Hz) in mono or stereo mode, at 8 bits or 16 bits encoding;
b) Standard black and white and color spectrogram of any part of the speech signal, with 3 distinct zooming tools (down to 1 sample resolution), 8 levels of bandwidth and 8 available analysis windows, 3 hierarchical levels of zooming;
c) Powerful fundamental frequency and intensity analysis (3 standard methods – spectral comb, AMDF, harmonic selection) with all user adjustable parameters;
d) Prosodic morphing, user graphically defined modification of the prosodic parameters of natural speech (fundamental frequency, intensity, syllable duration, pauses);
e) Easy insertion of text, bookmarks, comments. User defined speech section highlighting;

WinPitch also complies with the MDI Windows standard (Multiple Document Interface), and allows all functions to be concurrently applied to multiple speech signals.

3.3. Alignment units

The storage of spoken language resources should be based on the selection of a natural alignment unit. In C-ORAL-ROM all the textual information is tagged simultaneously with respect to prosodic parsing and utterance limits, therefore each prosodic unit corresponding to an utterance can be easily and directly aligned to its acoustic counterpart, thus ensuring a natural and meaningful text - sound correspondence.

This step is quite controversial at two levels. It implies on one side that the notion of utterance should be preferred to other possible linguistic notions as a natural alignment unit and that, on the other side, the criteria for the identification of utterances in a spoken language corpus are reliable.

As far as the first question is concerned *word based alignment* (that has been preferred for example in the *Spoken Dutch Corpus*) has low significance in spontaneous speech, and it is hard to be pursued for prosodic reasons. In spontaneous spoken language words are co-articulated in prosodic units and the acoustic effect of a word based alignment is perceptively unnatural.

Moreover, the alignment becomes significant from a linguistic point of view once it is defined with respect to a compositional linguistic domain, that is ranked over the word level description. Therefore the alignment problem is linked to the definition of the language structure in the spontaneous spoken language domain.

The C-ORAL-ROM approach is based on the idea that while Written language is characterised by a textual organisation based on syntax, Spoken language is mainly characterised by utterances, having a pragmatic nature and corresponding to *communicative acts* (Quirk, et al., 1985; Biber, et al., 1999; Cresti, 2000). In facts *sentence* based (or *clause* based) alignment turns out strongly *underdetermined* in spontaneous spoken texts. For example, considering textual information, the following dialogic turn is apparently one sentence:

*SEC: che macchina l'è codesta Punto
%tra: [which car is this Punto]
%sit: in a garage, a secretary looking for some
information for fixing a car

On the contrary the relevant acoustic information reveals that the dialogic turn is compound by two utterances, which can receive the following paraphrases: "I' m wandering which kind of car is this one. Is it a Punto ?" .

In other words the two utterances define two meaningful units for a linguistically relevant alignment, while the syntactic approach will lead to a meaningless alignment from a linguistic point of view.

Therefore textual information does not determine a significant alignment unit in spoken language, where not textual information is frequently required and, as the previous example shows, a meaningful alignment unit may not have a clause or sentence structure. So syntactically based alignment is at least underdetermined.

The relevant linguistic events (utterances) must be selected in the speech continuum through the full appreciation of the acoustic and pragmatic information. This conclusion, however, leads us to the second question.

A definition of utterance *as a speech continuum from one silence to one silence* has been frequently proposed, even as an objective mark allowing the automatic detection of utterance limits on the acoustic signal. However it must be stressed that the notion of utterance as a speech continuum from one silence to one silence is *together too week and too strong* for the representation of natural speech and therefore it does not allow any prevision on spoken corpora segmentation. In particular we can highlight the following:

a) segments of sound wave that are between two sound breaks frequently are not utterances;
b) in spontaneous speech frequently utterances start and/or stop with no break in the sound wave.

The quantitative relevance of both properties in spontaneous speech cannot be stated with precision but only guessed. For ex from 20% to 50% of utterances (depending on the text gender) of spontaneous speech corpora have a topic unit (Signorini, 2001). A topic cannot be an utterance but is frequently in between two silences (see. the example below).

Similarly the second utterance of the previous example is not preceded by a temporal break. The frequency of new utterances that start with no temporal break (or less than the voiceless part of a stop consonant) has not be counted but it is of course a very high percentage in spontaneous speech.

In conclusion the notion of utterance as a speech continuum from one silence to one silence is *together too week and too strong* for the representation of natural speech and moreover it does not allow any prevision on spoken corpora segmentation even from a statistic point of view.

3.3.1. Prosodic tagging

The segmentation of spoken texts into utterances corresponding to speech acts can be based on prosodic properties that are highly identifiable at the perceptual level.

In C-ORAL-ROM the prosodic tagging of the transcribed text it is not a *transcription* of the intonation, as for example ToBi, or MARSEC. that specifies the intonation profiles according to a phonological typology.
In C-ORAL-ROM prosodic tagging specifies on the text each perceptively relevant prosodic break in the speech continuum (prosodic parsing):

a) *Tone units with a not terminal contour,* reported every time a non terminal prosodic break can be perceived in a word sequence by a competent speaker: / (single slash)
b) *Terminal contours* (utterance limit) reported every time that a terminal prosodic break can be perceived by a competent speaker): // ? (double slash or question mark)

The previous example will be transcribed as follow in C-ORAL-ROM:

*SEC: che macchina l'è / codesta // Punto ?
%tra: [which car is / this // Punto ?]
%sit: in a garage, a secretary looking for some
information for fixing a car

Crucially terminal breaks indicate the prosodic completion of each utterance.

The definition of *utterance* in C-ORAL-ROM is theoretically defined. Given that intonation parses the speech continuum with relevant F0 movements we assume that the identification of utterances in the sound continuum is linked to the detection of perceptively relevant F0 movements. Also very traditional studies of prosody have noted that there is no such thing as an utterance without a profile of *terminal intonation* (Karcevsky, 1931; Crystal, 1975). Therefore the systematic correlation between terminal contours and utterance limit is an efficient heuristic method for speech segmentation.

However, at the theoretical level, we must consider that perception is highly sensitive to voluntary F0 variation ('t Hart et al., 1990) and that every utterance in spoken language from one side is the voluntary accomplishment of a speech act (Austin, 1962) and from the other it is necessarily parsed in one or more tone units.

The background theory of the C-ORAL-ROM project (Cresti, 1994, 2000) links the two properties: the voluntary F0 variations do not simply scan the utterance, but rather express functional values that are necessary to the accomplishment of speech acts. For this reason the selection of textual units corresponding to an utterance can be based on prosodic properties. In particular, as we did in the previous example, it is possible to identify an utterance each time the prosody makes it possible to perceive the completion of a speech act; i.e. intonation permits the pragmatic interpretation of the text (*Illocutionary criterion* Cresti, 1994, 2000). The illocutionary criterion has been successfully applied to both the corpora of Adult Spontaneous Speech and Infant

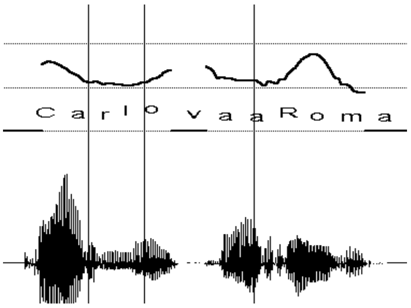
Speech allowing their tagging in utterances (see. Moneglia & Cresti, 1997).

The identification of functional values for prosody is also in some sense traditional (Bally, 1950; Halliday, 1985). For example it has been noted that, within the possible tone units, the tone information which enables one to identify the illocution, or modality, of the utterance lies in a specific scansion unit (Martin, 1978).

The theoretical approach we are referring to systematically links the study of such values to the study of spontaneous speech. The melodic pattern which scans an utterance can be simple (composed of a single tone unit) or complex (in which case it is made up of two or more tone units linked melodically together).

Non terminal tone units corresponds to the scanning of an utterance by means of a complex pattern: the type of which is discriminated at the perceptual level on the base of its form (*intonation pattern*, ' t Hart, et al., 1990). In principle each perceptively relevant tone unit conveys a specific functional value (*informational patterning*; see Cresti, 1994; Crest & Firenzuoli in press). For example the first tone unit of the following utterance is a Topic (*prefix* contour) and is followed by an information unit (with a *root* contour) allowing the identification of the illocutionary value of the utterance (*Comment*).

Carlo / va a Roma //
[Charles / is going to Rome //]



The results obtained on the basis of the application of the illocutionary criterion are crucially confirmed in the macro-syntactic theory of spoken language (Blanche-Benveniste, 1990) for which the syntactic *noyau* coincides with the tone unit bearing the illocutionary value.

C-ORAL-ROM Corpora represent the variety of speech acts performed in everyday language use and enables the description of their prosodic and syntactic structure in the four Romance Languages, from a quantitative and qualitative point of view.

4. References

Austin, L.J., 1962. *How to do things with words*, Oxford: Oxford University Press.

Anderson, A., M. Bader, E. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, R. Weinert, 1991. The HCRC map task corpus. *Language and Speech*, 34: 351-366.

Bacelar do Nascimento, F., (ed.), 2001. *Portugues falado: varietades geograficas e sociais*, Lisboa: CLUL & Instituto Camoens.

Bally, C.,1950. *Linguistique générale et linguistique française*, Berne: Francke Verlag.

Berruto, G., 1987. *Sociolinguistica dell’Iitaliano contemporaneo*, Firenze: NIS.

Biber, D., 1988. *Variation across speech and writing*, Cambridge: Cambridge University Press.

Biber, D., S. Johansson, G. Leech, E. Finegan (eds.) 1998. *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Biber D., S. Johansson, G. Leech, S. Conrad, E. Finegan (eds.) 1999, *The Longman grammar of spoken and written English*. London: Longman.

Biber D. 2000. Corpus based analysis of grammar: variability in the form and use of English complement clauses. In M. Bilger (ed.), *Corpus, Methodologie et applications linguistique*. Paris: Champion, 224-237.

Bilger, M. , 1997. Corpus de portugais & d’espagnol. *Revue de l’Association Français de linguistique appliquée*. 2:27-38.

Blanche-Benveniste, C. (ed.), 1990. *Le français parlé: ètudes grammaticales* . Paris: Editions du CNRS.

Blanche-Benveniste, C. (ed.) in press. *Corpus du Français parlé. Echantillonages*. Paris: Champion.

Cresti, E., 1994. Information and intonational patterning in Italian. In B. Ferguson, H. Gezundhajt, Ph. Martin (eds.) 1994. *Accent, intonation, et modéles phonologiques*. Toronto: Editions Mélodie. 99-140.

Cresti, E., 2000. *Corpus di italiano parlato*, vol. I- II, CD-Rom, Firenze: Accademia della Crusca

Cresti, E., V. Firenzuoli in press. L’articolazione informativa topic-comment e comment-appendice: correlati intonativi, In *Atti delle XII° GFS (Macerata 15 Dicembre 2001*). Macerata: Università di Macerata Press.

Crystal, D., 1975. *The English tone of voice*, London Edward Arnold.

De Mauro, T., F. Mancini, M. Vedovelli, M. Voghera 1993. *Lessico di frequenza dell' italiano parlato* Milano: Etass Libri.

Firenzuoli, V., 2000. Nuovi dati statistici sull’italiano parlato. *Romanische Forshungen*, 13: 213-225.

Firenzuoli, V., in preparation. *Repertorio dei profili intonativi di valore illocutivo in un corpus di italiano parlato*, Ph.D. thesis, Firenze: LABLITA.

Gadet, F., 1996. Variabilité, variation, varieté: le Français d’Europe. *French Language Studies*, 6:45-58.

Gibbon, D., R. More, R. Winski (eds.), 1997. *The handbook of Standards and Resources for Spoken language Systems*. Berlin: Mouton & de Gruyter.

Givon, T. (ed.), 1979. Discourse and Syntax. In Givon T. (ed.), *Syntax and Semantics*, vol. 12. New York: Academic Press

Halliday, M., 1985. *Spoken and written languages*. Oxford: Oxford University Press.

’t Hart, H., R. Collier, A. Cohen, 1990. *A perceptual study on intonation. An experimental approach to speech melody*. Cambridge: Cambridge University Press.

Karcevsky, S., 1931. *Sur la phonologie de la phrase*, in Travaux du Cercle linguistique de Prague, IV.

Labov, W., 1966. *The social stratification of English in New York City*. Washington D.C.

Lavacchi, L., C. Nicolas, 2000. *Dizionario Spagnolo Italiano*, (CD-rom Edition). Firenze: Le Lettere.

MacWhinney, B., 1994. *The CHILDES project: tools for analyzing talk*. Hillsdale: Lawrence Erlbaum Associates.

Martin, Ph. 1978. Questions de phonosyntaxe et de phonosémantique en Français, in *Linguisticae Investigationes*, II: 93-126.

Miller, J. & Weinert, R. 1999., *Spontaneous Spoken language*, Oxford: Clarendon Press.

Moneglia, M., in press. I corpora dell'italiano parlato di LABLITA: Criteri di costituzione, unità di analisi e comparabilità dei dati linguistici orali. In E. Burr (ed.), *Atti del VII° Convegno internazionale SILFI*. Pisa: Cesati.

Moneglia, M., E. Cresti, 1997. Intonazione e criteri di trascrizione del parlarto, in U. Bortolini E. Pizzuto (eds), *Il progetto CHILDES Italia*, Pisa: Del Cerro.

Miller, J., R. Weinert, 1999. *Spontaneous Spoken language* . Oxford: Clarendon Press.

Quirk, R., S. Greenbaum, G. Leech, J. Svartvik, 1985. *A comprehensive Grammar of the English Language.* London: Longman.

Rossi, F., 1999. *Le parole dello schermo*. Roma: Bulzoni.

Signorini, S., 2001. *Caratteristiche sintattiche e frequenze dei topic in un corpus di parlato italiano*, Tesi di Laurea, Univerity of Florence.

Stirling, J., I. Fletch.r, R. Mushin, L. Wales, 2001. Representational issues in annotation: Using the Australian map task corpus to relate prosody and discourse structure. *Speech Communication*, 33:113-134.

Tizzanini, G., in press. L'articolazione dell'informazione. Dati quantitativi di un corpus di italiano parlato. In E. Burr (ed.), *Atti del VII° Convegno internazionale SILFI*, Pisa: Cesati.

The Spoken Dutch Corpus, http//lands/let.kun.nl/cgn/edesign.htm

WINPITCH, 1995. http://www.winpitch.com