Multilevel corpus analysis: generating and querying an AGset of spoken Italian (SpIt-MDb).

Renata Savy*, Francesco Cutugno^, Claudia Crocco*

*Dept. of Linguistic and Literature, University of Salerno, Italy
^ Dept. of Physical Sciences, University "Federico II", Naples, Italy
*Via Ponte Don Melillo, 84084, Fisciano (SA)- Italy
^Complesso Universitario di Monte S.Angelo, Via Cinthia, 80126, Napoli, Italy
`rsavy@unisa.it, cutugno@na.infn.it, ccrocco@unisa.it`

Abstract

In this paper we present an application of AGTK to a corpus of spoken Italian annotated at many different linguistic levels. The work consists of two parts: a) the presentation of AG-SpIt, a toolkit devoted to corpus data management that we developed according to AGTK proposals; b) the presentation of corpus' structure together with some examples and results of cross-level linguistic analyses obtained querying the database (SpIt-MDb). As this work is still an ongoing investigation, results must be considered preliminary, as a 'demo' illustrating the potentiality of the tool and the advantages it introduces to validate linguistic theories and annotation systems. Currently, SpIt-MDb is a linguistic resource under development; it represents one of the first attempts to create an Italian corpus labelled at various linguistic levels (from acoustic/sub-phonetic, to textual/pragmatic ones) which can be queried in the interrelations among levels.

1. Introduction

As stated by Bird & Harrington's Editorial (2001), annotated corpora provide a link between speech and language technology research, and linguistics. Corpora are growing in dimension and their future usability will depend on the possibility to have tools properly projected to support their generation, annotation, management and querying. Particular attention has to be paid in the development of well structured databases in which information retrieval could be easy for tasks as multilevel linguistic analysis.

1.1. The AGTk theorical framework

It is widely accepted that Annotation Graphs (AG) introduced new important tendencies in the field of speech corpora annotation. The standard proposed by Bird and colleagues (Bird & Liberman, 2001; Maeda et al., 2002) contributes to the solution of many problems affecting formal descriptions of multi-level annotation framework as lack of linearity, i.e. the missed temporal correspondence between speech events and linguistic units, difficulties in the definition of a proper linguistic hierarchy among annotation levels, unification of different annotation standards and/or cross-reference among different metadata systems.

Even if the question of which could be the optimal querying system remains unanswered, information retrieval is widely and easily practicable both using xml querying tools like XQL or XQuery and SQL version of the AGsets, especially when the complexity of the query falls into a set of well defined use cases within the AG formal framework.

A further important feature that allows a wide and fructuous use of the AG standard for annotation is the possibility to import previously collected data, formatted according to different specification, into AG-xml. The AG Toolkit provides users with many primitives for the direct conversion from many usual formats of annotation to AG, however generation of new scripts, specifically produced to include in the graphs new data with unusual structure, is not a difficult task.

1.2. AGTk and corpora of spoken Italian (SpIt)

Spontaneous speech corpora collection has become a major task in Italian linguistic research only in the recent years. Several projects, ranging in the last decade and involving many Italian universities, have lead to the production of a considerable amount of annotated speech materials, in order to offer the scientific community the data and the tools necessary to carry on multi-level analysis on spoken Italian. The annotation guidelines and the definition of data architecture have sometimes been refined from a project to the following. In other cases, new data have been produced that cannot be directly integrated with the previously collected ones. Given this not unusual scenario, the formal framework proposed by Bird and colleagues seems to be ideal to prevent further data dispersion and to guarantee compatibility between already collected data and annotated materials that will be added in the future to the Italian repertoire of speech corpora.

In particular, in our research, we decided to experience AG method on a small sample of speech materials collected and updated in various phases within the frame of several national projects[1]. In the *CLIPS* project (Corpora e Lessici di Italiano Parlato e Scritto), a large corpus was partially coded at different segmental levels (i.e. acoustic, phonetic, phonological, lexical). The corpus contained a great amount of task-oriented dialogues (i.e. Map Task and Spot the Difference dialogues; Anderson et al. 1992, Pean et al.,1993). In the *IPar* project (Italiano Parlato), one of the CLIPS dialogues was provided with other annotations, for instance at the syntactic, prosodic, and information levels (Albano Leoni & Giordano 2006), in order to get a better comprehension of the linguistic dynamics of speech through a cross-level analysis. Further

---

[1] The projects (funded by the academic institutions involved, the Italian government and/or the EU) were co-ordinated by the University of Naples. The corpus will be soon available free at *www.clips.unina.it*.

annotations (at the syllabic, morphosyntactic and pragmatic levels) of other dialogues have been recently provided, or are currently in progress (Castagneto et al, forthcoming; Savy & Voghera, forthcoming).

Altogether, we have 3 dialogues and 13 annotation levels (and others can be added). All annotations follow specific criteria, reflecting theoretical assumptions and responding to coding standards. Furthermore, each annotation level selects its own base unit, depending on linguistic factors (see below, § 3.1)

The pilot study presented here is part of a third project, called 'Parlare italiano' (Voghera & Cutugno, in this volume), and it constitutes the basis for further implementations of the database through multilevel labelling of a larger sample[2]. The project aims at creating, on the Internet, a portal devoted to spoken Italian. The web site will contain corpora, tools for the automatic speech analysis and quantitative linguistic studies. SpIt-MDb (see below) will be subsequently implemented and developed, and it will be one of the tools available through the portal.

2. Software development

On the basis of this framework we implemented a new tool for the generation of AGsets, the visualization of speech data together with relative annotation and information retrieval features. The AG-SpIT (AG-Spoken Italian Tool) is an integrated software environment presenting an architecture based on five main modules:

1) a *converter* which operates on annotations and labeling usually coded in TIMIT format, including temporal information (such as phonetics, syllabic, prosodic, etc.), and outputs a complete AGset in XML format;

2) a *modifier* able to add to the previously generated AGset annotations not directly time-dependent (such as PoS tagging, syntactic and pragmatic labelling);

3) a *parser* accepting as input the AGset and producing an equivalent SQL database (SpIt-MDb),

4) a modified version of *Wavesurfer,* in which annotation data included into the AGset can be showed in a multiple frame environment together with the speech waveform and with the typical signal analysis tools;

5) a *Query Generator* operating on the SQL database in order to retrieve and observe data correlations among the various linguistic levels of annotation available.

Next sub-sections will describe in details modules and the environment features.

2.1.   First module: TIMIT converter

The first module consists of a set of scripts to convert TIMIT files in a unique XML AGset. Each input file is usually associated to a level of annotation; in each file, couples of time marker are biuniquely associated to a label according to the format:

start time   *end time*   label

Temporal markers are not used as cross-reference among the various annotation levels, i.e. all temporal annotations are unbinded, furthermore levels are not expected to respect any alignment rule.

In our corpus, sub-phonic events, phonic sequence, word sequences (orthographic and "standard phonological" transcriptions), intonational labelling, etc. are examples of TIMIT annotations. This module uses the AG library. A user-guided interface is built in order to facilitate the conversion to people not managing software codes. The resulting AGset is completely compatible with the AG standards.

2.2.   Second module: AGset modifier

The second module adds new annotations to the AGset created using the previous one. It is specifically thought for XML annotation deriving from corpus analyses that are, at least in principle, not time-dependent (as syntactic, morphosyntactic, pragmatic, etc.). Our aim is to attempt a (sometimes problematic) temporal alignment between speech signal and linguistic structures listed into these annotations.

The module accepts as input the XML annotation file and the AGset, and outputs the updated AGset. Temporal information are usually retrieved from the sequence of word markers already present in the AG file linking to TIMIT files containing words. However, adjustments are needed when the correspondence between linguistic constituents and word sequences is not biunique (as often it happens in syntactic annotation), and when the constituent domain involves only word parts (like morphemes in morphosynctactic annotation; see below §3.1). A report is produced at every step of the update process. The report has the duplex function to indicate errors and/or misalignments between the different levels of annotations that the program compares, and to list all the cases in which problems not depending on the method arise.

According to Cotton & Bird (2002) recursive paths typically produced in syntactic annotations have been represented within the AGxml document making use of equivalence classes.

An ambitious future aim is the creation of a universal converter able to "read" the local grammar of the annotation file (for example using its related DTD) and to parse it on the basis of a set of user requirements.

2.3.   Third module: AGset parser

The third module, making a wide use of AGLib, parses the XML AGset and produces an SQL database whose conceptual model respects the standard proposed by Bird. The DBMS chosen is the open source MySQL. It offers, among other useful tools, an IDE in which it is possible to execute SQL queries, and it has primitives for Tcl-Tk for the development of database management software. Connections to database, multiple access to different sub-corpora, back-up procedures are directly manageable within the software integrated environment.

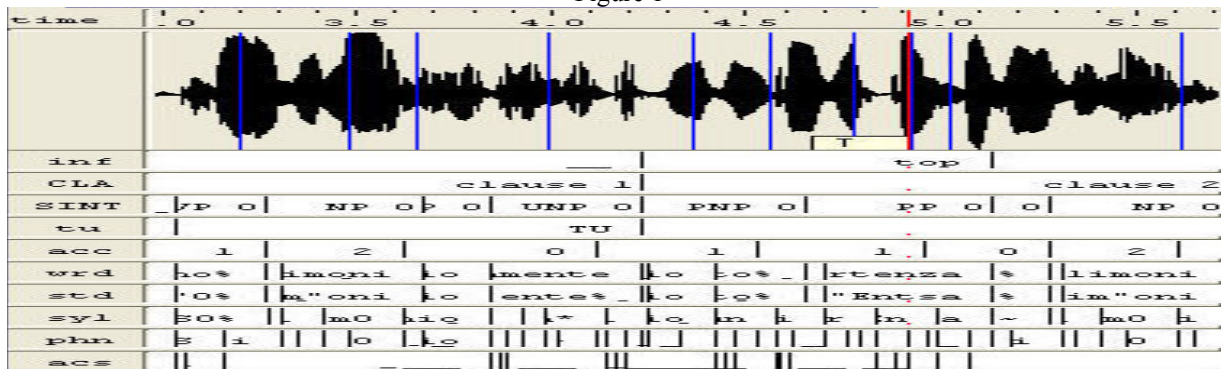[2] The project is co-ordinated by the University of Salerno.

Figure 1

Figure 2

2.4. Fourth module: graphic and audio environment

The fourth module is based on the core of Wavesurfer integrated with a set of tools that allow us to perform various operation on the AGset: we can select one or more AG from the AGset, visualize relative acoustic waveform, choose which annotation level we want to add into the panes, visualize instant event pointing directly to the waveform, and so on. Multiple audio tracks and relative annotations can be contemporarily presented to the user. All the speech analysis tools available via the Snack library in the Wavesurfer context are available in our environment too (see Fig. 1).

2.5. Fifth module: query generator

AG data can be accessed by means of a guided user interface integrated into the AG-SpIt environment. The Query Generator requires the filling of a simple form; the resulting query produces a request expressed in a controlled form of natural language. In other words, the user expresses a query composing a sentence in Italian; all the constituents of this sentence are driven by the interface; the semantic content of the resulting utterance reproduces an arc-pattern command for the selection of the relative corpus portion, respecting the given requests. Arc-patterns are the prototypical form of query in AG framework: they provide the user a way to select linguistic phenomena taking into account the timeline development of the acoustic speech dynamics. Queries based on arc-patterns, in their simplest form typically appear as:

First step:
select the start and end instants (anchors), *in a specified annotation level*
where the label field contains a given string
and

show other annotations
[included] | [overlapped] | [coincident] | [etc]
with the previous one

Second step (*Kleene closure):
Once you have obtained the list of linguistic phenomena required in the first step,
show all sequences of non-specified length of labels of any other level of annotation included in every previously selected item

These queries would be solved in a relatively easy way using AGQL (Bird et al., 2000) but, unfortunately, no practical implementation for this standard is presently available. Alternatively, the use of AG -> SQL conversion presents many limitation especially for what the *Kleene closure concerns.

Our system allows the formulation of queries as described in the first step, intersecting two different levels of annotations on the basis of their relative temporal position over the time line.

This kind of query outputs a list of labels and their relative temporal extensions (arcs). Because of SQL limitations, second step is then possible only for one arc at time and for only one further level: the interface allows single line selection in the output table and the indication of the level of further interest.

All the query results can be finally exported towards other applications like SQL tables or MS-Excel.

3. Linguistic applications

This section deals with several linguistic applications of AG-SpIT. Before proceeding with these examples, we first briefly illustrate how this tool can help to deal with several aspects relevant for studying spontaneous speech.

3.1. Preliminary remarks

Research has demonstrated that a multi-level approach can provide a better understanding of both spontaneous speech and linguistic organization in general. Speech exhibits deep interaction among different linguistic levels. The linguistic structure in spontaneous speech is strongly affected by the pressure of pragmatic strategies, including the information organization of discourse (see for instance Givòn, 1979; Halliday, 1967; Voghera, 1992; Cresti, 2000; Lambrecht, 1994). Moreover, research on speech has pointed out the crucial role prosody plays as an interface between segmental and higher linguistic levels and its involvement in syntactic and information phrasing, focus marking, and discourse structuring (Pierrehumbert & Hirschberg, 1990; van Donzel, 1999).

Obviously, the investigation of these cross-relations requires a complex managing of the data connected to different levels of analysis. AG-SpIt allows multi-level visualization and queries that make the management of cross-level linguistic data easy and intuitive. Furthermore, as mentioned in § 2, it is possible to build an open database, in which new data can be inserted at any time.

The potentiality of the tool was tested on a sample of three dialogues which were *autonomously* annotated at several linguistic levels (cf. §1.2). All of the coding formats used in labelling are directly managed by AG-SpIt. The *autonomy* of the annotation goes together with to the definition of *autonomous coding units*: i.e. base units used on each linguistic level were chosen in agreement with theoretical and technical premises relevant for that level. This means that coding standards and annotation units varied across the different levels considered.

Let us considered an example concerning with the annotation of the orthographic (WRD), morphosyntactic (PoS) and syntactic levels (SYNT). Although for all of these levels the relevant coding unit is in fact the "word", the definitions of "word" adopted are not the same among all annotations. In particular, the annotation of graphic words such as "prendimi" ("take me" or "take for me"; composed of a verb and a clitic particle) is different on the WRD and the PoS-SYNT level. On the former level the unit "word" corresponds to a "phonological word", while, on the latter, it corresponds to a "morphological word". Therefore, the word "prendimi" consists of one unit on the WRD level (i.e. one phonological word), while it consists of two units (i.e. two morphological words) on the PoS-SYNT level.

3.2. Cross-level queries

In this section we show the results of a number of queries made with SpIt-MDb. In these queries, prosody is taken as the point of reference, in accordance with its aforementioned role of interface among different linguistic levels. As seen in §2.1, The prosodic coding had three different levels:

1) the level of the intonation phrasing (TU boundaries);

2) the level of intonation target points, containing INTSINT tone labels (Hirst & Di Cristo, 1998);

3) the level of accent degree, containing prominence labels for the lexical words in the TU.

Table 1 shows the queries for two or more linguistic levels:

Table 1

Level involved in the query	Unities
intonation phrasing/lexical	T(one) U(nit) / word
intonation phrasing/rhythmic	TU /syl(lable)
intonation phrasing/segmental	TU / phon(etic segments)
syntactic/intonation phrasing	Cla(use)-ph(rase) / TU
intonation target points/ accentual /syntactic	Tone labels / prominence labels / syntactic ph
intonation phrasing/information phrasing/syntax	TU / Information Unit / syntactic ph

It is possible to select for each level a specific element to be put in relation with another element from a different coding level. For instance, on the syntactic level, it is possible to select the phrase or the clause as unit, and put it in relation with the TU or with the tonal targets (INTSINT labels).

3.3. Provisional linguistic results

A detailed analysis of the linguistic data which were gathered with SpIt-MDb, exceeds the goal of this pilot study. However, we obtained some provisional results that are promising and that show the tool's relevancy for linguistic analysis. Principally, our results indicate a) the type of data that can be gathered by the tool, and b) the cross-relations existing between different linguistic levels. In total six queries of increasing complexity were done, which are presented here.

The first three queries give results concerning minimal, maximal and average number of words, syllables and phonetic segments in the TUs. They are outlined in Table 2:

Table 2

TU length (elements)								
Syl			Words			Phon		
Max	Min	Avg	Max	Min	Avg	Max	Min	Avg
25	1	4,5	15	1	4	50	1	10

Interestingly, more than 50% of TUs coincide with a single word (frequently monosyllabic - max. 4 phones): all these cases are typical spontaneous speech phenomena such as disfluencies, self-repairs, false starts, filled pauses, fragmenting the prosodic continuity. However, short structures (3-10 syllables; 3-10 words) generally prevail over the longer ones.

From the fourth query, some remarkable results on the relations between clause and TU can be derived when confronting them with previous studies on the correspondence between TU and clauses (or other syntactic structures; see for instance Crystal 1969). According to Sornicola (1981), in spontaneous spoken Italian TUs correspond mostly to "phrase-structured sequences"; in Voghera (1990) and Caputo (1991), instead, more than 45% of the TU are said to contain one clause. For English, a similar result is found in Cruttendern (1986).

If we take the TU as pivot in the query (left side of table 3), we get a picture of the relations between syntactic and prosodic constituency which is partly different from the one outlined in the above mentioned works. In fact, only 24% of the TU correspond to a clause. Furthermore, almost 40% of the TU is co-extensive with a syntactic constituent smaller than a clause (i.e. one or more phrases). Our result is more in line with Sornicola (1981), even if the correspondence between phrase-structured sequences and TUs is found in less than half our sample.

Finally, only 2,6% of the TUs is composed of more than one clause.

Table 3

TU		Clause	
TU=cla	24.4%	cla>TU	30.4%
TU<cla	39.9%	cla=TU	46.5%
TU>cla	2.6%	cla<TU	4.9%
TU≠cla	33,1%	cla≠TU	48.5%

The partial discrepancy between our results and those reported by other authors can possibly be explained by the different definitions of tone units maybe used in these studies, or by the type of speech that was analyzed (task oriented *vs.* face to face dialogue etc.). However, despite this discrepancy, the crucial relation between clause and TU is confirmed by the results on the correspondence between clause and TU (right side of table 3). In fact, if we reverse the perspective and take the clause as pivot in the query, we see that more than 46% of the clauses correspond to a TU. A 30% of the clauses are furthermore spread over more than one TU; finally, only about 5% of the clauses are smaller than a TU. Overall, these results suggest that the TU frequently corresponds to a clause or to a syntactic constituent smaller than a clause (about 40% of TUs correspond to a phrase). On the other hand, the correspondence between TU and syntactic structure larger than a clause is rather unusual. Given this remarkable pattern, we believe it could be interesting to further investigate this result in future research.

The fifth query involves three linguistics levels. It shows the syntactic location of the main stress in the TU, and its possible correspondence with an intonation target point. The data show (cf. Table 4) that in most instances the main stress corresponds to the lowest (*Bottom*) or the highest (*Top*) pitch value in the intonation contour.

Table 4

	accent 2	B	T
Tot	*676*	**40.1%**	**32.0%**
VP (verbal phrase)	12.6%	8.9%	20.8%
UNP (unprocessed)	**34.5%**	41.0%	19.4%
REP (repetitions)	4.4%	4.8%	3.2%
PP (prepositional ph.)	**18.6%**	21.4%	19.4%
PNP (noun predicate ph)	4.4%	5.5%	4.2%
NP (noun phrase)	**25.4%**	18.5%	32.9%

The most frequent combination links the B to the main stress present in the TU. The combination of main stress and T is also frequent, while other combinations are relatively rare. Future research should focus on the position of main stress and intonation target point in the TU.

Main stress is predominately located on syntactically "unprocessed elements", i.e. isolated elements (spatial and temporal adverbs, conjunctions and interjections) that do not belong to any syntactic phrase (Cutugno & Voghera 2004). The high correlation between main stress and these phrase-external elements can be related to the prosodic autonomy of these "unprocessed elements". In fact, they are often realized as autonomous TU (almost 67% of the TU consists of an "unprocessed element"). As we said before, most of the "unprocessed elements" are spatial or temporal adverbs. Then, their syntactic and prosodic isolation can be easily related to their function in information structure. Spatial and temporal adverbs, in fact, occur at the beginning of the utterance as topics.

Main stress can also be located on NP, and PP; the combination between main stress and VP is less frequent. These data support the idea that in spoken Italian, from the prosodic point of view, VP is a weak constituent (see for instance Marotta 1984, Caputo 1991, Savy, 2001). However, in task-oriented dialogues, dehontic VP very often occur at the beginning of the clause and/or TU. When a VP is accented, it is usually associated with a T tone, very frequent in this position.

The sixth and last query involves syntactic, prosodic and informational levels. First, we explored the relations between information units, syntactic units and prosodic phrasing. Then, the relations between syntax and prosody (tonal targets and phrasing) in topic units was analysed more in dept, taking into account both the level of tonal target and the level of the prosodic phrasing. Our results show, in agreement with Giordano & Crocco (2006), that in most cases the TU and the information unit boundaries coincide (69%); or that, at least, one of the boundaries on one level corresponds to a boundary on the other level (27%). The complete lack of coincidence in the boundaries of the units of the two levels is rather infrequent. Therefore the data supports the hypothesis of a close relationship between information and intonation structure, although this should not be understood to be a complete identity. The analysis of the interrelation between information units and syntactic phrases adds other details to this analysis (see table 5).

Table 5

%	NP	PP	VP	UNP	tot
Topic	40,8	28,6	10,2	17,3	96,9
Comment	22,3	12,7	20,7	34,9	90,7
Appendix	19,2	29,5	19,2	26,7	94,5

Among the information units[3], the *comment* is the most frequent. From the syntactic point of view, it corresponds usually to a VP or NP, although other realizations (i.e. UNP and PP) are also possible. *Topics* correspond normally to non-verbal phrases (NP, PP and UNP): the correspondence with VP is rather unusual. Also the *appendix* is usually non verbal: it corresponds mostly to PPs; less commonly, it can be realised as NP or VP. From this analysis, a regularity emerges both in the relation between prosodic and information phrasing, and in the relation between informational role and syntactic realisation. Analysis of the syntax and prosody of topics gave rise to other interesting results. About 86% of the topics corresponding to a NP or a PP have a tonal target T (Top) on the head of the phrase (see Crocco & Savy, forthcoming). This can be explained by the fact that the topic units we considered occur at the beginning of the utterances (see above), where the speaker's pitch range and the melodic movements are usually expanded.

Table 6

Syntactic Head's Tone	Tone Type	%
High	T, H	31
Falling	TB, TL, TD, HB, HL, HD	42
Rising	UT, ST, LT, BH, BU, LH, SU	13
Rising-falling	UTB, LTB, LHB, UTD, LTL	8
Fallino-rising	TLH, TBH, HBU	3
Flat	S	3

---

[3] For a complete sketch of the theory we refer to, see Cresti (2000).

However, if we consider the global melodic trends found on those units (falling or rising movements), falling tones prevail, especially when the topic NP or PP corresponds to a TU. Rising tones are instead more frequent on topics smaller than a TU.

Further analysis of other types of speech could complete this sketch.

4. Conclusions

Even if this is only a preliminary study, encouraging results emerged.

First, our implementation of AG was effective in making also complex queries.

Then, even if SpIT still needs to be optimized, its reliability is already sufficient to apply in linguistic research: the queries provided promising linguistic results, integrating different levels of analysis. This integration is made possible by the AG: this tool does not request that the data are integrated *a priori*, during the creation of the corpus. The structure of the data can be instead defined *a posteriori,* in progressive steps during the creation of the database with AGTK.

5. Attributions

Although the authors largely cooperated, §§ 1.1 and 2 have to be attributed to F. Cutugno; §§ 1.2, 3.1, 3.2 have to be attributed to R. Savy; § 3.3 has to be attributed to C. Crocco. We wish to thank L. D'Anna, C. Del Sorbo and A. Del Prete for their contribution to software development.

6. References

AGlib: http://agtk.sourceforge.net

Albano Leoni F., Giordano R. (2006), (eds), *Italiano Parlato: analisi di un dialogo*. Napoli: Liguori.

Anderson A.H., Bader M., Gurman Bard E., Boyle E., Doherty G., Garrod S., Isard S., Kowtko J., McAllister J., Miller J., Sotillo C., Thompson H., Weinert R., (1992), The HCRC Map Task Corpus, HCRC, *Research Paper* 29, Univ. of Edinburgh.

Bird S., Harrington J. (2001), Editorial: Speech annotation and corpus tools, *Speech communication (special issue),* 32, pp.1-4.

Bird S., Liberman M. (2001), A formal framework for linguistic annotation, *Speech Communication*, 33:23-60.

Bird, S., Buneman, P., Tan, W.C. (2000), Towards a query language for annotation graphs. In *Proceedings of the Second International Conference on Language Resources and Evaluation,* pp. 807-814.

Caputo, M.R. (1991), *Aspetti prosodici dell'italiano parlato. Un'analisi spettroacustica,* Unpublished Master Degree thesis, Università di Napoli 'Federico II'.

Castagneto, M., Savy, R., De Leo, S. (forthcoming), Schemi di annotazione pragmatica a confronto: il test delle differenze. In *Proceedings of the II International Congress on 'Comunicazione Parlata'*, Napoli, February 23-25 2006.

Cotton, S., Bird, S. (2002), An Integrated Framework for Treebanks and Multilayer Annotations. In *Proceedings of the Third International Conference on Language Resources and Evaluation,* Paris: European Language Resources Association, pp. 1670-1677.

Cresti E. (2000), *Corpus di Italiano Parlato.* Firenze: Accademia della Crusca.

Crocco C., Savy, R. (forthcoming), Attraverso la prosodia: sintassi e informazione nel dialogo. In *Proceedings of the II International Congress on 'Comunicazione Parlata'*, Napoli, February 23-25 2006.

Cruttenden A. (1986), *Intonation*. Cambridge: Cambridge University Press.

Crystal, D. (1969), *Prosodic Systems and Intonation in English.* Cambridge: Cambridge University Press.

Cutugno F., Voghera M. (2004), AN.ANA.S.: analisi sintattica e annotazione XML a contatto. In *Proceedings of the National Conference 'Il parlato Italiano'*, CD-Rom, Napoli: D'Auria Editore.

Giordano R., Crocco C. (2006), Articolazione intonativa e articolazione informativa. In Albano Leoni F., Giordano R., (eds), *Italiano Parlato: analisi di un dialogo*, Napoli: Liguori.

Givòn T. (1979), From Discoure to Syntax. Grammar as a Processing Strategy, *Syntax and Semantics*, 12, pp. 81-112.

Halliday M.A.K. (1967), Notes on Transitivity and Theme in English, part 2, *Journal of Linguistics*, 3, pp.199-244.

Hirst, D., Di Cristo, A. (1998), (eds), *Intonation Systems. A Survey of Twenty Languages.* Cambridge: Cambridge University Press.

Lambrecht K. (1994), *Information structure and sentence form.* Cambridge: Cambridge University Press.

Maeda, K., Bird, S., Ma, X., Lee, H. (2002), Creating Annotation Tools with the Annotation Graph Toolkit. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, Paris: European Language Resources Association, pp. 1914-1921.

Marotta, G. (1984), *Aspetti della struttura ritmico-temporale in italiano,* Pisa: ETS.

Pean V., Williams S., Eskenazy M. (1993), The design and recording of ICY, a corpus for the study of intraspeaker variability and the characterisation of speaking styles. In *Proceedings of Eurospeech* 1993, Berlin, pp. 627-630.

Pierrehumbert J., Hirschberg J. (1990), The Meaning of Intonational Contours in the Interpretation of Discourse. In Cohen, Morgan, Pollack (eds), *Intentions in Communication.* MIT Press, pp. 271-310.

Savy R. (2001), L'interfaccia tra livelli di analisi del parlato: rapporti tra riduzioni sementali e schemi prosodici. In *Proceedings. of XXXIII Congresso SLI*, Roma: Bulzoni, pp. 309-328.

Savy, R., Voghera, M. (forthcoming), Sintassi in prosodia. In *Proc. of the II International Congress on 'Comunicazione Parlata'*, Napoli, February 23-25 2006.

Sornicola R. (1981), *Sul parlato*, Bologna: Il Mulino.

Van Donzel M. (1999), *Prosodic aspects of information structure in discourse.* Le Hague: Holland Academics Graphics.

Voghera M. (1990), Gruppi tonali e strutture sintattiche dell'italiano contemporaneo, *The Italianist*, 10:175-203.

Voghera, M. (1992), *Sintassi e intonazione dell'italiano parlato*. Bologna: Il Mulino.

Voghera, M., Cutugno, F. (in this volume), An observatory on Spoken Italian linguistic resources and descriptive standards.

Wavesurfer: www.speech.kth.se