

Overlaps in AVIP/IPAR, the Italian Treebank of Spontaneous Speech

Abstract. This paper will present work carried out on the 60,000 words Italian Spontaneous Speech Corpus which collapses together two corpora, the first one called AVIP, under national project API - the Italian version of MapTask, the other one called IPAR. We will present in particular the parser, to produce syntactic structures of overlapped temporally aligned turns. The paper will argue in favour of a joint and thus temporally aligned representation of overlapping material to capture all linguistic information made available by the local context. We argue that this is different from simply aligning on a time scale linguistic annotations on separate layers of representation. In fact, we are interested in collapsing overlapping linguistic material within a single representation in order to capture pragmatic inferences. This will result in a syntactically branching node we call OVL which contains both the overlapper's and the overleepee's material (linguistic or non-linguistic). The paper will also comment on data from our Written Text Treebank called VIT and compare them with the Oral Treebank.

Keywords. Corpus Annotation. Syntactic Structures. Semantic Structure. Prosody.

1. Introduction

This paper presents work carried out at the University of Venice for the creation of tools for the annotation of spoken Italian which allow the user to work in a format fit for the visualization of the results in multilevels representation in commercial browsers. The specific topic of this paper will be the characterization of overlaps along the lines of what has been done in MATE project and other international projects in progress like the MEETING project [1]. In the AVIP/API dialogues the quantity of overlapping speech is very high, as has been reported in [2;3]. At an international level, even though everybody agrees on the relevance of the phenomenon, there is no universal agreement on its representation from the linguistic point of view, in particular as concerns syntactic structure both at constituent and functional level.

In our approach we will investigate the place at which the overlap occurs, whether it is at a TRP, Transitional Relevance Place in which case also syntactic completion of some kind was involved - this was defined, for lack of a better linguistic basis, whenever a predicate was present [4;5]. Or in case it is in a Nontransitional Relevance Place: these two terms also hinged on the semantic completeness of the already uttered turn by the current speaker. Special care will be devoted to the related question of the recognition on the side of the overlapper of the incoming TRP and the irrelevance of the completion of the turn on the side of the current speaker, so that an interruption in the middle of the turn could still be interpreted as collaborative.

A viable syntactic-semantic definition of TRP will be formulated on the basis of the semi-automatically built treebank for the Italian National Project on spontaneous speech based on the Italian MapTask and the Spot Differences Dialogues [6].

This treebank will not only allow us to derive a precise definition of TRP but also classify all overlaps accordingly. This is due to a peculiar feature of the constituency which contains a distinct major constituent for overlaps, called OVL, where are located both the overlapped and the overlapping linguistic material in the actual location in the turn in which the interruption took place.

As will be described in detail below, the problem of the annotation of overlaps has required a new coding of all the corpus AVIP/API in order to recover the temporal alignment of the phenomenon under study and to collapse the linguistic material in a single layer of representation, preserving though the syntactic and semantic information associated to each segment. This is totally different from what is being proposed by work carried out both in Annotation Graphs [7], Emu [8], VIQTORYA [9].

2. Overlaps

Overlaps may be defined as a speech event in which two people speak simultaneously by uttering actual words or in some cases non-words, when one of the speakers, usually the interlocutor, interrupts or backchannels the current speaker. This phenomenon takes place at a certain point in time where it is anchored to the speech signal; but in order to be fully parsed and subsequently semantically interpreted, it needs to be referred semantically both to a following turn and to the local turn where it may produce conversational moves to repair what has been previously said by the current speaker.

One of the distinctive characteristics of naturalistic conversation (in contrast to monolog situations) is the presence of overlapping speech. Overlapping speech may be of several types, and affect the flow of discourse in various ways. An overlap may help to usurp the floor from another speaker (e.g., interruptions), or to encourage a speaker to continue (e.g., back channels), or simply end up just in an attempt at usurping the floor without success (Vain Interruption as defined in [10]). As a preliminary and tentative pragmatic definition we may define an overlap as being normally a physical event that happens in a single time unit in which two or more speakers want to communicate different and non-coincident communicative intentions. Exception made for rare cases in which the two or more speakers intended to say the same thing in the same time unit (see [11]).

Speaker overlaps, are directly observable in our data, since by definition overlaps occur at points of simultaneous speech on more than one of the (individually recorded) channels, besides their explicit indication in the ortho-phonetic transcription thus transliterated into the orthographic transcription. What we are interested in is finding out whether there is any correlation between the onset of overlaps and their possible characterization from the point of view of syntactic structure, which we have proposed to treat by introducing a node of discourse constituency called OVL (overlap), from where the two temporally aligned components of overlapping, the

overlapper and the overlappee stretch of speech/text, branch. The typologies proposed in the literature and those suggested by [10] will be verified in relation to their treatment at the level of syntactic constituency.

2.2 Overlaps and orthography: realignment and time irreversibility

In the original MapTask overlaps over two consecutive turns were simply marked off in blue colour: the words in blue overlapped. However, whenever there was more than one overlap in a single turn things became unclear, as shown in the following example taken from the materials made available on the web:

GIVER: **no do--** all right **okay, we'll we'll forg--.**

FOLLOWER: **I'm going** I'm going right... **I'm going right** towards the **yacht club?**

GIVER: **we'll forget** about the yacht club just now.

Here we are led to consider the Giver's "no do" to overlap with something previously pronounced by the Follower: on the contrary, this overlaps with the Follower's "I'm going"; then the Giver's "okay, we'll we'll forg--." Overlaps with the Follower's "I'm going right". Finally the Follower's "yacht club" overlaps with the Giver's "we'll forget". As can be gathered, there is no real motivation of separating turns which have strictly interconnected materials apart from the need to have a linear description. And as a matter of fact linearizing in the case of overlaps is twice wrong: phenomena which should belong to one and the same time unit are represented by the orthography as belonging to two separate time units. The colour is then used to rescue the temporal dimension.

The decision taken in the Italian MapTask was to follow the original transcription schema and conventions: in particular, overlaps are fully marked in the local speech aligned orthographic transcription, by introducing the index of the turn containing the overlapping material, which however is not visible and should be looked up in the following turn. In addition, two #s are introduced at the front of the turn index and at the end of the overlapped speech as shown in the following example:

Dialogue 2.

p1#94: no <sp> cioè sì c'ha<aa> <mh> <sp> una specie di tappo

p2#95: sì #<p1#96> c'ha un ta+ tappo <sp># , sì

p1#96: #<p2#95> di funghetto# <lp> c'ha prima una base un po' altina

Dialogue 2.1

p1_94: no, cioè sì c'ha, una specie di tappo.

p2_95: sì ov1_42 di funghetto < c'ha un ta_ tappo - >, sì.

Turn 95 contains an overlap which is introduced and erased from the following turn and indexed as shown in 4.1 version of the dialogue: the convention being that the ov1_42 index is followed by the overlapper's speech intruding in the overlappee's turn. The material being overlapped then follows the open '<' and the close of the overlap is marked by the closing '>'. In this way the orthography linearizes the bidimensional event of the overlap by keeping the linguistic material within the same turn as adjacent text rather than scattering it in different turns. The ownership of the

material by one of the speakers is guaranteed by its local respective position within the boundaries of the overlap: the *ov_N* starting symbol and the '>' at the end. It is important to notice that the two words are respectively pronounced by a woman and a man, the intruder utters with a rising tone: the implicit communicative intention is that of producing a better indication of the shape of the object currently under discussion and trying to get the other speaker to accept it.

There are two internal repairs caused by the overlap: the first one is "sì"/Yes as a reaction of Speaker 2 to a first definition of the shape "tappo"/cork, which is however taken only as being suggestive "una specie di"/a kind of, of a better yet to be defined final shape. And in the Speaker 2 turn, the repetition of "tappo" which is intentionally interrupted by recovering the turn role and suggesting the most appropriate shape, "di funghetto"/of a little mushroom.

3. Overlaps and syntax

As said above, overlaps challenge all criteria of linguistic representation which require the input sentence to be mono-dimensional, i.e. to contain the utterance of one single speaker. This fact is semantically essential in order to guarantee the linguistic representation to be interpretable. On the contrary, overlapped linguistic material, i.e. sentences which contain at the same time linguistic material coming from two or more participants in the dialogue are not only hard to parse: they might also constitute an obstacle to semantic interpretation [12].

As in most robust parsers, we use a sequence or cascade of transducers: however, in our approach, since we intend to recover sentence level structure, the process goes from partial parses to full sentence parses. Sentence and, below it, clause level are crucially responsible for the right assignment of arguments and adjuncts to a governing predicate head. This is clearly paramount in our scheme which aims at recovering TRPs by referring solely to syntax. Differently from what has been previously done in the English MapTask (see [13;14]), we erased all non-linguistic or paralinguistic elements from the original transcription, and added punctuation in accordance with pause interpretation.

Our proposal will then be articulated as follows:

- A. TRP coincides with sentence level or whenever main complements have been parsed and we are left with adjuncts;
- B. Non-TRP coincides with all the remaining cases, i.e. when the overlap starts between the specifier and the head of a constituent when still in preverbal position; or else whenever the overlap is positioned at the constituent boundary but the main governing predicate has not been parsed yet.

Case A. thus constitutes the semantically valid option corresponding to a projectable smooth TRP without overlap; on the contrary case B. constitutes the non-semantically viable option where the interlocutor does not have enough semantic content to project a TRP and simply wants to prevent the current speaker from continuing his talk.

Consider the previous example Dialogue 2, Speaker 2 utterance, whose syntactic structure is reported below,

Dialogue 2.b

da(turn(p2_95),cp(intj(si'), ovl(overlap(ov_42), spd(pd(di), sn(n(funghetto))), par(<), f(ibar(expl(c), vc(ha)), compc(sn(art(un), abbr(ta_), sn(n(tappo))))), par(par), overlap(>)), punt(virg), cp(intj(si'))), punto(.))

The realignment of all turns has given as a result a certain number of empty turns, i.e. all those turns which had been artificially built by simply containing overlapping material which had been already uttered by the current speaker before the previous turn was elapsed.

The need to represent linguistic information related to two speakers in the same syntactic structural representation, which is both semantically and pragmatically strongly intertwined has a lot of theoretical implications. In other words, to allow for feedback to take place between the Discourse and Sentence level of grammatical relations, we need discourse level phenomena to be adequately represented by sentence grammar. This is certainly the case we are tackling now: overlaps take place at a discourse level, however their import is deeply grafted into sentence grammar, by conditioning interpretation from taking place.

Coming now to overlaps, which as said above are on average occurring 1 every four turns, have been organized as follows:

- Overlaps occurring between specifier and head;
- Overlaps occurring in a parallel and unintentional simultaneous way;
- Overlaps which are semantically empty and are computable as backchannels;
- Overlaps at a higher constituency level, i.e. at sentence level or after main predicate and main complement has been computed.

In Table 1. below we show both absolute and percent values of all overlaps distributed in the three locations where the Italian MapTask has been recorded. We marked most deviant values in bold and italics. As can be easily gathered, Rome/Naples are the seat where in absolute terms most overlaps have occurred; it is also where, in absolute and relative terms, most semantically empty overlaps occurred. Bari, is the seat where in absolute terms the least overlaps have occurred: it is also the place where in relative terms the least semantically empty overlaps occurred; in addition to that, it is the place where the most specifier/head occurred and the most TRP relatable overlaps occurred. Eventually, Pisa – with four dialogues – is the seat where most parallel overlaps occurred.

Tab 1. Total overlaps data in AVIP/API Dialogues classified

Sites / Overlaps	Overlaps Totals	Specifier/constituent Head Overlap	Parallel Overlap	Semantically Empty Overlaps	Higher Constituency Level Overlaps
BARI	142 9.44%	53 31.2%	11 0.78%	25 17.6%	53 31.2%
NAPOLI	909	221 24.31%	97 10.67%	333 36.33%	258 26.38%
PISA	264	64 26.12%	58 21.96%	70 26.51%	72 27.27%
Rome	189 12.56%	42 22.22%	24 12.69%	77 40.74%	46 24.33%
Totals	1504	380	190	505	429

Means		25.26%	12.63%	33.57%	28.52%
--------------	--	--------	--------	---------------	--------

4. Overlaps, TRP and Dropping

Another important indicator of the actual import of an overlap is the relation intervening between an overlap and the completion of the turn by the current speaker. An overlap that also marks a dropping of turn by the current speaker who yields his turn to the overlapper can be computed differently according to whether the overlap takes place at the end of the turn or not. Turn-taking rules would relate TRP with end of turn and that would be the regular place for an overlap to occur. This is due to the fact that in normal conversational interaction speakers would be in the conditions to forecast when the current turn is ending and would produce an overlap past the TRP to speed up the attainment of the communication task. On the contrary, dialogues by speakers who use overlaps turn internally would contain a lot of cases of Continuing Conversation in presence of Overlaps: these we call Non Dropping Overlaps (hence NDOs). If we divide up all Overlaps into two categories: Competitive vs. Collaborative we end up with the following general subdivision,

- Competitive Overlaps – Parallel + Spec/Head
- Collaborative Overlaps – Semantically Empty – Higher than Constituent

It is a fact that NDOs may only occur in a competitive situation: either within what we defined Parallel Overlap or within a Spec/Head Overlap. We may interpret the occurrence of NDO as an indication of a collaborative attitude between the interactants: in presence of an overlap, people continue speaking. As we saw in Tab.1 and Tab.2 above Turns containing more than one Overlap are 98 overall. To these cases of NDOs we add all cases in which speakers alternate short and long pauses with overlaps during a long turn without the overlapper actually usurping the floor.

We computed NDOs for all dialogues and the overall picture we get is shown in Tab.6 below. As can be seen, Bari has the least number of NDOs, Napoli on the contrary has the highest number thus confirming our previous conclusion. Naples conversational style has as a specialty the exploitation of overlaps as a means to make dialogues more communicative, most redundant and least efficient. Nonetheless, this is accepted as a rule by conversants of the same regional variety and is regarded as an effective tool.

Tab 2. DisaggregatedOverlaps data in all Dialogues in absolute values

Sites / Overlaps	Overlap Totals	Competitive Overlaps	NDOs	% NDOs	% NDOs wrt. Competitive
Bari	142	64	14	9.89	21.87
Napoli	909	318	275	30.25	86.47
Pisa	264	122	32	12.12	26.22
Rome	189	66	19	10.05	28.78
Totals	1504	576	340	22.6	59.02

If utterances containing overlaps may be represented syntactically, then it should also follow that they should be interpretable semantically. In fact, this may be true for all overlaps containing linguistic material which is related to its context, right or left. However, when the linguistic material does not relate locally to any portion of the utterance it is hard to define a strategy. In the former case, the interpretation follows from a treatment of overlapped material as belonging to the current clause – as a fragment - or to a previous utterance – as an ellipsed fragment. The semantics of this utterance can be represented in a flat logical form where the DA (Dialogue Act) contains an OVL to which the governing predicate “HAVE” is applied twice, as follows,

```
Dialogue 2.b
DA(turn(p2_95),
  OVL(prop(si(y1),
    tappo(x1),
      AVERE(y1,x1)),
    prop(funghetto(x2),
      sì(y2),
        AVERE(y2,x2))))))
```

OVL is thus treated as a modality operator which has scope over two propositions.

6. Conclusions

We have presented work carried out to annotate the AVIP/IPAR spontaneous speech Treebank of Italian. We highlighted its peculiar features in taking care of overlaps, which have been time aligned and then parsed. In order to produce adequate syntactic constituency for such structures we decided to introduce a specialized node, OVL which includes the linguistic material produced by both the overlapped and the overlappee. We have given theoretical and practical pragmatic reasons for that choice. Finally we presented quantitative data from the Treebank and compared them to VIT, our Treebank of written Italian. Results show expected differences in utterance length and typologies which qualify spoken vs. written use of language.

References

1. Morgan, N., D. Baron, J. Edwards, D. Ellis, D. Gelbart, A. Janin, T. Pfau, E. Shriberg, and A. Stolcke. 2001. The ICSI Meeting Project. In *Proceedings of the Human Language Technology Conference*, San Diego, pp.10-18.
2. Delmonte R., Bristol A., Chiran L., Bacalu C., Tonelli S., (2004), “Parsing the Oral Corpus AVIP/API”, in *Atti del Convegno Internazionale “Il Parlato Italiano”*, Napoli, Università di Napoli, 20-40.
3. Delmonte, R. (2004), Strutture Sintattiche dall'analisi Computazionale di Corpora di Italiano, in Anna Cardinaletti (a cura di), "Intorno all'Italiano Contemporaneo" Franco Angeli, Milano, pp. 187-220.

4. Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson, 1974. A simplest systematics for the organization of turn-taking in conversation. *Language* 50(4): 696-735.
5. Ford, C.E. and Thompson, S.A. 1996. Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E.A. Schegloff & S.A. Thompson (eds) *Interaction and grammar*, Cambridge: Cambridge University Press, pp. 134-184.
6. Delmonte R. 2003. Parsing Spontaneous Speech, in Proc. EUROSPEECH2003, Pallotta Vincenzo, Popescu-Belis Andrei, Rajman Martin "Robust Methods in Processing of Natural Language Dialogues", Genève, pp, 16-23.
7. Cotton S. and S. Bird. 2002. An Integrated Framework for Treebanks and Multilayer Annotations. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pages 1670–1677. ELRA.
8. Cassidy S. and J. Harrington. 2001. Multi-level annotation in the Emu speech database management system. *Speech Communication*, 33(12):61–77.
9. Steiner I. and L. Kallmeyer. 2002. VIQTORYA – A Visual Query Tool for Syntactically Annotated Corpora. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, pages 1704–1711. ELRA.
10. Bazzanella, Carla 1994. "LE INTERRUZIONI", in *Le facce del parlare*, Cap 8, Firenze/Roma: La Nuova Italia, pp.176.
11. Wheatley, B., G. Doddington, C. Hemphill, J. Godfrey, E.C. Holliman, J. McDaniel, and D. Fisher, "Robust Automatic Time Alignment of Orthographic Transcriptions with Unconstrained Speech", *Proc. ICASSP-92*, Vol. I, 533-53.
12. Shriberg, E.; A. Stolcke, and D. Baron. Observations on overlap: Findings and implications for automatic processing of multi-party conversation. In P. Dalsgaard, B. Lindberg, H. Benner, and Z. Tan, editors, *Proc. EUROSPEECH*, vol. 2, pp. 1359–1362, Aalborg, Denmark, 2001.
13. Hindle D. 1993. "Deterministic parsing of syntactic nonfluencies", In *Proc. ACL*, pages 123-128.
14. McKelvie, D., 1998. "The Syntax of Disfluency in Spontaneous Spoken Language", *HCRC Research Paper HCRC/RP-95*, Edinburgh.