A corpus-based study on syntactic and phonetic prosodic phrasing boundaries in spontaneous Italian speech

Renata Savy, Miriam Voghera

Department of Linguistic and Literary Studies, University of Salerno, Italy

{rsavy; voghera}@unisa.it

Abstract

This work focuses on the relationship between prosodic and syntactic domains in order to investigate whether there is a preferred syntactic domain of tonal units and whether there is a preferred prosodic domain of syntactic constituents. Two independent analyses of phonetic prosodic boundaries of syntactic constituents and of syntactic structure of prosodic constituents in Italian spontaneous dialogues were carried out. On the one hand, our results confirmed data from previous studies on non-isomorphism of syntactic and prosodic constituents. On the other hand, new data are presented, which show that: 1) the coextension of prosodic and syntactic phrasing is favoured by specific syntactic structures, mainly phrases or minimal sentence structures; 2) informational and pragmatic factors as well as turn-taking frequency strongly constrain prosodic (and syntactic) phrasing; 3) the same acoustic-phonetic cues are involved in prosodic phrasing of different syntactic constituents.

Index Terms: speech, Italian, prosody, syntax, phrasing.

1. Introduction

The relationship between prosodic and syntactic phrasing is a well established topic within the scientific debate, and very recently new data have been collected on Romance languages [1, 2, 3]. Most results pertain the analysis of laboratory or read speech, in which phonetic as well as other linguistic and register variables are put under control [2, 3]. Comparatively few are studies which take into account a wider-ranging set of data, including different speech registers and dialogic material, and even fewer are on Italian. However, the analysis of spontaneous speech is valuable, on the one hand, to study the complex interplay between phonetic and phonological prosodic features and, on the other hand, to assess the role of textual and contextual variations. We present here the first step of a long-term program, which aims at the description of the prosody-syntax interface in Italian spontaneous dialogues, starting from the prosodic phonetic segmentation and its relation with both phonological and syntactical phrasing. We focus here on the first nucleus of data concerning the relationship between phonetically segmented prosodic units and their syntactical constituency.

According to Prosodic Phonology the prosodic constituents are defined both on the basis of boundary marks and as dominion of different phonological phenomena. The output is a prosodic hierarchy of different layers. There is a general agreement on the non-isomorphism between prosodic and syntactic constituents, as claimed in several classical and recent works [4, 5, 6, 7, 8]. However, some correspondences between the hierarchy of prosodic constituents above the word and syntactic hierarchy is registered. Major Phonological Phrases (MaP) easily correspond to syntactic phrases and Intonational Phrases(I) to clauses:

ex. 1: $[(\text{Mangerò})_{\text{MaP}}\ (\text{panini col salame})_{\text{MaP}}]_{\text{IP}}$
(I) will eat salami sandwiches

While this correspondences are frequently found in laboratory and read speech, no comparable amount of data on spontaneous speech is available. In fact, the analysis of prosodic and syntactic phrasing in spontaneous dialogues must face several theoretical and empirical problems.

Firstly in spontaneous speech it is not always clear how constituents belonging to different layers must be detected, since the relationship between the phonetic correlates of prosodic units and their phonological representation is not so well defined and the same phonetic markers can be assigned in turn to Phonological or Intonational Phrase boundaries.

Secondly, further difficulties raise in dialogic texts where the detection of syntactic unit boundaries within the speech stream can be very problematic, since most of the utterances do not present the canonical sentence form and/or many turns are constituted by nonsententials [9], i.e. verbless non-elliptical utterances. In such cases both phonetic and syntactic phrasing can be difficult.

Finally, prosodic phrasing also depends on informational as well as pragmatic factors, so that according to some scholars prosody is basically information dependent [10].

In this paper we offer some empirical data on both phonetic and syntactic segmentation. The result is a comparison between prosodic and syntactic phrasing in Italian speech based on an acoustic analysis of syntactic and tone unit boundaries. This allows to investigate whether there is a preferred syntactic domain of phonetic tonal units and whether there is a preferred prosodic domain of syntactic constituents. Our findings can represent a useful basis for detecting which correspondences are phonologically relevant in order to support prosodic phonology hierarchy.

2. Method

2.1. Corpus

We analysed three task-oriented dialogues from the Clips corpus [11] (www.clips.unina.it), elicited by the "Spot the difference" game technique and produced by six speakers of three different regional varieties of Italian. Each dialogue has a duration of about 15 minutes and consists of a minimum of 192 and a maximum of 361 turns. Due to the specific type of interaction (basically a question/answer dialogue), this kind of texts show a high frequency of turn-taking and turns are relatively short (mean length=8 words), although they can have a very variable length (from 1 to 50 words). Moreover many words are monosyllabic.

The entire corpus has been analyzed at the prosodic and syntactic level separately (cf. §2.2) and consists of 842 turns, 3577 TUs and 1409 clauses.

2.2. Annotation criteria

2.2.1. Syntactic Analysis Criteria

The syntactic analysis has been carried out using AN.ANA.S. a constituent-based system for syntactic annotation linearly aligned to the verbal sequence to match the syntactic analysis with the level of prosodic annotation [12, 13]. AN.ANA.S. uses XML to represent the annotations and allows the organization of linguistic units within a hierarchical structure, which comprehends different layers; in this article we have considered the phrase and the clause levels only. The annotation distinguishes among different types of phrases: NP, VP, PP and PNP, the predicative phrase depending on a copular verb. Three types of clauses are recognized: main, dependent and verbless. We tagged as verbless clauses all autonomous utterances, which cannot be considered elliptical clauses, such as clauses with VP ellipsis and sluicing-stranding structures [9, 14].

2.2.2. Prosodic Analysis Criteria

The prosodic annotation consists of three different levels: 1) phrasing; 2) phonetic tonal tagging; 3) stress/accent tagging. All annotations are in TIMIT format. In this paper we focus on the phrasing level only, in which the time-aligned sequence of Tone Units (TU) was marked.

TUs were analyzed and segmented on the basis of some acoustic criteria which are considered to be evidences of prosodic cohesion; they may refer to local or global prosodic events and can co-occur in a prosodic group [15, 16, 17, 18]:

- presence of a (potential) final pause
- coherence in global trends of f0 and energy (e. g. declination of both f0 and energy)
- parametrical reset at the beginning of a new TU
- pre-pause lengthening

Since our aim is to detect which syntactic constituent matches the phonetic segmentation more frequently, TU are isolated as phonetic units, regardless of their phonological constituency, and no *a priori* correspondence between phonetic TU and IP or MaP constituents was made, neither phonological description of the boundary tones is provided.

2.3. Coding and querying

The annotated dialogues were codified according to the AG standard [19], and structured in a database (SpIT–MDb [20]), which allows automatic querying on the relationships between different annotation levels. We then asked separately for the correspondence

- of [clause] or [XP] to [TU] labelled strings
- of [TU] to [clause] or [XP] labelled strings

Moreover, TU and clause left and/or right boundaries misalignments were retrieved (see §3.1).

3. Results

Data from the syntactic and prosodic point of view are presented separately, while the relationship between the two sets of data will be discussed in the next section.

3.1 The syntactic point of view

The data in Table 1.a suggest that the TU is not the preferred prosodic domain of clauses, since in only 41% of cases there is a one-to-one correspondence and clauses can be smaller than a TU or span over more than a TU (Table 1.b).

Table 1. a) *Prosodic extension of Clauses;* b) *Number of TUs per clause (for C > TU)*

1.a

Boundaries	**% of clauses**
Clause = TU	41%
Clause > TU	41%
Clause < TU	10.4%
Misalignment	7.5%
Total	100%

1.b

TU/clause	**% of clauses**
2	15.3%
3	5.1%
4	2.1%
>4	1.2%
>1, <2	17.6%
Total	41.3%

However, data in Table 2 show that the cases of misalignment are only 7,5%, while in 75% of cases clauses and TU left boundaries coincide, and in 65% of cases both the left and right boundaries coincide, whatever the number of TUs (ex. 3-6). The higher coincidence of left boundaries confirms that the Align-Left constraint is at work [5], regardless of the clause length and constituency.

Table 2. *Clause and TU boundaries*

Syntactic/Prosodic constituency	**Left boundary coincidence**	**Right boundary coincidence**	**%**
Clause = TU	+	+	41%
Clause > TU	+	+	24%
Clause < TU	+	-	10%
Clause > TU	+/-	-/+	18%
Misalignment	-	-	7.5%

ex. 3: /[*le ruote sono rotonde*]$_{CLA}$/$_{TU}$
(the wheels are round)
ex. 4: /[*scendendo*/$_{TU}$ /*verso il gatto*]$_{CLA}$/$_{TU}$
(going down/ toward the cat)
ex. 5: /[*hai detto*]$_{CLA}$ [*hai il cuore tu*]$_{CLA}$/$_{TU}$
(you said you have the heart)
ex. 6: /[*i puntini in basso a destra*/$_{TU}$ /*ce n'ho sette*]$_{CLA}$ [*di cui*/$_{TU}$ /*due triangoli*/$_{TU}$ /*e uno puntato al centro*/$_{TU}$]$_{CLA}$
(the little dots on bottom right/ I have seven of them of which two triangles and one drawn at the center)

Among the factors that can favour the one-to-one correspondence between clauses and TUs (Cs=TUs), we considered both clausal length and constituency. Surprisingly in our data length does not play a relevant role, since only 43% of Cs=TUs consists of just one phrase. Two syntactic factors are more relevant: dependency and clause internal structure. 97% of Cs=TUs is a main clause (both verb and verbless), while only 3% of Cs=TUs is a dependent clause, mainly relative clauses or *perché* (*because*) clauses. As far as the internal structure is concerned, Cs=TUs are mostly verb clauses (53%). This is particularly striking if we consider that in spontaneous dialogues short verbless utterances, such as typical *yes/no* answers and discourse markers, are very frequent. In Table 3 we report the most frequent syntactic structures in Cs=TUs.

Table 3. *Most frequent syntactic structures in Cs=TUs*

Frequency Rank	**Syntactic structure of *Cs=TUs***
1	Copular *to be* (ex. 7)
2	Existential *to be* (ex. 8)
3	Declarative (S)VO (ex. 9)
4	Imperative VO (ex. 10)
5	Declarative SVO (ex. 11)

ex. 7: [*le ruote sono rotonde*] (the wheels are round)
ex. 8: [*e poi sotto al sasso ci sono tre lineette*]
(and then under the stone there are three little lines)
ex. 9: [*c'ha un grande quadrato*] (it has a big square)
ex. 10: [*fai questo percorso*] (follow (lit. do) this route)
ex. 11: [*la chiusura c'ha una specie di manico*]
(the fastener has a sort of handle)

However, such structures have different frequency: copular and existential *to be* clauses represent 60% of total, while SVO declarative clauses represent just 10% of total.

3.2 The prosodic point of view

If we look at the data from the prosodic point of view (Table 4), half of phonetic TUs is smaller than a clause, while 39% of TUs has the extension of a clause:

Table 4. *Syntactic constituency of phonetic TUs.*

Constituency	**% of TUs**
TU = Multi-Phrase Clause	27.5%
TU= Single Phrase Clause	11.7%
TU < Clause	49.4%
TU > Clause	6.9%
Misalignment	4.4%
Total	100%

TUs corresponding to single-phrase clauses consist in both verb (VPs; ex. 12-13) and verbless clauses (NPs and PPs; ex. 14-15):

ex. 12: */aspetta* $_{VP=CL}$/$_{TU}$ (wait);
ex. 13: */ho capito* $_{VP=CL}$/$_{TU}$ (I have understood);
ex. 14: */sulla destra* $_{PP=CL}$/$_{TU}$ (on the right);
ex. 15: */le dita del bambino* $_{NP=CL}$/$_{TU}$ (the fingers of the child).

In 50% of cases TUs smaller than a clause correspond to one or more syntactic phrases (Table 5):

Table 5. *Syntactic constituency for TU<clause*

Constituency	**% of TUs<Cl**	**% of total TU**
1 XP	28.6%	14.1%
>1 XP	21.3%	10,6%
Spoken elements	50.0%	24,7%
tot	100%	49,4%

The phonetic TUs corresponding to single phrases are costituted mostly by NPs and PPs, with or without modifiers:

ex. 16: */la spada*$_{NP}$/$_{TU}$ (the sword);
ex. 17: */due triangoli*$_{NP}$/$_{TU}$ (two triangles);
ex. 18: */la parte finale*$_{NP}$/$_{TU}$ (the final part);
ex. 19: */all'estrema destra*$_{PP}$/$_{TU}$ (at the estreme right).

Phonetic TUs wider than one phrase consist of branching NPs and PPs (ex. 20), coordinate phrases (ex. 24), but mostly in phrases plus other material, such as adverbs or disfluencies (ex. 21, 22, 23).

ex. 20: */il polsino della manica della signora*$_{NP}$/$_{TU}$
(the cuff of the sleeve of the lady)
ex. 21: */lo specchietto*$_{NP}$ *invece*$_{Adv}$/ (the mirror instead)
ex. 22: */alla parte orizzontale*$_{PP}$ *proprio*$_{Adv}$/$_{TU}$
(on the orizzontal part just)
ex. 23: */la seconda*$_{NP}$ *un po' meno*$_{Adv}$/$_{TU}$
(the second one a little less)
ex. 24: */quella di destra*$_{NP}$ *e quella di sinistra*$_{NP}$/$_{TU}$
(that on the right and that on the left)

Summing up, if we consider the data in Table 4 and 5, in more than 36% of cases TUs corresponds either to single-phrase clauses or to single or multiple phrase structures.

Phrase length does not play a crucial role in phrasing, since the TUs=Phs have an extension from 1 to 15 syllables:

ex. 25: */uno due tre quattro cinque sei sette puntini*/$_{TU}$
(one two three four five six seven little dots)

Finally, 50% of the total of TUs<clause (25% out of total TUs) coincides with typical spoken elements, such as discourse markers, interjections, retreat and repair sequences, disfluencies:

ex. 26: ***/sono/***$_{TU}$ */constano di due dita/*$_{TU}$
(they are / consist of two fingers)
ex. 27: ***/cioè/***$_{TU}$ *dal naso alla bocca c'è una linea/*$_{TU}$
(/that is/ from the nose to the mouth there is a line)
ex. 28: */eheh bravo/* $_{TU}$ ***/perfettamente/***$_{TU}$
(eheh good/ perfectly)

4. Discussion

The independent analysis of phonetic prosodic boundaries of syntactic constituents and of syntactic structure of prosodic units allows some interesting considerations. In fact, the two perspectives offer a partially different state of affairs.

While clauses do not have a preferred prosodic domain, TU phonetic boundaries tend to coincide with phrase boundaries [21]. This difference could be attributed to different length of the two syntactic constituents: normally clauses are longer than phrases. Yet, length does not seem to play a decisive role in clause or in phrase prosodic phrasing, since both Cs=TUs and Phs=TUs are not necessarily the shortest. Only 43% of Cs=TUs consists of just one phrase and only 28% of just one word, and Phs=TUs can go from 1 to 15 syllables.

Other factors appear more relevant, such as the syntactic structure for clauses and the informational and dialogic structure for phrases.

The two most frequent structures in Cs=TUs are predicative and existential *to be* clauses. Both structures can be considered minimal sentential structures, in which the semantically null *to be* has the only function to express tense and aspect information. In fact, according to [22], existential clauses can be considered as inverse copular clauses. This 'minimal sentencehood' seems to favour a single phonetic phrase. Besides, the totality of Cs=TUs is constituted by main clauses, while rarely a dependent clause corresponds to TU.

As far as TUs=Phs are concerned, an important role is played by information structure: topic elements tend to be prosodically isolated [10, 23] (ex. 29), although topicality is not necessarily a condition of isolation as shown in previous studies on spontaneous dialogues (cf. [24]). Moreover prosodic segmentation frequently signals comment NP, regardless of their syntactic status. More remarkably comment NP (TUs=Cs) are frequently signalled by prosodic segmentation (ex. 30-31).

ex. 29: */la chiusura/*$_{TOPIC//TU}$ */c'ha una specie di manico?/*$_{TU}$
(the fastener /it has a kind of handle?)
ex. 30: */Le dita?/* $_{COMMENT/TU}$
(the fingers?)
ex. 31: */poi/ /le dita del bambino/* $_{COMMENT/TU}$
(then/ the fingers of the child)

Other pragmatic and textual factors contribute to prosodic phrasing in spontaneous dialogues: the tight turn exchange constitutes a constraint for prosodic (and syntactic) segmentation; questions (ex. 30), answers and acknowledgements (ex. 32) alternate, making the text a fragmented structure; re-planning elements (ex. 33, 26), disfluencies, repetitions, discourse markers (ex. 34, 27), phatic expression (ex. 35) further amplify text discontinuity.

ex. 32: */perfetto/* $_{TU}$ [verbless clause]
(perfect)
ex. 33: ***/il cavallo là/***$_{TU}$ */il cavaliere/* $_{TU}$ */ce l'ha la spada*?/ $_{TU}$
(The horse there/ the horseman has he the sword?)
ex. 34: ***/allora/***$_{TU}$ */vogliamo partire dalle finestre della casa*?/ $_{TU}$
(so/ want (we) start from the windows of the house?)
ex. 35: ***/vabbè dunque/***$_{TU}$ */fammi pensare/* $_{TU}$
(well then/ let me think)

Phonetic prosodic phrasing marks the autonomy of these structures, but on the basis of our data no claims on their phonological constituency are allowed: from the acoustic point of view, the same phonetic cues are involved in the phrasing of possible different prosodic constituents. Moreover, it must be kept in mind that a great amount of TUs corresponds neither to a clause nor to a phrase, but to re-planning sequences, adverbial and discourse markers, whose syntactic status is debatable: this makes the syntax-prosody interface analysis more tentative.

In conclusion, it is not easy to identify boundary correspondences between (phonological-)prosodic and syntactic constituents on the basis of phonetic correlates. Our data on spontaneous spoken dialogues confirm that syntactic constituents and prosodic units are non-isomorphic, whatever the extension and the hierarchical level are considered. Moreover, not all the prosodic segmentations delimit syntactic constituents, but can isolate also typical spoken elements. Nevertheless, this kind of analysis can constitute a basis for further and deeper investigations on the relationship between acoustic/phonetic structures and phonological representation as well as on syntax-prosody interface.

5. Conclusions

Spontaneous spoken dialogues present similar regular features, cross-linguistically shared, depending on the on-line productive/receptive processes, which entail a low degree of pre-planning by the speaker and a low degree of selective attention by the receiver. This produces texts, which are basically discontinuous, i.e. made of short chunks of speech not always hierarchically structured, but rather adjoined to one another. Both syntax and prosody contribute to the phrasing of the speech stream, but their relationship is not always straightforward. Prosody plays a crucial role in phrasing speech chunks that are relevant for both speaker and hearer, but not necessarily syntactically coherent: textual, pragmatic and informative constraints are simultaneously at work. Due to this complex set of factors, prosodic phrasing does not always maps onto syntactic constituents, as the high number of TUs=typical spoken elements shows. The syntactic analysis of prosodic phrasing reveals that the constituent which coincide more frequently with a TU is the phrase, regardless to its length or internal structure. On the contrary, clauses do not map into a preferred prosodic domain, unless they are minimal sentential structures, which tend to coincide with TU boundaries [21].

These results suggest that TU boundaries are not dependent on syntactic constraints, but the prosodic and syntactic coextension can be favoured by specific syntactic structures.

Further steps are however necessary to complete the analysis of our material. Stress/Accent degrees and tonal patterns are essential elements in shaping and delimiting prosodic phrase boundaries. Moreover, the analysis of f0 level and different kinds of boundaries is necessary to investigate the degree of coherence *across* TUs, which can better show if and how TUs are organized in higher prosodic (linguistic) constituents. Consequently the study of the interplay of different prosodic levels and syntactic phrasing will be our next goal.

6. Acknowledgements

The present research was realized within the project *Parlare italiano* (www.parlaritaliano.it), funded by the Italian Ministry of Education, University and Research. We thank Rosa Giordano and Mariapaola D'Imperio for useful comments on a previous version.

7. References

[1] Nespor, M., Vogel, I., Prosodic Phonology, Dordrecht: Foris Publications, 1986.

[2] D'Imperio, M., Elordieta, G., Frota, S., Prieto, P., Vigario, M., "Intonational phrasing in Romance: The role of syntactic and prosodic structure", in Frota, S., Vigario, M., Freita, M. J. (eds.), Prosodies, Berlin, Mouton de Gruyter, 59-97, 2005.

[3] Frota, S. D'Imperio, M. Elordieta, G. Prieto, P. Vigário, M. "The phonetics and phonology of intonational phrasing in Romance". Segmental and Prosodic Issues in Romance Phonology, ed. by P. Prieto, J. Mascaró, and M.-J. Solé. Amsterdam/Philadelphia, John Benjamins:, 131-153, 2007.

[4] Selkirk E., Phonology and Syntax: the Relation between Sound and Structure, Cambridge Mass.: MIT press, 1984.

[5] Selkirk E., "The Syntax-phonology Interface", in Smelser, N.J., Baltes, Paul B. (eds.), Intern. Encycl. of the Social. and behavioral Sciences, Oxford: Pergamon, 15407-15412, 2001.

[6] Selkirk, E.. "Sentence phonology", in Frawley, W., Bright, W. (eds.), The Oxford Intern. Encycl. of Linguistics, 2nd edition, New York and Oxford: Oxford University Press, 2003.

[7] Truckenbrodt, H., The syntax-phonology interface, in Paul de Lacy (ed.), The Cambridge handbook of phonology, Cambridge: Cambridge University Press, 435-356, 2007.

[8] Frota, S., Prosody and focus in European Portuguese: Phonological phrasing and intonation, New York, Garland Publishing Inc, 2000.

[9] Progovac, L., Paesani, K., Casielles, E., Barton, E. (eds.) The Syntax of Nonsententials, Amsterdam/Philadelphia: Benjamins, 2006.

[10] Cresti, E., Moneglia, M. (eds.), C-Oral Rom. Integrated Reference Corpora for Spoken Romance Languages, Amsterdam/Philadelphia, Benjamins, 2005.

[11] Cutugno F., Savy R., "CLIPS: diatopic, diamesic and diaphasic variations of spoken Italian", paper accepted at the 5th Corpus Linguistics Conference, Liverpool, 20-23 July 2009.

[12] Voghera, M., Basile, G., Cutugno, F., Fiorentino, G., "Sintassi in AN.ANA.S.", in F. Albano Leoni, R. Giordano, Italiano parlato. Analisi di un dialogo, Napoli, Liguori, 189-211, 2005.

[13] Voghera, M., Cutugno, F., "AN.ANA.S.: aligning text to temporal syntagmatic progression in Treebanks", Proceedings of the 5th Corpus Linguistics Conference, Liverpool, 20-23 July 2009, in press.

[14] Giordano, R., Voghera, M., "Frasi senza verbo: il contributo della prosodia", in Ferrari, A. (ed.), *Sintassi storica e sincronica dell'italiano*, Atti del Conv. Intern. SILFI, Basilea 2008, in press.

[15] Caputo M.R., "Aspetti prosodici del processo di segmentazione nel parlato spontaneo", in *Atti del XX Convegno Nazionale dell'AIA*, Roma, Esagrafica, 361-366, 1992.

[16] Ladd, D., Intonational phonology, Cambridge: Cambridge University Press, 1996.

[17] Gussenhoven, C., The Phonology of Tone and Intonation., Cambridge: Cambridge University Press, 2004.

[18] Hirst, D.; Di Cristo, A., "A survey of intonation systems". In Hirst D., Di Cristo A. (eds.), Intonation Systems. A survey of twenty languages, Cambridge: Cambridge University Press, 1-44, 1998.

[19] Bird S., Liberman M. "A formal framework for linguistic annotation," Speech Communication, 33:23-60, 2001.

[20] Savy R., Cutugno, F., Crocco, C., "Multilevel corpus analysis: generating and querying an AGset of spoken Italian (SpIt-MDb)". Proc. of 5th Intern. LREC Conf., 1654-1659, 2006.

[21] Voghera M., "Gruppi tonali e strutture sintattiche nell'italiano parlato spontaneo", The Italianist, 10: 175-203, 1990.

[22] Moro, A., Existential Sentences and Expletive There, in Blackwell Companion to Syntax, , Padova, UNIPRESS, 1997.

[23] Hedberg, N., Sosa, M., "The Prosody of Topic and Focus in Spontaneous English Dialogue", in Büring D., Gordon M., Chungmin L. (eds.), Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation, Dordrecht, Kluwer Academic Publishers, 101-120, 2006.

[24] Crocco C., Savy R., "Topic in dialogue: prosodic and syntactic features". In Proc. of Interspeech '07, Antwerp, CD-Rom, 2007.