The Realization of Semantic Focus and Language Modeling

RODOLFO DELMONTE

Istituto di Linguistica – Università di Venezia

Ca' Garzoni-Moro – S.Marco, 3417 – VENEZIA(I)

ABSTRACT

Italian is a language in which discourse level informational strategies are easily detectable at sentence level. When arguments of a certain predicate do not constitute new information they are adjoined as clitic to the front of the verb; subject arguments constituting the theme of a discourse or text are left unexpressed. All relevant information on the contrary is highlighted by means of a variety of structural means: these are usually accompanied by phonetic signals mostly at the level of intonational contours. Semantic focus can be characterized by phonological structure, syntactic structure and pragmatic or full semantic representation. Only emphatic and contrastive focussing requires pragmatic or full semantic representation: this is not generated by available grammatical components of rule systems for speech synthesis, currently presented in the literature. The two remaining levels of representation, the phonological and the syntactic ones, enable a system of synthesis by rule to realize focus structure in most cases. Relevant semantic information is passed on to the syntactic component from the lexicon, which must be highly articulated. The remaining components activated in a system for synthesis by rule are the morphological and the phonological ones.

Phonetically speaking, the focussed constituent can be characterized by a peak with Low or High tone, aligned with word-stress, accompanied by a preceding H/L tone and sometimes followed by a L tone in coincidence with an Intonational Group boundary. Intonational Groups (IGs) constitute the higher phonological structure and are defined on a syntactic-semantic level, as the root sentence including the higher S node and its complements and modifiers. Moreover, we found out that to obtain a satisfactory definition of focus the highest-lowest peak in F_0 value is not sufficient as an acoustic correlate. Focus is defined as a relation over two adjacent tonal assignments, in terms of the rate/s of change of the F_0 curve.

INTRODUCTION

In a previous paper[1]we distinguished between Phonological Focus (FF) which gives rise to unmarked Focus Assignment Rules (FAR), and Logical Focus (LF) which gives rise to marked FAR. The former constitutes a case of default sentence level rule which associates a certain basic pitch contour with each Intonational Group(IG). Basic intonational contours of a certain language are usually defined generalizing over a set of illocutionary types (or tunes as defined in[2]) which are language-specific. In Italian there are at least the following: declaratives, questions, exclamatives and parentheticals. IGs constitute the higher phonological structure and are defined on a syntactic-semantic level, as the root sentence including the higher S node and its complements and modifiers.

Logical Focus (LF) is conceived as the pitch induced by syntactically governed discontinuities of constituents which can – and usually are – affected by discourse level rules, as to their interpretation. These structures are however detectable at sentence level and give rise to a syntactic representation in which grammatical functions are assigned to constituents which do not occupy their canonical position in superficial or constituent structure. FF and LF generate focus structures which define the boundary of a sense unit at a discourse grammar level: with FF focus structure includes the arguments of the predicate as they are normally associated by lexical frames, where syntactic or functional subcategorization, selectional restrictions and other feature information is listed for each lexical entry. In the case of LF this is also taken into account, plus the marked structures of Italian in terms of syntactic discontinuities. No pragmatic or extragrammatical knowledge is required, however, since no emphatic or contrastive structures are generated by the rules.

We take for granted that the system will generate an adequate structural description of marked structures(but see[3]). In order to investigate its relations with an acoustic-phonetic model of focus structure wee built a test set made up of sentences inlcuding the following structural types:
1. Neutral declarative followed by a subordinate hypothetical clause;
2. Topicalized version of 1.

3. Clitic left dislocation version of 1.
4. Clitic right dislocation version of 1.
5. Sentence with an Extraposed Subject NP;
6. Sentence 1 with Postposed Subject;
7. Sentence 1 with Inverted Subject;
8. Cleft construction;
9. Wh- question;
10. Yes-no question.

Sentences have been read aloud by an expert phonetician who repeated them until he judged to have performed the best rendering. F. and short-term power (both on a log scale) were computed each 10 ms at the CSC of the University of Padua.
Sentences are listed below with underneath their phonological marking:

1. Gli industriali devono pagare i decimali
   HL*   H*   H*LL%
se vogliono che le trattative continuino.
                              L%

2. I decimali gli industriali devono pagare
   H*L   L%
se vogliono che le trattative continuino.
                              L%

3. I decimali gli industriali devono pagarli
   H L*   H L*   H*L L%
se vogliono continuare le trattative.
                         L%

4. Gli industriali devono pagarli i decimali
   H L*   H* L   L%
se vogliono che le trattative continuino
                              L%

5. Questo accordo non possono accettarlo
   H L*   H*   H   H*LL%
i sindacati
   L%

6. Devono pagare i decimali gli industriali
   H   H   H   H*L   L%
se vogliono continuare le trattative.
                         L%

7. Devono pagare gli industriali i decimali
   H   H   H   H*L   L%
se vogliono la continuazione delle
trattative.
   L%

8. Sono i decimali che gli industriali non
   H*L L%
vogliono pagare.
   L%

9. Chi hanno detto che hanno intenzione di
   H   H   H*L L%   aiutare gli
industriali?
                L%

10. Hanno detto che avrebbero aiutato i
    H   H*L   H* H   H*
terremotati gli industriali?
    H H*L   L H%

Sentences 1. and its variants can be translated roughly as follows: "The industrialists must pay the decimals if they want the negotiations to continue"; sentence 6 as follows: "The unions cannot accept this agreement"; sentence 9 as follows: "Who did the industrialists say they intended to help?" and finally 10 as follows: "Have the industrialists said they intended to help the earthquake victims?"
As to the underlying phonological model, the reader is referred to [1].

In Pierrehumbert system [2], only two tones in combination make up the intonational contour specification: T* where T=H,L, the star indicates alignment with the prominent syllable. As a first approximation we adopt P's binary notation, plus the tone associated with IG's boundaries: T%=H,L where H% is usually associated with yes/no questions and L% marks the end of non-interrogative IG's. As in her system, when focus is associated to a prominent syllable there is a couple of tones which appear, as follows: [+Focus]---> H*l; HL*. It must be noted that the other two allowable sequences (L*H, LH*) are less frequent in Italian, or belong to emphatic and contrastive utterances which we do not take into account here. Also, we did not see the need for introducing a phrase accent, which should accompany the final nuclear pitch accent as happens in English.

ANALYZING THE DATA

From a linguistic point of view we can divide sentences into two parts: the one following and the other preceding the focussed constituent. First of all we look at the sentence section following the focussed constituent, which on a first approximation we take it to coincide with the rightmost T*T/TT* tonal marking. The portion to be considered varies remarkably from one sentence to another: it is constituted by a subordinate clause in sentences 1. and 3.; the subordinate clause plus what remains of the major clause, once the topicalized constituent has been fronted, in sentence 2. the subordinate clause plus the constituent which has been extraposed, either the subject or the object NP of the main clause, in sentences 4. 6. 7.; the presupposed relative clause attached to the clefted constituent in sentence 8.; the extraposed NP subject in sentence 5.; and the right dislocated NP object in sentence 4.
All this sentence material can be treated homogeneously from an intonational point of view even though it contains syntactic and semantic elements differing quite markedly from one another. These components of the intonational structure can be opposed to the material which precedes the focussed constituent/s which we discuss below. The phonetic characterization of post-focus linguistic elements can be defined as follows: there is a downstep pattern in the F. contour which reflects a somewhat global line starting from an upper limit and reaching a baseline value about 5 half-tones below it (hereafter h.t.)[1]. The declination line associated with each such portion of F. patterns does not lend itself easily to defining a constant decaying rate. In fact, lowering seems to apply randomly to prominent/non-prominent syllables looking at sentence stretches of a certain syllable length. Local variations may take Phonological Words[2] as their domain, with the only restriction that local F. jumps cannot override F. jumps of the previous PW. The first sentence, declarative, is made up of two IG's, the first of which ends with the main sentence and the second with the subordinate clause. Only the main clause contains focussed material, i.e. the assertion of the underlying semantic proposition; the subordinate expresses an hypothesis based on given information. In sentences moving the focussed constituent to the front, after FAR has applied, the declination line is set at approximately .5 h.t. above the final value(L%). Also these sentences (see 2 and 3) are made up of two IG's, the first of which ends with the main sentence and the second starts with the subordinate clause. The only noticeable difference from the simple declarative consists in the decrease in the final lowering at the end of the first IG: the degree of final lowering is higher in the simple declarative than in its marked variants and this is due to differences in semantic representation. In the former case, the main clause contains an assertion and the whole proposition constituted by the main predicate and the subject of predication are elements of focus structure: the pitch range correlated with the main sentence is higher than the one correlated with the subordinate clause. Marked variations of this utterance concentrates the predication onto a single constituent which marks focus structure: in sentence 2 it is the object NP, as in sentence 8; it is the VP in sentence 6 and the subject NP in sentence 7, and so on. As in [5] focus is the representation of the variable x such that $P(x)$, where $P(x)$ is a predication in x corresponding to the dominant or main Verb. What is needed then is a coindexing rule to associate the predication with the entity in focus, as in [6]: Coindex NP and X where X = an AP, PP, NP, VP or S. Coindexing tells us which thing x is being predicated about. In case of sentence 2, a topicalization, what we have is:

2i. $[_{NPi}$ decimali$]_1$ $[_{VP}$devono pagare gli industriali$]_1]$
$$X$$
$$FOCUS=X=[_{NPi}\text{ decimali}] \qquad P(X)$$

As Berwick rightly remarks: "there is certainly not much in the way of constraint in this proposal. What is missing is the machinery telling us which NP's and X's are to be coindexed"(ibid.,53). This would require discourse structuring rules, obviously; but at sentence level a lot can be done in Italian on the basis of syntactic structuring, as discussed above.
We are left with the portion of the intonational contour which precedes the focussed constituent. From a phonetic point of view, to achieve a satisfactory definition of focus it is not sufficient to look at peaks in the pitch contour.

Variability in the topline -- or the maximum value for the F. contour in a phonological phrase = which can be constituted either by a peak, H, a maximum, or a fall to a very low pitch, L, a minimum in the pitch range of a given intonational contour in absolute terms, do not constitute the correlate of the focussed constituent. Other factors not relatable to focus can contribute to the creation of peaks, such as the length of the utterance or the beginning of a new discourse. We found and verified in synthesis experiments, that the steepness of the dipping following/preceding the focussed segment (usually a syllable), i.e. the rate/s of change or number of h.t. for the segment/s constituting the sequence relevant to the definition of focus structure, is the viable discriminating correlate of focus. In this way focus is defined as a relation over two adjacent tonal assignments, in terms of the steepness of the dipping of the F. contour. If we look at our examples, we find easily that in the first portion of the sentence there are two or even three combinations of T*T/TT* - and indeed, potentially there could be an infinite number. Only if we adopt our criterion we can account for sentences in which two or more constituents seem to be structurally marked and semantically relevant in the overall informational structure. This is the case of sentences 3,4,5 in which a constituent is moved to TOP position or is left/right dislocated and is bound to a resumptive clitic within the sentence, as shown in:

iii. $[_{S'}[_{TOP}[_{NP}$Questo accordo$]$ $[_{S}$non$[_{NP}$e$]$ $[_{VP}$possono$[_{VP}[_{V}$accettar$][_{c1}$lo$]]]][_{NPi}$ sindacati$]]]$

The constituent in TOP does not count as new information as is the case with topicalized sentence 2. Rather, it qualifies as secondary focus even though it has been fronted: primary focus is associated with the VP and is marked as H*LL% at the offset of the IG.
The grammatical representation is thus confirmed by the data we collected in that focus is characterized by three features: a L/H peak/fall, aligned with word-stress, accompanied by a preceding or trailing H/L tone followed by a L tone in coincidence with an IG boundary(not necessarily), the steepness must be the highest in the sentence. If we look at the steepness we have the following data: in sentence 3. HL*=6 h;t. whereas H*LL%=8 h.t.; in sentence 4. HL*=3 h.t. but H*L=7 h.t.; in 5. HL*=4 h.t. but H*LL%=8 h.t.; in 2. the steepness associated with H*L=9 h.t.; in 8. H*LL%=8 h.t.; in 6. H*L=9 h.t. and finally in 7. H*L=9 h t.

We shall concentrate now on the two interrogatives: the wh- question in 9 and the yes/no question in 10. As to 9 we note that the wh- word constitutes the questioned object and the NP subject "gli industriali"

is extraposed beyond three sentence boundaries, as shown below:

ii. chi [shanno detto] che [shanno intenzione] di [saiutare] .

The intonational contour of the wh- question is clearly identifiable in that it doesn't possess a final peak nor a single peak at the onset: in wh- questions all the fronted constituent is in focus and is raised to a H plateau. What follows is a very steep F. drop: 10 h.t. in our examples. This pattern sharply separates the remaining sentence portion, which is uttered on a low declination line. It must be remarked that wh- words do not possess word stress unless they are contrastively emphasized – so they build a PW with the following head in this case the word "detto".

On a semantic level, wh- questions are partial questions and the H portion of the sentence is solely constituted by the questioned material, the remaining part of the question no longer constituting part of the question because presupposed or already known. On the contrary, yes-no questions are total questions and the whole sentence is uttered homogeneously on a H level.

1 This characterization of F. variations in terms of half tones has been suggested to me by G.A.Mian and G.Tisato; each half-tone

corresponds to ΔF/F=6%.

2 We define a Phonological Word as a structural component of IGs made up of one stressed lexical element, the head of the PW, preceded by as many unstressed lexical elements as there are within a Phonological Phrase Phonological Phrases in turn correspond to major syntactic constituents(see Selkirk, 1984).

BIBLIOGRAFIA

[1] Delmonte R.(1983), A Phonological Processor for Italian, in Proc. 1st Meeting European Chapter of the ACL, Pisa, 26-34.
[2] Hirschberg J., J.Pierrehumbert(1986), The Intonational Structuring of Discourse, Proc.24th Annual Meeting of ACL, Columbia Univ.New York, 136-144.
[3] Delmonte R., G.A.Mian, G.Tisato(1986), A Grammatical Component for a Text-to-Speech System, Proc.ICASSP 86, Tokyo, 45.2.1,2407-2410.
[4] Selkirk E.(1984), Phonology and Syntax, The MIT Press, Cambridge Mass.
[5] Berwick R.(1983), Introduction: Computational Aspects of Discourse, in M.Brady & R.Berwick(eds), Computational Models of Discourse, The MIT Press, Cambridge Mass., 27-106.
[6] Williams E.(1980), Predication, Linguistic Inquiry, 11(1), 203-238.
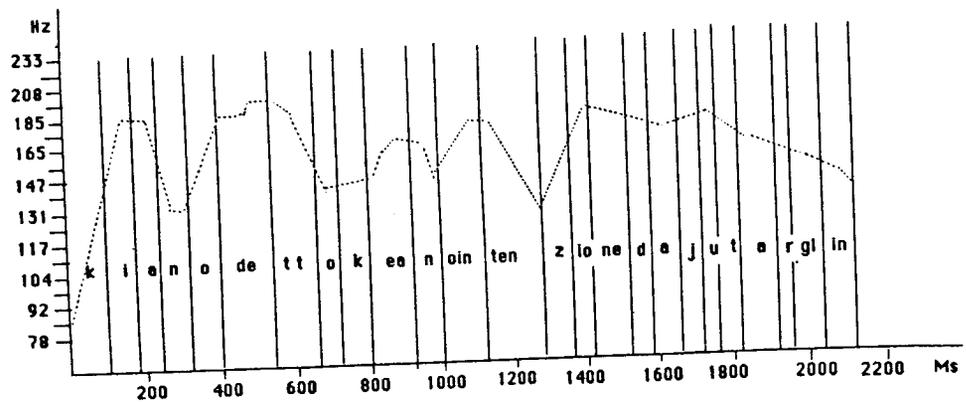
Fig.1 F° plot of wh- question
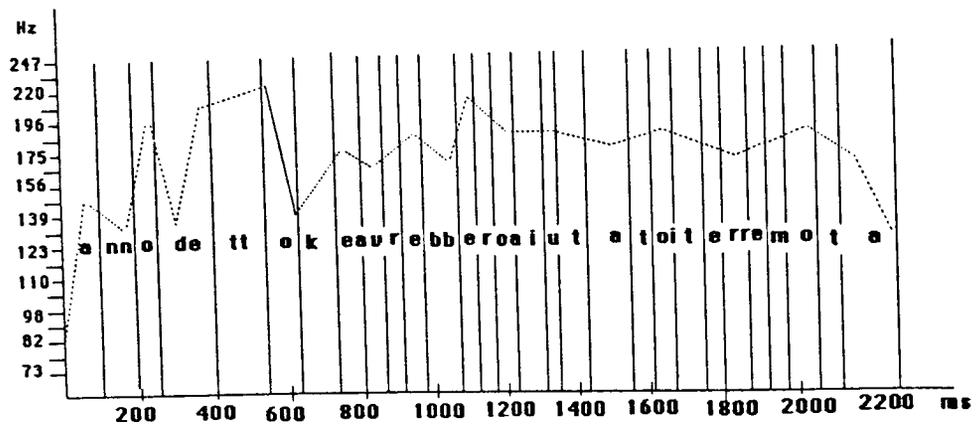
Fig.2 F° plot of yes/no question

Se 24.1.4