

Sentence modes and prosodic phrasing

Patrizia Bonaventura and Per Aage Brandt***

*Department of Communication Sciences

**Department of Cognitive Sciences

Case Western Reserve University, Cleveland, OH

{pab72; pab18}@case.edu

Abstract

The goal of the present study was to test the hypothesis that specific prosodic features are systematically associated with modes of enunciation, and that such prosodic features are independent from the semio-syntactic structure of the sentence. Modes of enunciation correspond generally to sentence types: Assertive, Imperative, Interrogative and Affective (Brandt, 1971; 2008). Semio-syntactic structures are defined as elementary grammatical structures composed minimally of a head and one of the core complements, Subject, Predicative, Object, Dative and Agent. Modes of enunciation are associated with semio-syntactic structure at the level of linearization of meaning, when words and morphemes are selected to communicate the semantic content. Intonation contours associated with different modes of enunciation, were categorized based on the Tilt labeling system. Two groups of 3 sentences with two different semio-syntactic structures, were realized with the four different sentence modes. All sentences were read once by 4 professional actors.

Preliminary results seem to show a systematic expression of the sentence mode at the prosodic level, across speakers, for Assertive, Imperative and Affective mode (Anger) only. Recurrent intonation contours appeared not to be exclusively related to the mode of enunciation, but to be typical to the semio-syntactic structure, supporting the hypothesis that prosody connects phonetics to semantics, or semantics to phonetics, through grammar.

1. Introduction

The goal of the present study is to test the hypothesis that specific prosodic features are systematically associated with modes of enunciation, and to verify whether such prosodic features are dependent from the semio-syntactic structure of the sentence.

Modes of enunciation correspond grossly to sentence types; in particular, according to Brandt's theory (1971), four modes can describe all possible variations in enunciation: Assertive, Imperative,

Interrogative and Affective. Different modes of enunciation convey information from semantic levels of language organization, which reflect in intonation.

However, also another level of language organization, namely, the semio-syntactic, or stemmatic one, reflects its information content in intonation contours. Semio-syntactic structures are defined as elementary grammatical structures composed minimally of a head and one of the core complements, Subject, Predicative, Object, Dative and Agent (Brandt, 2004).

The more general hypothesis tested in the present study, is that prosodic structure connected to Modes of Enunciation (Brandt, 1971; 2008) is determined by information generated at the levels of organization of meaning, namely, the "Semantic frame structure" level (III), "Context" level (IV) and "Illocution" level (V) (see 1.1). Information content generated at these levels would be expressed, at the Linear level (I) of organization of language, by the Modes of Enunciation, through their association with specific prosodic characteristics. The goal of the present study is to verify whether modes of enunciation are systematically associated with some specific prosodic features, in terms of intonation contours measurable by acoustic parameters, by taking into account also the effects of the Stemma level (II), which carries information about the semio-syntactic properties of the sentence, and also generates content that surfaces in prosody.

The present study was designed to isolate the effects of the Semantic /Semiotic/Illocution levels and the effects of the Stemma level on the resulting general intonation contours, in order to identify prosodic effects possibly expressing only information relative to Modes of Enunciation.

Before describing the methodology adopted, the architecture of the Brandt's (1971; 2008) model of organization of levels of language, meaning and thinking, will be described (see 1.1); the relationship between modes of enunciation and prosody will also be outlined in more detail as well as the relationship between semio-syntactic structures and prosody (see 1.2).

1.1. Levels of organization of meaning and language

The hypothesis adopted for this study, that prosody connects phonetics to semantics, or semantics to phonetics, through grammar, is based on a general model of organization of levels of meaning and language (Brandt, 1971, 2008). According to this model, five levels of organization have to be defined, in describing the relation between language and meaning:

I) Linear structure, or concatenation of lexematic and morphematic entities under a prosodic profile.

II) Grammatical node-structure ("stemmatic structure") accounting for constituent syntax; this level is interpretable in both directions: the node cascades can be linearized by rule-based projection, and the node cascade can be interpreted semantically. It integrates morphology.

III) Semantic frame structure, utterance dimensioned, accounting for agency, motion, change, and exchange. This level structures semi-equivalences between different constructions. It integrates lexical entities (words of word classes).

IV) Phenomenological structure (Context level): experiential domains of concepts shared by speakers, underlying frames (physical, social, communicational, mental; higher order contexts).

V. Epistemic structure (Illocution level): the speaker's and hearer's actual topic for thinking, related to a narrative circumscribing the situation of thinking and speaking.

The model postulates that the information encoded at the levels of organization of meaning (Semantic Frame, Context and Illocution, or III-V), is expressed through the modes of enunciation.

Modes of enunciation, in their turn, surface at the Linear level (I) of production of speech acts, through prosody; in particular, the hypothesis is tested here that the F0 contour of the sentence would reflect the information relative to the modes of enunciation, as well as also information conveyed by the Stemmatic level (Bonaventura and Brandt, in preparation).

The focus of the observation for the present study is to identify F0 contours related to the meaning structure; therefore, variations due to linguistic structure have been controlled by selecting only two simple semio-syntactic structures, a Transitive, and a Ditransitive, for the sentences in the corpus.

1.2. Semio-syntactic structures and prosody

Semio-syntactic structures are defined according to the semio-syntactic model adopted as a theoretical framework in the present study (stemmatic syntax; Brandt, 2004): this model assumes that we cognize language in the act of using it, and we intuitively understand that the content of sentences is a semantic whole that forms a part of a larger whole of thought. The grammatical structure of a sentence, i.e. its syntax, is integrated in a whole of thought or meaning: in this sense, syntax is a meaningful "instance" in the architecture of language. Syntactic structures are accessible to the language user, and they are projected into and extracted from the structures of the "instance" of phonetics. A similar process of projection and extraction connects it to the "instance" of thinking we usually call semantics. The stemmatic view of language assumes that projection and extraction are essential processes at work between all three basic "instances"—phonetics, syntax, and semantics. Stemmatic representations of the grammatical relations shaping a sentence as a network of variably interdependent components, words and phrases that are networks of words in the same sense, can be represented by graphic models that reflect our intuitive understanding of grammatical relations (Brandt, 2004).

Grammatical sentences and phrase constructions in a wide range of languages can be relevantly analyzed in this format to form a canonical complement cascade, a so-called stemma, whose nodes are stably linked to a finite set of semantic types of complementation, invariant across different constructions and even different languages. The stemmatic syntactic model represents basic semantic operations of a construction – phrase or sentence, as a cascade of operations of complementation preceded by an initial element, a 'head' that serves as an anchoring reference for the operators ('marks') that determine the linear form of phrases (constructions) and sentences (constructions of constructions).

The generation of phrasal structure is based on an order of dominance presided by a finite verb, under which the complements (considered as "actants" and "circonstants") are organized. The semantic nodes corresponding to the complements under the finite verb (head) are generated according to the following order:

1. Subject complement
2. Predicative complement
3. Object complement
4. *Telos* complement (i.e. indirect object, as dative)
5. *Arche*' complement (i.e. "agent", or origin of action)

6. *Topos* complement (i.e. “time” and “place” adverbial expressions)
7. *Logos* complement (i.e. adverbial categories of “logical” determination, or “manner”)
8. Junctive complement (expresses “coordination” or “juxtaposition”)

Minimal semio-syntactic structures, as the ones used in the corpus for this experiment, are defined as elementary grammatical structures composed minimally of a head and one of the core complements, Subject, Predicative, Object, Dative and Agent.

The stemmatic syntax perspective assumes the primarily non-referential character of signified meaning and therefore recognizes the “instance” of linguistic semantics as grounded in the human imaginary and in human acts of communication, including gesture, prosody, and the production and exchange of lexicalized utterances. The fundamental role of (stemmatic) syntax is thus to let language combine linear (sequential) order and conceptual (iconic) order into constructions with both phonetic and semantic properties. Prosodic intonation of constructions can be considered as a phonetic indicator of specific syntactic structure; so that differences in syntactic organization will correspond to different prosody.

Prosody connects phonetics to semantics, or semantics to phonetics, through grammar. In particular, the deeper levels of organization of meaning (semantic, contextual and illocutionary see 1.1), are assumed to express their informational content through modes of enunciation.

1.3. Modes of Enunciation and prosody

The inscription of subjectivity in grammar is a prominent motivation for the existence of prosodic features of manifestation, since a natural prediction will let the foregrounding or backgrounding of subjectivity be a plausible reason for theatrical intonations referring to the speaker.

Four modes of enunciation (speaker’s cognitive attitude to content) are particularly prominent: the **Imperative** (1), the **Interrogative** (2), the **Assertive** (3), and the **Affective** or exclamative mode (4). These modes can be characterized by two semantic features: (1) + subjectivity, – reality of content; (2) – subjectivity, – reality of content; (3) – subjectivity, + reality of content; (4) + subjectivity, + reality of content. (1 – 4) are strongly differentially marked in sentence prosody in most languages. The morphology of mode – opposing the subjunctive (S) and the indicative (I) modes – is an indirect version of the same categories; so in the Romance languages, the

binary morphological contrast S/I expresses the quaternary semantics of modes. The subjunctive mode expresses in Latin (1) and (2), in French (1) and (4), while in Spanish and Italian it can express (1), (2), and (4). Our question is whether the indirect manifestation of mode of enunciation is also, or can also be, marked by specific prosodic features comparable to the situation in direct manifestation.

The distinction between modes of enunciation may concur, together with corresponding semantic and syntactic structures, to determine the intonation contour of a sentence.

The primary goal of the present study is to investigate how intonation contours are associated with given modes, and whether intonation contours that seem to be associated with certain modes of enunciation, are also dependent on specific semio-syntactic structures.

The secondary goal of this study is to identify linguistically meaningful parameters in intonation, and their relative acoustical correlates, that can be used in automatic synthesis of prosody, to predict sequences of pitch accents related to sentence modes.

1.4. Intonational models and prosodic categories

Linguistically meaningful information which might be significant for interpretation of an utterance’s intonation, is articulated at several levels: syntactic structure complexity and ambiguity, distinction between new/given, focus/comment, presupposition/assertion information, as well as content relative to speech acts, or sentence types. A model of intonation flexible and comprehensive enough to account for all the information deriving from these levels of linguistic and meaning structure, is not available.

Most common phonological categories for description of the pitch accents (Ladd, 1996; Pierrehumbert, 1980; Beckman et al., 1993), do not allow to use labeling and categorizations for the purpose of automatic stylization of prosody: some categories are not associated with F0 correlates (e.g. level accents), and the criteria for categorization are too subjective to provide a reliable system that can consistently be used for automatic labeling of intonation contours in natural continuous speech (Taylor 2000b). Furthermore, some types of labels are used to represent a great variety of intonation contours that include very different linguistic information.

In the present study, an attempt is made to define more systematically the levels of information reflected in intonation, and to associate each level to specific phonetic parameters, that can be consistently

identified even by an automatic labeling system. In order to define these phonetic parameters, a system that accounts for some phonological distinctions, but also provides precise tools for continuous measurement of acoustic parameters was selected, the Tilt model of intonation (Taylor, 1995; 2000b). The Tilt model adopts some of the labeling categories proposed by the Pierrehumbert (1980) model, but redefines the intonational space in terms of parameters, measurable as amplitude, duration and tilt values (Taylor 2000b).

2. Method

2.1. Data collection and Subjects

Two groups of 3 sentences with two different semio-syntactic structures, were associated with the four different modes of enunciation (e.g. Assertive: 'He bought a book'. Interrogative: 'Did he buy a book?' Imperative: 'Buy a book'). The minimal semio-syntactic structures were Transitive, composed by nodes 1-3 (see 1.2), and verb as a head (H), and Ditransitive, composed by nodes 1-3-4 (Fig. 1 and 2).

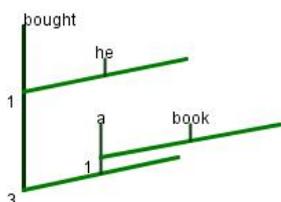


Fig. 1. Stemmatic representation of sentence: "He bought a book" (Transitive)

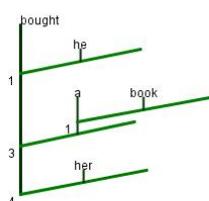


Fig. 2. Stemmatic representation of sentence: "He bought her a book" (Ditransitive)

All sentences were read once by 4 professional actors, and were elicited with instructions to obtain expression of all modes by the same speaker. In particular; the Affective mode was represented by four variations: Anger (generic) Anger (disappointment), Surprise and Endearment.

The sentences were recorded in the Case Speech Production Lab sound booth at Case Western Reserve University, and were pronounced at 3-

seconds intervals. The subjects were 3 men and 1 woman, native speakers of English. The dialectal origin of the participants was not controlled.

2.2. Data analysis

The hypothesis that intonation contours are influenced a) by modes of enunciation (ME) and b) by semio-syntactic structures (SSS) in predictable ways, has been tested by describing the intonation contours in the short sentences, as produced by the different subjects. In particular, hypothesis (a) was tested by comparing intonation contours by mode of enunciation, for sentences with the same semio-syntactic structure (SSS), across subjects; hypothesis (b) was tested by comparing intonation contours for same mode of enunciation, produced by different subjects, across semio-syntactic structures, to verify whether systematic variations in the F0 contour occurred for same ME across SSS. Also, presence of possible consistent associations between intonation contours and nodes of semio-syntactic structures has been observed.

Intonation contours were described by using a set of labels proposed in the Tilt intonational system (Taylor 2000b), based on Pierrehumbert bitonal system (Pierehumbert, 1980); Tilt markers were selected, as the Tilt intonational system allows extraction and measurement of acoustic parameters corresponding to each label, to be possibly used for synthesis applications.

Application of the results to automatic prosodic stylization for synthesis is the secondary goal of this research, and if systematic associations were found between ME or SSS and some intonation patterns, in future studies the categories can be associated with quantitative parameters, like amplitude (as excursion in frequency of rising and falling pitch accents) duration and tilt measure in rising-falling contours, to be used for automatic extraction of prosodic information.

In particular, the following set of markers was used, for this preliminary analysis: a = normal pitch accents; c = minor and level accents; ac = rising /falling accents; ab = rising accent at boundary; acb = rising /falling accent at boundary; cb = falling accent at boundary. Some of these categories were created, based on Tilt parameters, as they resulted necessary for the description of the speech sample.

The sentences were hand-labeled, by the speech processing tool Praat. In particular, three levels of tagging were performed for this study: (a) phonetic, by SAMPA symbols (Well, 1997) (b) prosodic, by Tilt-based symbols (c) semio-syntactic, reporting the nodes corresponding to each syllable nucleus marked by an accent label (see 1.2, and Figg.

1 and 2). An example of the labeling system is shown in Fig. 3:

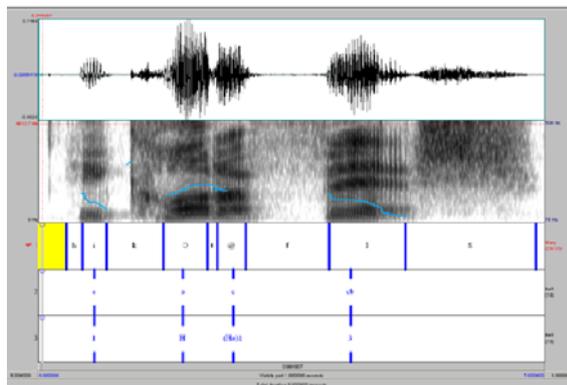


Fig. 3: Example of labeling system: tier 1 = phonetic (SAMPA symbols), tier 2 = prosodic (Tilt labels), tier 3 = semio-syntactic (nodes, see 1.2)

3. Results

Pitch contours and boundaries from all sentences pronounced by the four subjects were labeled according to the system described in 2.2, and contours associated with the different sentence modes by all speakers, at least in 50% of the occurrences, were considered as “recurrent pattern”. Tables 1 and 2 report the results for the Transitive and Ditransitive semio-syntactic structures, respectively.

In Table 1, when only part of the contour was recurrently used, the contour was included in [] and the varying pitch accents have been marked by ‘x’. Table 2 reports the distribution of the sentences, by semio-syntactic type and mode, and the frequency of occurrence of the most recurrent intonation contours by ME and SSS.

ME/SSS	Transitive	Ditransitive
Assertive	x – [c-c-cb] 1-H-1-3	x-[c- c-c-acb] 1-H-1-4-3
Imperative	[a-c-cb] H-1-3	x-[c-c-cb] H-3-1-4
Interrogative	No recurrent pattern	No recurrent pattern
Affective (surprise)	No recurrent pattern ([c-a-c-acb] 42%)	x-x-[c-c-acb] 1-H-1-4-3
Affective (anger-generic)	x- [a-c-cb] 1-H-1-3	x-[c-c-cb] 1-H-1-4-3
Affective (anger-disappointment)	x- [a-c-cb] 1-H-1-3	No recurring pattern (x-x-[c-c-cb]: 45%)

Affective (endearment)	No pattern	x-x-[c-c-cb] 1-H-1-4-3
------------------------	------------	---------------------------

Table 1. Association of Modes of Enunciation (ME) with intonation contours, by semio-syntactic structure (SSS, either Transitive or Ditransitive)

ME/ # of sentences	Transitive (21 sentences x 4 speakers)	Ditransitive (21 sentences x 4 speakers)
Assertive (6 sentences)	60%	73%
Imperative (6 sentences)	75%	55%
Interrogative (6 sentences)	--	---
Affective (surprise) (6 sentences)	--	50%
Affective (anger-generic) (6 sentences)	58%	58%
Affective (anger-disappointment) (6 sentences)	83%	---
Affective (endearment) (6 sentences)	---	60%
Total sentences analyzed: 168	84	84

Table 2. % of association of intonation contours described in Table 1 with Modes of Enunciation

The results show that association of an intonation contour with a specific mode occurs, for both SSS, at least 50% of the occurrences, for Assertive, Imperative, and Affective mode with Anger (generic) connotation; in particular, the highest percentages are associated with the Imperative (Transitive) and Assertive (Ditransitive) modes. Affective mode set as Surprise shows a recurrent pattern with 50% frequency, only for Transitive, whereas Affective as Anger with Disappointment shows the highest consistency for the Transitive SSS, but no recurrent pattern for Ditransitive. Finally, the Affective mode with Endearment shows a recurrent pattern only for the Ditransitive. No recurrent pattern for any SSS, was shown for the Interrogative mode.

A general observation about the results is that percentages of agreement in use of F0 contours are not very high (50%-83%), showing a great range of variability among subjects in the use of intonation contours by mode.

Recurrent patterns do appear and they seem to be typical of different modes (e.g. x-x-[c-c-acb] for

Affective/Surprise (Ditransitive) vs. x-[a-c-cb] for Affective/Anger (both connotations, Transitive). However, no similar patterns across SSS have been found, possibly indicating that meaning information is somehow determined by the higher grammatical level of encoding of language.

The most salient characteristic of the preferred contours in Fig. 1 seems to be the fact that for the Transitive structures, the contours are associated with the H-1-3 sequence at the SSS level, whereas for the Ditransitive structure, the predominant association is with the last 3 nodes of the structure, 3-1-4.

In terms of the description of the function of prosody by the cognitive semio-syntactic model of linguistic analysis (Brandt, 1971; 2008), the two patterns H-1-3 and 3-1-4 can be identified with the two components of the meaning structure of a natural proposition: 'Agents Accessing Objects' (Object clause: H-1-3) and 'Modifying them in view of some final Destination' (Dative clause: 3-1-4). According to the general description of the situational meaning as: (A access O, achieving O → O*, with goal D) (Brandt, 2008), it appears that the meaning information "A access O" might surface as consistent intonation patterns by mode on the H-1-3 component of the semio-syntactic structure; on the other hand, it seems that the other component of the situational meaning "achieving O → O*, with goal D", might surface in the preferred contours associated by mode with the semio-syntactic structure 3-1-4.

These preferences might indicate that the association of prosodic features (or sequences of pitch accents within a certain intonation contour), might be determined by the meaning information, generated at the Level III of the cognitive semio-syntactic model (Brandt, 2008), and filtered through the semio-syntactic structure (Level II, Grammar), that determines the form of its linear representation at the phonetic level. This scenario seems to support the hypothesis that prosody connects phonetics to semantics, or semantics to phonetics, through grammar (see 1.2).

4. Discussion and conclusions

The preliminary results were based on a manual categorization and the restricted number of the samples did not allow statistical analysis. Therefore the results of this study are considered indicative of a trend that might be investigated more in deep in the future.

In particular, other analyses that need to be pursued are testing presence of identified recurrent contours in a larger corpus, produced by both actors

and non-actors. In fact, the controlled conditions of the study (use of actors to consistently pronounce the different modes of enunciation), might not reveal different trends that can occur in pronunciation by non-actors or in spontaneous speech. Also, the dialectal origin of speakers should also be controlled, to minimize socio-cultural effects that might affect expression of the sentence modes.

Further research should focus on measurement of the acoustic Tilt parameters corresponding to the pitch accents in the contours that have shown to recur more systematically, in order to evaluate the range of variation of the amplitude, duration and tilt values for each pitch accent across subjects, for each contour: such analysis would provide a more detailed description of the timing and shape properties of each contour category, as well as of the intersubject variability in their realization.

Finally, an observation of linguistic interest would be the analysis of the correlation of the rhythmic structure of the sentence to the intonational and semio-syntactic labeling. The metrical structure of the utterance might be related to the syllabic (and morphological) composition of the words, and changes in the prosodic patterns at the syllabic level, might also intervene in connection with the prosodic changes related to the meaning and semio-syntactic structure levels.

5. References

- [1] Beckman, M. E. and Ayers, G. M. (1993). Guidelines for ToBI labelling. Technical Report Version 1.5, Ohio State University.
- [2] Brandt, P. A. 2008. "Thinking and language. A view from cognitive semio-linguistics", paper presented the Speech Prosody 2008 Conference, Sao Paulo, May 6-9, 2008.
- [3] Brandt, P. A. 2004. Toward a Cognitive Semiotics. In Fastrez, P. (ed.), *Recherches en Communication*, 19, *Dossier Cognitive Semiotics*, Université Catholique de Louvain
- [4] Brandt, P. A. 1971. Mode, textualité: note sur la modalité romane, *Revue Romane* VI, fasc. 2, Copenhagen.
- [5] Ladd, D. R. D. (1996). *Intonational phonology*. Number 79, in *Cambridge Studies in Linguistics*. Cambridge University, Press, Cambridge, UK.
- [6] Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. PhD thesis, MIT, Published by Indiana University Linguistics Club.
- [7] Wells, J.C.], 1997. 'SAMPA computer readable phonetic alphabet'. In Gibbon, D., Moore, R. and Winski, R. (eds.), 1997. *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter.
- [8] Taylor, P. A. (1995). The rise/fall/connection model of intonation. *Speech Communication*, 15:169–186.
- [9] Taylor, Paul A. 2000b. Analysis and synthesis of intonation using the tilt model. *Journal of the Acoustical Society of America*, 107(3):1697--1714.