

The definition of a descriptive space of Italian prosodic forms: The CALLIOPE model

Licia Sbattella^{1,2}, Roberto Tedesco¹ and Sonia Cenceschi¹

¹ Department of Electronics, Information and Bioengineering, Politecnico di Milano, Italy

² Fondazione Sequeri Esagramma Onlus, Italy

{licia.sbattella|roberto.tedesco|sonia.cenceschi}@polimi.it

CALLIOPE (Combined and Assessed List of Latent Influences On Prosodic Expressivity) is a model that aims at categorizing all prosodic forms. Following Cresti (2000), we will call sentences as Information Units (UI). Each UI has a surjective correspondence with a specific prosodic unit, and conveys a specific informative intention. With our model, we aim at providing a list of all possible factors that affect the prosody, and consequently the interpretation, of every UI. The CALLIOPE model defines a multidimensional “space”, where each “dimension” represents a characteristic influencing the vocal paralinguistic components of UIs. Each characteristic is actually a categorical variable, assuming values in a set of labels; for example, the dimension Emotion contains the labels *sadness, joy, fear*, etc. The following table shows the 11 dimensions of our space.

Dialog		Background		
F ₁ . Structure F ₂ . Linguistic modality F ₃ . Intonational focus F ₄ . Rhetorical form	F ₅ . Motivational state F ₆ . Speech mood F ₇ . Spontaneity F ₈ . Emotion	F ₉ . Subjective expressiveness skill	F ₁₀ . Social context	F ₁₁ . Language, dialect or local accent form

Each UI is thus associated to a “point” into this space; more formally, a generic UI_i is associated to a tuple T(UI_i) composed of 11 labels:

$$T(UI_i) = (l_1, l_2, \dots, l_{11}) : l_i \in F_i, 1 \leq i \leq 11$$

Not all combinations are possible or commonly used: for example, it's not possible to produce a sentence like “Giovanni ama Maria!” (D'Imperio, 2002) as *exclamatory* (a label of F₁) and *ironic* (a label of F₄) at the same time; this means that several tuples are never observed or (which is the same) our space contains several points where no UIs can be associated. On the other hand, we argue that, for a generic UI, it is possible to select a precise label for each dimension of our model. It's important to underline that each tuple T(UI_i) corresponds intuitively to more than a single UI, while the opposite is not true: a single UI can't be associated to more than one tuple (surjectivity).

The CALLIOPE dimensions are divided into two groups: Dialog and Background; the first group lists 8 dimensions strictly related to the interaction situation, where each one takes value from predefined and closed label sets; the Background group considers only 3 dimensions, which do not depend on any specific interaction and take values from predefined but open label sets.

Usually, experiments on intonation start choosing a particular and narrow topic, selecting participants, grammatical forms, social context, parameters to be analysed, etc. Looking at these choices from the CALLIOPE perspective, each experiment regards a well-defined “sub-space” of the model.

The CALLIOPE dimensions

(F₁) Structure: *Declarative, interrogative with 1 tonal unit, interrogative with 2 or more tonal units, interrogative disjunctive, echo questions, exclamative, vocative, lists.* Note that the structure of the UI does not always correspond to the same intonation for all languages. For example, not all languages use a final rise in intonation to indicate a question. These labels were organized according to Prieto et al. (2010-2014).

(F₂) Linguistic modality: 24 labels divided into five main groups, as explained in Kratzer (2012). According to Heusinger et al. (2011), “*Modality is the category of meaning used to talk about possibilities and necessities, essentially, states of affairs beyond the actual*”.

(F₃) Intonational focus: *presentational, contrastive, counterpresuppositional, definitional, contingent, reactivating, and identification.* These labels have been extracted by Gussenoven (2007) and carry a pragmatic function. Such labels fit well to intonational languages; about tonal languages, we are aware of on-going studies about the influence of intonational focus (Ouyang 2012).

(F₄) Rhetorical form: we consider only rhetorical forms that, we argue, influence the prosody of the sentence: *irony, aposiopesis, repetition*. The *non-rhetorical* label is also included, for UIs lacking any rhetorical form.

(F₅) Motivational state: as in the Liotti (2008) theory of SMI (Interpersonal Motivational Systems), motivations within interpersonal exchanges are analysed in an evolutionary perspective, taking into account psychotherapeutic dialogues. We can assign a motivational label to each UI, as shown in the AIMIT (Fassone et al., 2012) manual: *ranks of dominance or submission, care-seeking, sexual bonding, and caregiving*.

(F₆) Speech mood: *whispered, normal, and shouted* speech. We argue that such labels permit to represent common situations. Notice that the mood is often confused with Emotions and Linguistic modality. In our opinion, however, the mood represents a distinct, independent characterization of the situation where the dialogue is situated.

(F₇) Spontaneity: *spoken, read, and recited* speech. These labels reflect the typologies found in speech corpora.

(F₈) Emotion: as in Plutchik, R. (2001), we choose: *Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise*. We are aware that classifying emotions is controversial and no definite list exists as explained in Cowie (2000). We argue, however, that the group we propose –together with Linguistic modality– could describe the vast majority of cases.

(F₉) Subjective expressiveness skill: it's the level of personal skills the subject shows during the dialog. It does not indicate the level of schooling, which –we argue– does not affect the prosody but only the linguistic skills. We defined some labels (e.g., *without communication deficit*) but the set is obviously open.

(F₁₀) Social context: some special social contexts affect speakers' prosody. For example, a nursery teacher speaking with children, a politician explaining his plan in public, or a priest reciting a religious litany have different, emphasized or flattened prosody. We defined some labels (e.g., *daily situation*) but the set is obviously open.

(F₁₁) Language, dialect or local accent form: According to the main Universals in Language (Cristofaro, 2010), speech derives from the need to express global concepts (e.g., talking about past or future events) that are common to all humans. On the other hand, such phenomena have different effects on different linguistic systems (Sadock & Zwicky, 1985). For example, Cheang (2009) shows that the sound of sarcasm changes among different languages. Thus, we need to add to our model an explicit dimension where we represent each language variety as a label.

We are currently planning to start the construction of a prosodic corpus recording audio samples in collaboration with actors of *Libro Parlato Onlus*¹. The purpose is the acoustic characterization of a first narrow group of frequently used prosodic forms, to be used as models in clinical multimodal training sessions². We will start focusing on *standard Italian* (a label of F₁₁), as defined by Canepari (1986), recording professional actors *without communication deficits* (F₉) and simulating *daily situation* (F₁₀), thus defining a point into the Background subspace. About the Dialog subspace, we will restrict our samples to *non-rhetorical sentences* (F₄), *normal mood* (F₆), and *recited speech* (F₇).

References

- Cresti, E. (2000), Corpus di italiano parlato, Accademia della Crusca.
- D'Imperio, M. (2002), Italian Intonation: an overview and some questions, Probus (special volume on intonation in Romance languages), 14(1), pp. 37-69.
- Prieto, P., Borràs-Comes, J., and Roseano, P. (Coords.) (2010-2014). Interactive atlas of Romance intonation. Web page: <http://prosodia.upf.edu/iari/>.
- Kratzer, A. (2012), Modals and Conditionals: new and revised perspectives (Oxford Studies in Theoretical Linguistics), Oxford University Press.
- Heusinger, K. von, Maienborn, C., and Portner, P. (Eds.) (2011), Semantics - An international handbook of natural language meaning, vol.2, De Gruyter Mouton.
- Fassone G., Valcella F., Pallini S., et al. (2012), Assessment of Interpersonal Motivation in Transcripts (AIMIT): An inter- and intrarater reliability study of a new method of detection of interpersonal motivational systems in psychotherapy, Clinical Psychology & Psychotherapy, 19(3), pp. 224-234.
- Gussenhoven, C. (2007), Types of focus in English, in Lee, C., Gordon, M., and Büring, D. (Eds.) Topic and Focus; Studies in linguistics and philosophy, vol.82, pp. 83-100, Springer.
- Ouyang, I.C., Kaiser E. (2012), Focus-marking in a tone language: prosodic cues in Mandarin Chinese, in LSA Annual Meeting Extended Abstracts 2012, vol. 3, pp. 8:1-5.
- Liotti G., Monticelli F. (2008), I sistemi motivazionali nel dialogo clinico. Raffaello Cortina editore.
- Plutchik, R. (2001). The nature of emotions. Am. Sci. 89:344–350.
- Cowie, R. (2003), Describing the emotional states that are expressed in speech, Speech communication, 40(1-2), pp. 5-32, Springer.
- Cristofaro, S. (2010), Language universals and linguistic knowledge, in Jae Jung Song (Ed.) The Oxford Handbook of Linguistic Typology, pp. 227-249.
- Sadock, J. M. and Zwicky, A. M. (1985), Speech Acts Distinctions in Syntax, in Shopen, T. (Ed.), Language Typology and Syntactic Description, pp. 155-196., Cambridge University Press.
- Cheang, H.S., Pell, M.D. (2009), Acoustic markers of sarcasm in Cantonese and English, The journal of the acoustical society of America, 126(3), pp. 1394-405.
- Canepari L. (1986). Italiano standard e pronunce regionali. CLEUP.

¹ *Libro Parlato Onlus*, www.libroparlato.org

² See the LYV project, at the Politecnico di Milano, Polisocial Award 2016; www.polisocial.polimi.it/en/projects/.