

Prosody and expressiveness marking in Bàsàa oral discourse: The case of melisms (First results)

Emmanuel-Moselly MAKASSO

Laboratoire Parole et Langage, Université de Provence

emmanuel.makasso@lpl-aix.fr

Abstract

Though prosody is studied since at least several decades, it is still unclear what kind of information about the speaker's subjectivity and emotion is conveyed by prosody. This paper tries to focus on the role of prosody in marking speaker's expressiveness in large Bàsàa spontaneous speech corpora. Following Caelen-Haumont & Bel 2000, this paper considers that prosodic organization is twofold: on one hand, it is related to the overall organization of phrasing at the group and sentence levels, that is namely the intonation structure; and on the other hand, it conveys information about the affective status and expressiveness of the speaker, that Caelen-Haumont et Bel (2000) termed *melism*. In this work the Bàsàa language is concerned. Bàsàa is a tone Bantu language spoken in Cameroon, which has never benefited from any study using automatic processing. A corpus of conversational speech from radiophonic interviews was collected. Such recordings, coming from spontaneous dialogues, were acoustically analyzed by means of the PRAAT software. Specifically, a script (Caelen-Haumont and Auran, 2004) was employed to automatically extract F0 information from large corpora, especially in high peaks. The term *Melism*, borrowed from the domain of singing, refers here to large or maximal excursions of F0 on words (lexical or grammatical), spreading at times over an entire word. Semantically and/or pragmatically speaking, melisms constitute clues for discourse interpretation and more precisely, for interpretation of the speaker's communicative and informative intentions.

Index Terms: Prosody, oral discourse, expressiveness, melism

1. Introduction

Like it is the case for many other Bantu languages, the linguistic description and analysis of Bàsàa suffers a great lack as far as computing or automatic processing are concerned. This work tries to contribute to fill this gap, with an analysis of large corpora of oral spontaneous speech. More specifically, we analyze large Bàsàa speech corpora with the help of computing for prosody automatic generation, and then we analyze the sense conveyed by the prosodic features obtained from the processing. Over the last years, many questions have been addressed concerning prosody as well as the interface between prosody and the other linguistic and pragmatic domains. This phenomenon resulted in the organization of many conferences (Interfaces prosodiques, Interfaces Discours-Prosodie, Speech Prosody, Journées d'étude de la prosodie). However, many questions are still open. In this article, the role of prosody in conveying information on the speaker's expressiveness in spontaneous oral discourse is analyzed. This information refers to her or his communicative intention and affective status.

The basic assumption on which this study lies is that prosody organization is twofold: on one hand, it is related to the overall organization of phrasing at the group and sentence levels, that is namely the intonation structure; and on the other hand, it conveys information about the affective status and expressiveness of the speaker, that Caelen-Haumont et Bel (2000) termed *melism*. This work is grounded on the Bàsàa language. It is a tone Bantu language spoken in Cameroon, coded A 43 a by Malcolm Guthrie. We collected a corpus of conversational speech in radiographic interviews. Our recordings were acoustically analyzed by means of the Praat software and a specific tool running under Praat, MELISM.

2. Melisms

The term *melism* does not belong to the traditional linguistic or prosodic terminology, but to the singing one. It refers to a melodic figure spread over the duration a word, such that the number of notes perceived is higher than the number of syllables in the word. Melisms do occur in spontaneous spoken discourse, in such a way that this prosodic space is embedded in the intonation structure, while it is nevertheless mainly grounded on another specific prosodic structure, i.e. the affective layer. The melism space is thus an open tribune for subjectivity, because the speaker reveals himself as an individual, with his communicative and informative intentions. Caelen-Haumont (2002) distinguishes the concept of melism from that of intonation as such:

Prosody in our perspective is characterised by the interaction of two prosodic processes, one the support of linguistic structure, the other the breakdown of this support; [...] intonation actualises the forces of association, while lexical prosody actualises the forces of dissociation.

2.1. The MOMEL-MELISM script

The Momel-Melism script is a discrete melodic coding running under PRAAT (Boersma, 2001). See Caelen-Haumont & Auran (2004) for the script and details. This script enables to analyse melodic prominence with a sufficient accuracy. Generally applied to short linguistic sequences such as words, one can also apply it to longer ones, no matter their length.

The script requires:

- A segmentation of the sound file into relevant linguistic units for analysis (words)
- A stylisation of the pitch curve by identification of target points, which is realised by the Momel Algorithm (Hirst, 93). This algorithm breaks the pitch curve into both microprosodic and macroprosodic elements, depending on its length.

The Momel-Melism procedure produces textgrids with six tiers, as shown bellow, with from top to bottom:

- the Praat manipulation window displaying the speech signal and the stylised melodic curve
- the melism coding showing from top to bottom:
 - a segmentation into words
 - an annotation into monotonal or bitonal sequences
 - a recoding within the segmented sequence
 - a tonal target coding following the Melism procedure
 - a conversion into semitones
 - the value of F0 in Hz.

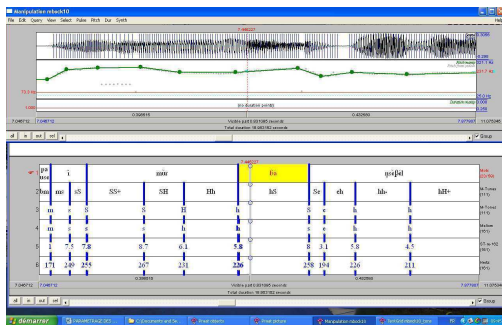


Figure 1: a sample of textgrid labelled by the Momel-Melism script

The Momel-Melism script generates automatically such textgrids. The procedure gives a phonetic representation of the speech signal in 9 levels: Acute (A), Supra (S), High (H), Elevated (E), Mid (M), Centered (C), Bottom (B), Infra (I) and Grave (G). Melisms are restricted to the first three highest levels, A, S and H.

2.2. The corpus

The corpus in this study consists in radiophonic interviews of 6 people, 1 female and 5 males. They were recorded either in a studio or inside houses. The recordings were then divided into wave files under Praat, with variable duration. Table 1 below shows the number of files and the duration par speaker.

The themes all the speakers were asked to develop concerned either themselves or the Bāsàa community the speakers belong to and to which they are strongly attached. These are:

- An anthropologist asked to describe the Bāsàa culture;
- A traditional ruler asked to discuss the Bāsàa tradition;
- A lady who acquired a traditional grade reserved only to men, and who explains what happened;
- A journalist who writes chronicles on social harm;
- A musician who has been arrested and convicted, and who wants to prove that he is not guilty;
- An elder of a family presenting the importance of a family congress.

Table 1. Number of files of our corpus

speakers	Nb of files	Duration/s
Jmfab	36	55~70
Fem_mbo	54	30~35
Mandel	32	25~30
Alad	20	25~30
Nkeng	41	30~35
Mbock	25	30~35

Indeed, all the speakers felt personally involved in the subject. This is why the hypothesis that these speakers by virtue of their personal involvement in the subject, give free rein to their emotion and subjectivity, and thus produce melisms, could be put forward.

2.3. Structure of melisms in our corpus

The melisms obtained from our corpus presented a great variability depending on speakers. From an amount of 2964 melisms, we got 1888 lexical items (ML) and 1076 grammatical items (MG). The following table 2 and Figure 2 display the distribution of melisms in accordance to the speakers:

Table 2. Distribution of melisms per speaker

speakers	ML	MG
Jmfab	649	318
Fem_mbo	551	357
Mandel	209	70
Alad	84	79
Nkeng	251	174
Mbock	144	78

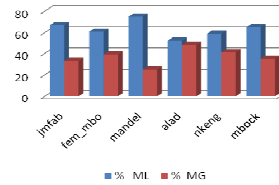


Figure 2: percentages of ML and MG per speaker

We consider that there are three features that conveniently characterise melisms: the tonal representation, the direction of the melodic curve and the type of melism structures.

- **Tonal correlates:** In the psychology and prosody domain, the positive valence phenomenon, which is the intrinsic attractiveness (positive valence) of an event, object, or situation (Frijda, 1986; Cowie and al, 2000b), is correlated to the F0 highest values.

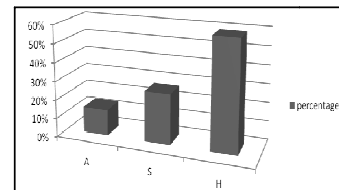


Figure 3: Tonal values of melisms

In those conditions, in agreement with this work, only three levels are considered as dealing with melisms: A, S and H. As for the French language, we can see from the graph (Figure 3) below that the H level is overrepresented as it is the most central target in the overall speakers' range, and thus, can be embedded on each ascending and descending slope. The A and S levels entail the highest F0 peaks, and the melisms

produced at that level grounded on the biggest effort are supposed to be highly informative.

- **The melodic curve direction:** It is very important to note here that most of the time (almost 80%) the curve is not flat such as a plateau. In most of the cases indeed, it shows downward or upward trends. This phenomenon tends to confirm the dynamic nature of melisms, because downward and upward movements imply also great increase or decrease of pitch energy.

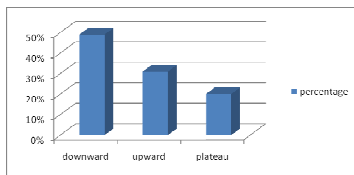


Figure 4: trends of the slope on melisms

- **Type of structures:** According to the results bearing on French prosody (Caelen-Haumont, 2007), 4 types of melisms can be found in our Bâsâa language corpus:
 - The M type which is a melism within a syntagm, without any phrase constraint, be it contextual, grammatical or phonological. This category represents mostly the full meaning of melism, because it is just a matter of subjectivity.
 - MC stands for a melism by contact, that is, a word that is close to a melism, i.e. just before/after it.
 - MP stands for melisms before pauses.
 - MF indicates melisms at the utterance final position, but not followed by a pause.

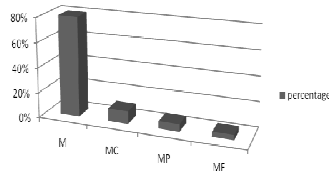


Figure 5: percentages of melism types

From the graph above (Figure 5), we can see the overrepresentation of the M category (80%). This fact means that the melisms being internal to the phrase without any contextual constraint are the true melodic expression of affectivity. It gives also evidence that our corpus is based on spontaneous speech.

3. Melisms in semantic networks

An example of a table is shown as **Erreur ! Source du renvoi introuvable.** Somewhat Caelen-Haumont (2002, 2004) and Morange (2005) demonstrated that melisms can be grouped together under an isotopy (Rastier 1989), which comprises lexical fields. The hierarchy of semantic analysis of melisms is presented below:

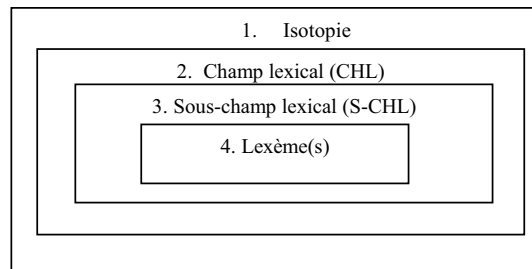


Figure 6: Semantic analysis level Hierarchy (Morange 2005)

Following this organization, the melisms (ML) in our corpus matched the discourse objectives of the speakers, presented in 2.2. This is the percentage of melisms in isotopies per speaker:

Table 3. Main isotopies of our corpus

Speaker	Isotopies	percentage
Alad	Judicial trials	74%
Fem_mbo	The traditional power	77%
Mandel	The present day society	67%
JMFAB	The traditional society	78%
Mbock	The traditional society	71%
Nkeng	The family congress :	79%

From the third column of this table, we can see that the large majority of ML (67% ~79%) belongs to the main isotopies which represent the discourse objectives. Every isotopy contains a certain number of lexical fields. Table 3 above presents the distribution of lexical fields in our corpus.

Table 4. Percentages of the speakers' lexical fields

Speaker	Lexical field	percentage
Jmfab	The family	37%
	Attitudes and behaviors	28%
	religion	13%
Fem_mbo	Traditional power	40%
	Politic or administrative power	25%
	Attitudes	12%
Alad	The young protégé	25%
	The judicial world	36%
	The speaker himself	13%
Mandel	Human relationships	17%
	The economical crisis	16%

	The political party	21%
	Avarice	13%
nkeng	The family congress	49%
	The family relationships	30%
	The traditional society	45%
Mbock	The Badjob group	26%

The above table 4 demonstrates the notion of semantic weight that makes certain lexical fields more important than others. This is to say that not all the lexical fields have the same value. The one with most occurrences may be more important than another with fewer occurrences. Likewise, all the lexical sub-fields do not have the same weight. In the following tables, we show the distribution of lexical fields into lexical sub-field, per speaker. One column is reserved to the main lexemes, those that are denotatively close to the main isotopy. The last column presents the number of occurrences those lexemes.

4. Pragmatic Analysis

It is at this level of analysis that we can see the speaker's expressiveness, because we analyze the intention hidden behind the semantic meaning of words both individually and intonationally grouped together in lexical fields. In her analysis of melisms, Caelen-Haumont (2002, 2007) demonstrated that a subjective dimension of messages can be discovered through certain local beliefs, specific to the speaker. These beliefs rely on two categories of values: circumstance values motivated by the actual discourse objectives, which comprise common cultural and social knowledge; and speaker's values which are the expression of her or his inner feelings, convictions, subjectivity and as a result the informative and communicative intentions of the speaker. These values are perlocutory aims, which can be understood as the actions that could derive from the speech act. In our corpus, as seen in table 4 above, most of the lexical fields focus on the Bāsàa culture (family, traditional power, religion). The way these lexical fields contribute to meaning is that the addressees should get

- For Jmfab, the notion of family constitutes the most important lexical field, especially how is designed the ideal Bāsàa family. As a lecturer, he gives his opinion about social matters like family relations, religion and the way to behave in society. He is a conservative and he wants people to behave like their ancestors did. Concerning family relations, he proposes that men should not feel equal to women, and that people should respect social hierarchy. Concerning religion, he wants people to adopt [ŋgámá] "divination", a traditional religion, instead of Christianity, recently brought by westerners.

- Fem_mbo is very attached to the position of [mbómbòk] *traditional ruler* she acquired and that is questioned. Her most important lexical field is traditional power. She wants to convince people that she deserves to carry that function, although it is reserved to men and that no woman carried it before. She wants to be respected because of her administrative and political favourable condition and by virtue of this, to be accepted as a traditional ruler.
- Mandel criticizes blight on society. He accuses the economical crisis to bring extreme poverty to people, and invites the political leaders to react. Concerning human relations, he denounces unfaithfulness and encourages friendship. He also criticizes a political party which has not produced satisfactory results in 15 years. In addition, he denounces avarice and invites people to share what they have.
- Alad wants to prove that he is not guilty of the charges against him. That is why his most important lexical field is the judicial world. He accuses the young man who walked with him, emphasizing that it is during his travelling that the young man caused the problem. Also, he wants people to believe that the judicial system is not honest because someone of his quality (a famous musician) would receive more consideration.
- Nkeng shows the importance for families to be together, like the congress of his own family. Family constitutes the most important lexical field of this speaker. But he criticizes the behavior of young people who lack consideration for elders. He also presents the importance of nephews, and he calls everybody to respect them.
- As for Mbock, traditional power is the most important lexical field. He wants people to respect the traditional hierarchy, particularly the position of [mbómbòk] *traditional ruler* that he is carrying. Also, he wants people to have more consideration for the traditional power that is innate, rather than for the power that has been acquired.

5. Conclusion

This paper is to our knowledge one of the first works grounded on Bāsàa data banks, and using computer tools to analyze the Bāsàa language and prosody. Grounded on our first results, this paper aims at showing how prosody could convey expressiveness in Bāsàa oral spontaneous discourse. Through melodic prominence on words termed as melism, it has been presented how prosody can contribute to the interpretation of messages, and more specifically, what are the underlying speaker's communicative and informative intentions. These intentions are comparable to perlocutory aims, that is to say the actions that the speaker induces (consciously or not) as awaits to be taken following his speech act. The following work will extend the distribution of melisms through the lexical fields and subfields.

In this work, we described the notion of melism as well as their kinds gathered from our corpus, their distribution and their main tonal targets (*A, S, H*). We then presented the distribution of the lexical melisms into lexical fields and isotopies, and finally, we introduced a preliminary discussion about the role of those melisms in marking the speakers' expressiveness. The results obtained here will eventually encourage some subsequent work with larger corpora.

6. References

- Caelen-Haumont, G. et Bel, B. « Le caractère spontané dans la parole et le chant improvisés : de la structure intonative aux mélismes », *Revue Parole*, 15/16, pp. 250-301, 2000.
- Caelen-Haumont, G. et Auran, C. « INTSMEL, un outil pour l'analyse des contours proéminents de F0 », *Bulletin PFC*, N°3, pp. 115-126, 2004.
- Caelen-Haumont, G. Labelling and structuring the F0 prominences using an automatic segmentation and annotation tool (MELISM), and statistical results, Proceedings of O-COCOSDA, Hanoi, Vietnam. 2007.
- Calliope (éd.) *La parole et son traitement automatique*, Masson, Paris, Milan, Barcelone, Mexico. 1989.
- Cowie R., « Describing the emotional states expressed in speech », *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, Proceedings on line, <http://www.qbc.ac.uk/en/isca/proceedings>. 2000.
- Cowie R., Douglas-Cowie E., Savvidou S., McMahon E., Sawey M., Schroder M., «Feeltrace” : an instrument for recording perceived emotion in real time », *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, Proceedings on line, <http://www.qbc.ac.uk/en/isca/proceedings>. 2000.
- Fuchs, C & al. *Linguistique et Traitements Automatiques des Langues*, Hachette Supérieur. 1993.
- Kerbrat-Orecchioni ,C. 1, éd. *L'énonciation. De la subjectivité dans le langage*, Colin U, Linguistique, 1999.
- Lacheret-Dujour, A., « Prosodie-discours : une interface à multiples facettes », *Nouveaux cahiers de linguistique française*, 28, pp. 7-40. 2007.
- MORANGE, S. *Approches structurale, prosodique, psychocognitive de quelques propriétés linguistiques subjectives de la parole : l'exemple de trois Montmartrois âgés*, Thèse de Doctorat, dir. Mary-Annick Morel, Univ. Paris 3 -Sorbonne Nouvelle. 2005.
- Nico H. Frijda, , , *The Emotions*. Cambridge (UK): Cambridge University Press, 1986. 207 p, 1986.
- RASTIER, F. *Sens et textualité*, Hachette, Paris. 1989.
- Simon, A. C. *La structuration prosodique du discours en français*, Berne, Peter Lang, 2004.
- Wichmann, A. « Prosody and discourse, a diachronic approach » in Auran et al. *Proceedings of Discourse-Prosody Interface symposium – IDP05*, Aix-en-Provence, Seridisc, Belgique, Cederom. 2006.