Voice diversity in conversation : a case study

Roxane Bertrand, Robert Espesser

Voice Diversity in Conversation: a Case Study

Roxane Bertrand & R. Espesser

Laboratoire Parole et Langage, CNRS, Aix-en-Provence
{Roxane.Bertrand ; Robert.Espesser}@lpl.univ-aix.fr

Abstract

This paper's aim is to present prosody as a polyphonic marker through the analysis of Direct Reported Speech excerpts pronounced by a female speaker in everyday conversation. Our results reveal the existence of global melodic differences between the three types of discourse analysed, eg Direct Speech, Direct Reported Speech which contains self-quotations and Direct Reported Speech containing other virtual enunciators than the source speaker.

1. Theoretical perspective

The voices to which we refer here echo the diversity of enunciative sources that a given speaker is often led to introduce in his own speech. This conception of enunciative source heterogeneity implies that the speaker is no longer conceived as the unique source of his discourse. We therefore admit that his productions are submitted to other influences as soon as the notion of self-conscious subject unicity has been challenged.

Authier Revuz (1982) makes a distinction between two types of heterogeneity. ″Constitutive heterogeneity″ finds its origins in works outside linguistics (Freud, Bakhtine, Mead) and concerns an internal reality depending on the subject. ″Shown heterogeneity″, linked to language activity performed by individuals, belongs more directly to the core of linguistics insofar as it is a clear reference to the plurality of voices that subjects invoke in their speech. In this study only the second type of heterogeneity will be considered. The term ″enunciator″ (Ducrot, 1984) will be used to mention the different voices staged by a speaker in his own speech.

The most vivid illustration of the presence of other voices in one's speech is the direct reported speech (DRS). De Gaulmyn (1992) defines the Reported Speech as the recorded broadcast of utterances previously pronounced by identified enunciators; a case that Vincent and Dubois (1996) name ″reproduction″. These authors propose other uses of the structure of reported speech which testify the diversity of facts it encompasses, among which we find : a/ ″pseudo-reproduction″, consisting in letting believe that the utterances have been pronounced already ; b/ ″realization″ which concerns a reported speech which is the prototype of several similar events; c/ ″invention″ which consists in making say utterances that have never been pronounced (Vincent & Dubois, 1996). These three uses cover the totality of cases examined here.

DRS, far from being a mere speech-reproduction strategy, constitutes a true speech production strategy. Quoting consists in adopting utterances emanating from someone else (to achieve argumentative goals for instance) to serve one's general speech project (Vincent & Dubois, 1996 ; Bertrand, 2001).

The speaker producing DRS can either quote other enunciators or himself / herself. Self-quoting supports the idea that it is rewarding to do so, since a simple direct speech might have been produced instead.

2. Objectives

Our objective in this study is to validate the role of prosody as a polyphonic marker through the analysis of F0 parameter. More precisely, we seek to characterize three discourse types :

-Direct Speech (DS)

-Reported Speech consisting of utterances emanating from another enunciator than the speaker (DRo)

-Reported Speech consisting of self quotations (DRsq).

3. Hypothesis

Voice changes (i.e. enunciator shift) correspond to vocal changes (i.e. shifts in pitch) (Grosjean, 1991). These voice changes are characterised by different involvement degrees from the speaker in his DS, DRsq and DRo.

We hypothesize that these involvement degrees are reflected in global variations of the F0 parameter (frequency distribution, temporal distribution, short term correlation).

4. Corpus and Methodology

We used excerpts from a one-hour familiar family conversation (Traverso, 1996) recording between three participants. 248 DRS were produced, 188 from the main female speaker (A) studied here.

The DRS were tagged manually on the speech signal. We retained pairs constituted of a DS immediately followed by a DRS. The duration of each pair varies between 1 and 4 seconds (the mean duration is about 2 seconds) for DS and DRS respectively. The duration of each element was equalized (the longer element is clipped to the shorter duration).

The pairs were also selected on a homogeneity criterion in terms of voiced sequences. The speech signal has been segmented automatically in voiced areas delimited by their own voiced boundaries or limited to a maximal duration of 140 ms. This value, quite close to the mean value of unstressed syllables in spontaneous speech (Astésano, 1999), provided a good adjustment between the number of these voiced areas and that of real syllables in the corpus, that is close to a syllabical sequencing. We call such voiced sequences ″pseudo-syllables″ (PS).

The selected pairs consisted of a 400 ms duration of voiced speach at least and 5 PSs to avoid too big a proportion of silent pauses. Morever, overlapping pairs also were

eliminated. These various constraints account for the fact that 50% of the initial DRS has been rejected.

91 pairs of DS versus DRS (among which we found 53 DRo and 38 DRsq) were part of the selection.

In this study, we considered only F0 parameter. F0 extraction combined three detection methods : comb function, AMDF and autocorrelation. F0 was computed every 10 ms.

Two types of measures were derived from the F0 files. The first type was the mean of the raw values of F0 (in Hz) of each PS. The second type is the value (in Hz) of a target-point (TP) obtained from a modelization routine. The F0 raw curve was modelized by a continuous and smooth curve deprived of microprosodical variations, which are not relevant for the analysis (Di Cristo & Hirst, 1986). According to Di Cristo (2000), this modelization process has two main advantages: it substitutes to the raw curve a smooth curve corresponding to the continuous perception of speech melody and it ensures the emergence of the target points indicating the significant anchoring points of this curve.

5. Statistical analysis

5.1 F0 values distributions

5.1.1 Descriptive statistics

Descriptive statistics were derived from PS and TP measures. These were mean F0, standard deviation (SD) of F0, lower and upper quartiles, interquartile range (IRQ), lower and upper 10th percentiles, inter 10th percentile range (IDR). Percentile measures were computed too because they were less tied to assumptions about distribution.

Table 1a: F0 mean and standard deviation of PS and TP.

	Item count	mean	SD
PS			
DS	896	248.69	66.08
DRo	605	249.41	66.31
DRsq	383	267.92	83
TP			
DS	515	264.7	82.5
DRo	348	261.5	79.3
DRsq	220	285	99

Table 1b: F0 quartiles and 10th percentiles of PS and TP.

	q <	q >	IQR	d <	d >	IDR	median
PS							
DS	205	274	69	186	340	154	231
DRo	204	280	76	188	344	155	232
DRsq	216	303	87	190	385	195	244
TP							
DS	207	304	97	186	385	199	240
DRo	205	297	92	183	366	183	235
DRsq	214	329	115	187	429	242	256

5.1.2 Comparison of the F0 values distributions

To compare these F0 values distributions which differ from the normal distribution, we used a Kolmogorov-Smirnov test.

Pseudo-syllabes

DS vs Dro: D = 0.0427; p = 0.5268
DS vs Drsq: D = 0.1155; p = 0.001554
Dro vs Drsq: D = 0.1357; p = 0.000355

Target-Points

DS vs Dro: D = 0.045 ; p = 0.78
DS vs Drsq: D = 0.112 ; p = 0.041
Dro vs Drsq: D = 0.12 ; p = 0.03

DRsq is quite different from both DS and DRo. DS and DRo do not distinguish from one another.

The differences were less obvious for TPs than for PSs because there were less TPs.

5.1.3 *Proportion test*

In a previous study (Bertrand & Espesser, 1998), we found a greater proportion of F0 values greater than 270 Hz in the DRS. This threshold was used here to compare the three distributions.

Table 2 : proportion of the PSs > 270 Hz

	PSs proportion	Total count PSs	count > 270 Hz
DS	0.26	896	237
DRo	0.28	605	172
DRsq	0.36	383	140

Proportion test

DS vs DRo p = 0.43 (ns)
DS vs DRsq p = 3.610^{-9}
DRsq vs DRo p = 0.009

Table 3 : proportion of the TP > 270 Hz

	TPs Proportion	Total count TPs	count > 270 Hz
DS	0.35	515	181
DRo	0.347	348	121
DRsq	0.445	220	98

Proportion test

DS vs DRo p = 0.967 (ns)
DS vs DRsq p = 0.02
DRsq vs DRo p = 0.0248

NB : the proportion of the F0 values greater than 270 Hz is more important for TPs than for PSs: a PSs segmentation does not eliminate intermediate values while the TPs modelization retains extreme values preferentially and backgrounds intermediates values.

These results show that DRsq stands out from DS and DRo which are indistinguishable. DRsq is characterized by an extension of the PSs and TPs distributions beyond 270 Hz.

5.2 Global variation of F0 values

5.2.1 Temporal density of the TPs

TPs being located essentially at the points where the curve changes, temporal density (number of TPs per seconds) is a cue of the variability of the F0 curve. Below are compared (t-test) the temporal density of the three populations.

DS vs DRo : 3.3 vs 3.6
t = 1.9 df = 115.7 p = 0.054
DS vs DRsq : 3.62 vs 3.31

t = -1.77 df = 70.1 p = 0.08
DRsq vs DRo : 3.62 vs 3.6
t = -0.09 df = 76.6 p = 0.92

There are no significant differences between the three discourse types. The mean temporal difference between two TPs varies from 275 to 300 ms.

5.2.2 Slope between two consecutive TPs

The slope between two TPs (in Hz/ms) is another cue of the variability of the F0 curve. As the mean slope is equal to 0 by construction, we compared the SD of the three populations.

Table 4 : SD of the slope of DRsq, DS and DRo

	DRsq	DS	DRo
SD	0.558	0.496	0.42

DS vs DRsq: F = 0.7904 p = 0.035
DS vs DRo: F = 1.378 p = 0.001
DRsq vs DRo : F = 1.7441 p = 3.844 e-06

The three SD are significantly different with a clearcut distinction between DRsq and DRo.

To illustrate these results, we calculated the frequency value of a TP located at 300 ms from a first 250 Hz TP (mean value for the speaker in this corpus) with a slope measure of one SD (typical value of the distribution).

Table 5 : semitone ratio of TP (i+1) to estimated TP(i)

	DRsq	DS	DRo
TP(i+1)/TP(i) semitone	8.86	8.05	7.07

5.2.3 Temporal evolution of the F0 values

As a cue of temporal continuity, we chose the correlation coefficient (R) between two consecutive PSs F0 values and two consecutive TPs F0 values too.

Table 6 : correlation coeff. (R) of two consecutives F0 values (PS and TP)

R	PS(i) PS(i+1)	TP(i) TP (i+1)
DS	0.449	0.048
Dro	0.637	0.25
Drsq	0.49	0.055

The table 7 shows the significance of the difference between the correlation coefficients.

Table 7: Comparison of Rs

	PS(i)PS(i+1)	TP(i)TP(i+1)
DS vs Dro	z: 5.11 p : 3.16 10^{-7}	z : 2.997 p : 0.0029
DS vs Drsq	z: 0.859 p : 0.39	z : 0.086 p : 0.93
Dro vs Drsq	z: 3.31 p : 9.10^{-4}	z : 2.3 p : 0.0207

DRo stands out quite significantly from the other two, both by a lower variability of slopes and a higher correlation coefficient between 2 consecutive PSs (or TPs). DS and Drsq do not really differ related to these parameters.

NB : Correlations are weaker for TPs as they are more temporally scattered than PSs.

6. Discussion

Our results contribute to validate the role of prosody as a polyphonic marker. It is clear that the F0 parameter is relevant to differenciate the three types of discourse (DS, DRsq, Dro) that have been considered.

In our previous results two melodic effects enabled us to distinguish between DS and DRS, these effects currently belong to the Drsq and Dro types.

DRsq differs from the two other types on the F0 values distribution of the PSs and TPs. Drsq is characterised by an extension of our female speaker's pitch range as standard deviation and interpercentile ranges show (table 1a, 1b). Upper percentiles increase more than others, which confirms the significant increase of the proportion of values greater than 270 Hz (table 2, 3). Morever, the slope measures between two consecutive TPs show a more important variability for DRsq which has more abrupt slopes. The combination of these two acoustic elements contributes to make DRsq more salient in the speaker's speech. Table 5 shows that a TP estimated from a typical slope of one SD is 1.8 semitone higher for DRsq than DRo, which is perceptible.

Such results confirm the position of Vincent and Dubois, who wondering about the purpose of using reported speech with self quotations, come to the conclusion that the content of these quotations is thus highlighted. The use of DRsq implies that what is uttered is relevant for our female speaker who thus draws the attention of other interactants to some speech segments by creating a complex system.

Dro is produced with smaller slopes in a narrower pitch range than it is the case for Drsq, hence a lesser variability. Dro has a higher correlation coefficient between 2 consecutive PSs or 2 consecutive TPs, which indicates a better degree of predictability between them. We can interpret this point as reflecting a linear dependance distributed among all PSs / TPs or, more interestingly, a dependance concentrated on sequences composed of 3 or 4 consecutive PSs, which realizes a micro-serie or a reccurrent motive composed of rather smooth and continuous moves. The presence of lengthening phenomena can create this effect, and we could thus oppose the 'melodical salience' (Caélen-Haumont, 1991) of DRsq to a 'temporal salience' of DRo. On listening to the corpus, we have been able to locate the reccurrent presence of a particular ″drawling″ inflexion in DRo.

We already have established (Bertrand, 2001) that DRS are relevant utterances for the speaker under study insofar as they are used to generate in her dialogue a specific self-image. It is then of prime interest to make them salient to co-participants. She makes use of two strategies : a/ she presents herself in a high position at discourse level (she posseses the information) and subjective level (a fighting lively and dynamic person), b/ she often introduces enunciators (DRo) in low positions (discourse position of asker, or subjective position of coward). The discourse-enunciative level, in spontaneous speech notably, is not easily separated from a level belonging to the affective. Under the general heading of focalisation, Di Cristo (1999) unites aspects of discourse salience which have been dealt with separately. According to Bolinger's view, any transformation transfer is mediatised by the affective in any case, by the interest that a given speaker

shows for their own message, and by the desire to impress their audience.

Our results confirm that different involvement levels can emerge by the staging of various focalisation systems (emphasis here), specific to DRsq and DRo which are ″cristallized″ (De Gaulmyn). The term 'focalisation' echoes Di Cristo's conception, who believes it plays a major part to identify and make salient speakers' involvement strategies in their speech.

DS with a small pitch range (table 1a and 1b) and a weak correlation coefficient (table 6) can be opposed to DRsq and DRo. We argue that DS are less marked because the speaker's involvement is not important as she seems more interested in focalizing DRS.

More generally, these results are in conformity with the notion of « enunciative instability » (Vion, 1998) characterised by a fluctuating discourse marked by sucessive strong and weak moments dependent on the degree of a speaker's involvement in his speech. This instability or ″enunciative breathing″ would correspond to an ″enunciative rhythmization process″ based on the reccurrence of melodic patterns in some relevant discourse elements (DRS). The DS, around which the two DRS types are to be found, could be considered as planning stages that our female speaker neutralizes, knowing she is about to produce DRS so as to make the latter utterly salient.

7. Références

[1] Authier-Revuz, J., 1982, Hétérogénéité montrée et hétérogénéité constitutive, éléments pour une approche de l'autre dans le discours, *DRLAV*, 26, 91-115.

[2] Ducrot, O., 1984, Esquisse d'une théorie polyphonique de l'énonciation, in *Le Dire et Le Dit*, Editions de Minuit, 165-191.

[3] De Gaulmyn, M.M., 1992, Grammaire du français parlé. Quelques remarques autour du discours rapporté, in Actes du Congrès de l'ANEFLE *Grammaire et français langue étrangère* Joussaud & Petrissans (dir.), Grenoble, ANEFLE, 22-23.

[4] Vincent, D. ; Dubois, S., 1996, *Le discours rapporté au quotidien*, Nuit Blanche Editeur.

[5] Grosjean, M., 1991, Les Musiques de l'Interaction, *Thèse de Doctorat de Psychologie*, Université Lumière, Lyon II.

[6] Caélen-Haumont, G., 1991, Stratégie des locuteurs en réponse à des consignes de lecture d'un texte : Analyse des interactions entre modèles syntaxiques, *Thèse de Doctorat d'Etat de Lettres*, Vol. 1, Université de Provence.

[7] Bertrand, R., 2001, Etre soi avec les mots d'autrui , *Faits de langues*, 23, (à paraître).

[8] Traverso, V., 1996, *La conversation familière. Analyse pragmatique des interactions*, Lyon, PUL.

[9] Bertrand, R. ; Espesser, R., 1998, Prosodie et discours rapporté : la mise en scène des voix , in Pragmatics in 1998 : *Selected papers from the 6th International Conference*, vol. 2, Verschueren, Jef (ed) Anvers, International Pragmatics Association, 45-56.

[10] Astésano, C., 1999, Rythme et Discours : Invariance et sources de variabilité des phénomènes accentuels en français, *Thèse de Doctorat de Phonétique*, Aix-en-Provence.

[11] Hirst, D.J., Di Cristo, A. ; Espesser, R., 2000, Levels of representation and levels of analysis for the description of intonation systems, in *Prosody : Theory and Experiment*, Merle Horne (ed), Kluwer Academic Publishers.

[12] Di Cristo, A. ; Hirst, D.J., 1986, Modelling French Micromelody : Analysis and Synthesis, *Phonetica*, 43, 11-30.

[13] Di Cristo, A., 2000, La problématique de la prosodie dans l'étude de la parole dite spontanée, *PArole*, 15-16, 189-249.

[14] Di Cristo, A., 1999, Le cadre accentuel du français contemporain : essai de modélisation, *Langues*, 2 (4), 258-267.

[15] Vion, R., 1998, De l'instabilité des positionnements énonciatifs dans le discours, in Pragmatics in 1998 : *Selected papers from the 6th International Conference*, vol. 2, Verschueren, Jef (ed) Anvers, International Pragmatics Association, 577-589.