Chapter 23. "Voices" and bodies

Claire Maury-Rouan | LPL, CNRS – Université de Provence, France

John Benjamins Publishing Company
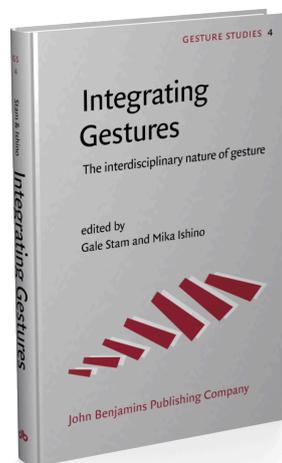
"Voices" and bodies

Investigating *nonverbal parameters of the participation framework*

Claire Maury-Rouan

LPL, CNRS – Université de Provence

According to interactional and dialogic linguistics, utterances may be seen as complex constructions in which alternate "voices", more or less identifiable in reported speech, or less transparent as in polyphony, may be heard. As vocal changes had been established in reported speech, it was hypothesized that shifts in facial expressions, gesture or posture paralleling shifts of footing might also be found. The analysis of videotaped data showed that two distinct formats: reported speech and polyphony in perspective shifts co-occurred with relevant nonverbal cues. Based on the degree of variation of accompanying vocal and visual parameters, three relevant types ('underplayed', 'animated' and 'lively') were found in reported speech. Perspective shifts were found to start with a pause, a shift in the speaker's posture (head tilt) and pitch range variation. Two distinct cases of perspective shifts were found, whether the speaker became (–gaze) or remained (+gaze), possibly indicating quite different mental forms of perspective shifts.

1. Polyphonic "Voices" inside speech

From an interactional point of view, discourse has often been described as the result of a joint construction of utterances. When speaking, we plan and adapt the form and content of our discourse in accordance with an internalized image of our addressees, remembering previous interactions and anticipating the decoding progression and possible reactions of our interlocutors.

Such a conception of speaking has been described from several angles. It is typically present in Bakhtinian linguistics (Bakhtine 1929, Bakhtine 1979, Voloshinov 1968) and in the various more or less Bakhtinian enunciative theories of *polyphony* developed by Ducrot (1984), Authier-Revuz (1982), Vion (1988) or Nölke (1993). In addition, Goffman's (1981) conception of *footing* is based on a quite similar viewpoint. These various theories describe a given speaker as not being the unique source of his

discourse. In Ducrot's model, for instance, a given speaker may stage several distinct voices (called "énonciateurs" by Ducrot) within his speech. In a very similar way, Goffman distinguished between three "figures" as being alternately responsible for the words being heard within the speaker's utterance: the "animator" (as the individual voicing the words – the "author" (as the identifiable source of the words, in the case of reported speech, for instance) and the "principal" (as the party who is socially responsible for what is said e.g. *science, religion*).

In a number of cases, production formats rely on clear structure marks, as in reported speech, quotations or speech repairs. But in other formats, the identification of the sources becomes a subtle and highly complex task, for instance, when some kind of speech sounds as though it is reported speech, but without any clear cues to an identifiable "author". Similarly, distinct voices may be heard when several stances are staged, or simply alluded to by a given speaker who animates a fictitious dialogue within his own speech. Such allusions sometimes amount to tenuous traces like the presence of the French particle *quand même* ('yet'). The many degrees in the ways this variety of sources may be displayed, hidden or implied in speech have been been described in Bakhtinian terms as "heterogeneity" under two forms: "shown" or "constitutive" (Authier-Revuz 1982; Vion 1998; Maury-Rouan, Vion & Bertrand 2007). In the same way, Goffman's production format allows a considerable expansion of the various figures animated in the speaker's utterances, thus presenting a highly complex *laminated* speaker, as Goodwin and Goodwin (2004) pointed out.

From a more interactional viewpoint, polyphony may have several advantageous effects. As they invite a variety of voices into their speech (or, in Goffman's words, as they produce constant changes in footing), speakers display a reflexive stance towards their utterances; this may give them the image of being broad-minded and considerate interactants. It also produces softening face work effects as it changes a harsh statement into a softened, open-minded proposition.

In their 2004 paper, Charles and Marjorie Goodwin commented on Goffman's Participation Framework as being exclusively speaker-oriented. In Goffman's model, they pointed out, "speakers and hearers inhabit different worlds," speakers being endowed with "rich cognitive and linguistic capacities" while other participants "are left cognitively and linguistically simple" (being solely classified as "ratified-non ratified addressees/over-hearers/bystanders.."), as the process of mutual monitoring is usually neglected in Goffman's study of the participation process, whereas "the process of designing an utterance for a particular type of addressee can shape its structure in powerful ways (...)" as "speakers reconstruct their sentences in their course to make them appropriate to different kinds of recipients" (p. 224).

Furthermore, Goodwin and Goodwin criticized the fact that Goffman's production format usually highlights the sole stream of speech – "over other forms of embodied practice that might also be constitutive of participation in talk". The same blindness could be found in most of the work currently published in the linguistic field of polyphony research and of Bakhtinian studies on interaction and speech, with the

exceptions of a few authors like Vincent and Perrin, according to whom "reported speech is woven into a series of non-verbal events" (1999: 291).

2.   "Voices" and bodies

This study investigates changes that might occur in speakers' faces, gaze, postures and gestures when such "voices" are being staged. At its present phase, it does not escape Goodwin and Goodwin's reproach of being exclusively focused on speakers, and solely attempts to offer an insight into the "other forms of embodied practices" (Goodwin & Goodwin 2004: 25) that might parallel participation in talk.

A number of studies have investigated the part played by prosody in various cases of reported speech, one of the most frequent (and most clearly identifiable) shifts in footing. In reported speech, Elisabeth Couper-Kuhlen (1998) explained, the unity within a single speaker of the three figures (author, animator and principal) dissolves, and they become independent. Couper-Kuhlen found that a set of phonatory cues helped speakers in identifying such shifts, as she demonstrated that vocal changes are often crucial for coherence in helping interlocutors to detect changes in footing because verbal prefaces (e.g.: "*he said...*") are currently omitted in spontaneous speech, and appropriate shifts in personal deixis (mainly personal pronouns) may be lacking, or ambiguous. The main indication of the staging of an alternate figure inside the animator's speech is then transferred to vocal deixis. This may include notable shifts in four levels of speech characteristics: loudness, pitch, tempo, and voice quality. In addition to identifying deictic functions, Couper-Kuhlen noted that such phonatory voice variations often convey different stances in the way the summoned figures are presented in the animator's discourse, a property confirmed by the reactions of interactants. Consistently, Bertrand (2002) found a correlation between different degrees of the speaker's involvement and overall variations of the F0 parameter.

The presence of a strong link between facial expressions and vocal qualities, and between facial, head or even slight hand movements and changes in vocal pitch and stress has been established through a number of studies, from Fonagy (1983), Bolinger (1985), Ekman (1979) and Ohala (1980). This leads us to hypothesize that shifts might be found in facial expression and possibly in other movements, gaze, gesture or posture, occurring in the same way as the vocal changes that parallel shifts of footing in reported speech. This insight has been reinforced by data from Goodwin (1984) and McClave (1998).

Goodwin (1984: 228–229) described together with changes in voice and intonation, several displays in a narrator's body position (involving clasped or separate hands, elbow position, head movements, gaze, leaning forward) as significant indications "about the nature and extent of her orientation to the conversation itself, as well as displays about the type of talk she is producing, and relevant differences within that talk" including parts of a narrative and reported speech. Relevant action patterns corresponded to the speaker's and the recipients' orientation and to the sequential organization of talk.

In a comprehensive study of head movements, McClave (1998) noted that speakers change their head positions to indicate different discourse levels.

> "Speakers change the orientation of their heads when switching from indirect to direct discourse. This is not simply a realignment or a reorientation toward the interlocutor, as (...) speakers at times break mutual gaze and turn their heads away from the recipient at the beginning of a quote". ". . . .The speaker seems to assume momentarily a new head position to give voice to another's words. Thus a different speaker is conceptualized as occupying a different space. The spatial shift signals the speaker's change in footing from narrator to that of a character in the narration and makes visible the change in narrative structure." (McClave 1998: 369).

3. Corpus and methods

In the present study, two corpuses were investigated for non-verbal correlates of shifts of footing. Corpus n°1 is an 80 minute video recording of an 82 year old lady, Ariane, interviewed about her life story, in the familiar setting of her sitting-room, in the presence of a couple of friends (alternate addressees during the interview) and a cameraman.

The second corpus, filmed in a research lab, recorded two young men (volunteers in the experiment) engaged in casual talk. They had been asked to talk about memories of unusual events in their lives.

To date, 36 out of the 80 minutes of the 'Ariane' Corpus (n°1) have been thoroughly analyzed, and 10 out of the 60 minutes of the 'Young Men' Corpus have been partially analyzed.

With a view to describe the parameters of nonverbal shifts, the verbal data from both corpuses were first analyzed following discourse analysis criteria, with special attention to three polyphonic discourse patterns: reported speech, speech repairs and verbally explicit perspective shift sequences. Relevant voice and gesture parameters were investigated in a later phase.

The nonverbal parameters of perspective shifts and of speech repairs were investigated in the total corpus (46 minutes). The nonverbal accompaniment of reported speech sequences was investigated in Corpus n°1 'Ariane' (36') only.

4. Findings

In two out of the three discourse patterns, reported speech and perspective shift sequences were found to display relevant nonverbal cues. Up to this point, speech repairs do not appear to have been marked by any relevant nonverbal aspects.

4.1   Reported speech

4.1.1   *Prosodic displays*

The prosodic cues of reported speech as described either by Couper-Kuhlen (1998) or Bertrand (2002) were found to be present in the corpus, although unevenly distributed. Three types of distinct reported speech animation seemed to stand out, corresponding to regular patterns in the corpus (even though some of the cases might have been better described in terms of degrees in a scale of prosodic marking).

Type 1: 'Underplayed' reported speech.
In a number of cases, where the presence of reported speech was clearly displayed through verbal *prefaces* ("he told me, she used to say..."), vocal parameters were found to be very tenuous, reduced to subtle tempo variations. No difference as to pitch range, volume, or voice quality was perceptible in the transitions between direct speech preceding the sequence of reported speech and the reported speech sequence itself: the only prosodic cue amounts to a tenuous shift in scansion marking the reported utterance.

Example 1

Example n° 1
    Type 1: example (from Ariane: 2:30):
    <preface>/reported utterance

_____

... un auteur, dont le nom me reviendra,   *<qui racontait> "je suis une enfant trouvée"*
... a writer, whose name I should remind, <who told   > "I am a foundling"

Type 2: 'Moderate' reported speech
In the second group, reported speech was more vividly animated: scansion was more distinctly present, and shifts in volume were perceived, but no changes were heard as to pitch or voice quality.

Type 3: 'Theatrical' reported speech.
In the last type of reported speech, the speaker gave life to "author" figures in a way close to a theatrical performance, not only displaying shifts in tempo and in volume but also adopting vocal quality and pitch changes as if actually impersonating the absent speaker's voice. This third type of reported speech type was found to be the most frequent in the corpus (as shown in Table1).

Example 2
Type n°3: example (from Ariane : 30:01):
(the narrator is telling how a spinster relative refused to be married to a distant cousin):

reported utterance /	<preface	> main speech/reported utterance

"non ! non !"	*< elle lui dit*	*> je la revois/"non ! non ! donnez-le à*
"no ! no !	"<she told him>	I can still picture her/" no ! no ! give him to
Marie-Thérèse!"		
Marie-Thérèse"		

4.1.2 *Visual parameters*

Type 1: In 'underplayed' reported speech, where scansion was the only varying prosodic parameter, the speaker did not display any change in facial expression, trunk or head position; no gesture was produced, and the gaze maintained the same direction as during the previous direct speech sequence. *Underplayed* reported speech thus appeared to be semiotically embedded within the direct speech speech sequence. The facial expression, its qualifying and/or commentary value, and the general body display also matched the main utterance. Were it not for the slight shift noted in scansion, *underplayed* direct reported speech would be in all respects similar to the way syntactically embedded indirect reported speech would be voiced and nonverbally performed. (In such cases, the use of direct reported vs. indirect reported speech might simply serve highlighting functions).

Type 2: In *moderate* reported speech, displaying shifts in volume in addition to distinct scansion, several shifts in visual nonverbal behavior were found: gestures matching the verbal meaning of the utterance were produced, giving more embodied life to the animated figure; consistently with McClave's 1998 findings, shifts in posture – typically head movements – often marked the beginning of the reported speech sequence. Gaze direction conformed to the behavior usually found in speakers (alternating look-away moments and control glances).

Type 3: In type 3, *theatrical* reported speech, together with voice quality shifts, affecting pitch range and vocal quality combined with the other variations in vocal parameters, facial expressions matching the impersonated voice pitch and quality appeared, in addition to posture shifts, head movement and gestures already found in type 2. As in the previous types, gaze direction alternated look-away moments and control glances.

4.2   Perspective shifts

Perspective shift was the other type of discourse pattern to appear with a distinct nonverbal accompaniment. In discourse, perspective shifts are observed when a speaker interrupts the development of a point to take into account an alternate viewpoint. This may be operated in a carefully organized rhetoric way, using adequate syntactic and pragmatic apparatu, as in written argumentation. In spontaneous oral speech, although instances of planned and pragmatically controlled perspective shifts may also be found, a number of them seem unprepared and simply seem to occur: the speaker,

Table 1. Vocal and visual parameters in the various types of reported speech

	Prosodic parameters: variation in:	Visual parameters variation in:	Total number of cases:	%
Type 1: *underplayed*	scansion	–	19	28%
Type 2: *moderate*	scansion, volume	gestures head movements shifts in posture	15	22%
Type 3 *theatrical*	scansion, volume, pitch, vocal quality	gestures head movements shifts in posture facial expressions	33	48%
			68	100%

while uttering a point, as if struck by an alternate viewpoint, stops and starts again wording the different opinion or standpoint. Thus, perspective shifts share some of the aspects of speech repairs, such as dysfluency, pause and restart on a new version, and both behaviors imply a reflexive attitude. However, a distinction seems to be justified, defining speech repairs as a search for better lexical adequacy in wording a consistent point of view as opposed to the contemplating of a distinct opinion or standpoint found in perspective shifts. In terms of polyphony, a distinct voice is heard within the speaker's discourse with the irruption of this new perspective, whether this voice can be identified as that of a former interlocutor, a virtual opponent or an imaginary or collective party in relation to whom the speaker's utterances are produced.

In this type of framework, the animated figure is not always clearly referred to deictically as opposed to most of the cases in which direct reported speech is used; the figure's identity may simply be sketched or inferable from the verbal organization of the utterance revealing the presence of alternating voices.

4.2.1 *Verbal and vocal marks of perspective shifts*
Several cases of such perspective shifts were found in both corpuses, verbally and vocally displayed in a similar way. Verbal displays currently include relevant introductive discourse particles such as the French *mais* ('but, however'), *enfin, bon* ('well', 'actually'), followed by the verbal utterance of the alternate point of view. In other cases, the indication of perspective shifts may be limited to tenuous allusions to alternate stances as in the use of some French adverbs: *finalement'* ('everything considered'*)* or *quand même* ('nevertheless').

On the prosodic level, perspective shifts are frequently marked by variation in pitch range (to a higher pitch range level in our corpuses).

4.2.2 *Nonverbal components of perspective shifts*
Although sharing a set of consistent visual displays, the perspective shifts in the corpus need to be placed in two distinct categories according to the way speakers embody the

utterance conveying the reflexive speech act (Table 2). In the first category, the speech stream suddenly stops; after a short pause, the speaker displays a shift in posture, typically consisting of a head movement, as the head is tilted backwards and to the side, and the face is oriented to a different direction the reflexive re-oriented part of the utterance begins. In this first case, the speaker is –gaze during this shift and remains so at least until the verbal utterance is almost completed. Examples 3, 4, and 5 illustrate this category.

Example 3
perspective shift (–gaze): example from Young Men (1:52 to 2:02)
(on *enfin bon* ('actually', 'well') it suddenly has occurred to the speaker that his mother might have been right after all):

_____

non par contre +++ ma mère elle faisait des +++ enfin bon + hé + je comprends hein + mais
                            (makes faces)   (head tilts backwards)
no my mother did some ++++ actually well I understand ok + but

_____

bah euh + j'étais pas propre longtemps hein
um + it took me a long time to get toilet-trained

In this example, the tilt of the head coincides with the utterance of the adverb *quand même* ('nevertheless') which apparently represents (and condensates) a previous dialogue with the speaker's mother's on her child raising methods.

Example 4
perspective shift (–gaze): example from Ariane (6:05 to 6:12)
(the speaker suddenly stops grieving on the hard way she was raised as a child, as it suddenly occurs to her that the Victorian education principles left her mother no choice : *mais elle pouvait pas* – 'but she could'nt'):

_____

Ah c'était + au moins on savait où on allait + mais elle pouvait pas faire autrement je pense

_____

Ah it was + at least we knew where we stood + but she couldn't act differently I guess
                                  (head tilts backwards)

      _____

   +++ *elle pouvait pas*
   +++ she just couldn't

Example 5
perspective shift (–gaze): example from Young Men (2:22 to 2:30)
_____

apparemment c'était radical quoi + mais putain c'était + quand j'y pense je me dis putain
apparently it seemed radically efficient but it was fuckin' + thinking of it I realize

_____   _____

c'était quand même vachement violent quoi
it was <u>nevertheless</u> fuckin' tough though
     (head tilts backwards)

In the second category, the interruption of the speech stream is less noticeable; a shift in posture including the tilting backwards of the head is present, but often less distinctly displayed; however, the most striking difference is the presence of a constant gaze towards the interlocutor's face.

Example 6
perspective shift (+gaze): example from Ariane (18:54 to 18:58)
_____

C'était + c'était + c'est vrai que ça a existé ça

It was + it was + it's true that such things actually did exist
          (head tilts backwards slightly)

Table 2. Distinct perspective taking displays

	Discourse strategy	Actually experienced
Prosodic parameters	Speech is interrupted (briefly) Pitch range shift (frequent)	Speech is interrupted (briefly) Pitch range shift (frequent) (more distinctly than in Discourse strategy)
Postural parameters	Posture Shift Head tilted backwards and to the side Face is re-oriented	Posture Shift Head tilted backwards and to the side Face is re-oriented (more distinctly than in Discourse strategy)
Gaze	+gaze	–gaze

5. Discussion and conclusion

The behavior schemes found in the two distinct types of perspective shifts display head movements very similar to the ones described by McClave (1998) and that we also found in reported speech sequences. Thus in perspective shifts, one might consider, as in reported speech, that "the speaker seems to assume momentarily a new head position to give voice to another's words"(McClave 1998: 369), even though in such cases the

different speaker might be described as a mere voice, a stance alluded to by the main speaker's utterance. Nevertheless, such a different speaker is conceptualized by the speaker as occupying a different space so that a spatial shift signals the speaker's change in footing to that of another figure, and makes the change in perspective visible.

According to studies on the relation of gaze direction to speech (Kendon 1967, Lee & Beattie 1998, Goodwin 1981) such a withdrawal of gaze is considered to coincide cognitively with planning phases as opposed to phases of feed-back control. Therefore, it is likely that a speaker, when engaged in thinking-while-speaking on being "struck" by a new idea or by the relevance of an alternate standpoint, will display such a –gaze attitude.

On the other hand, the presence of similar verbal, prosodic and nonverbal parameters, but with a speaker being +gaze, might amount to something more *prepared* and the utterance of the words expressing a reflexive standpoint should in this case be seen as more rhetorical, as a kind of discourse strategy: the staging of an alternate voice is thus used as one of the many ways of avoiding too much assertiveness.

In a –gaze speaker, these same cues should be seen as marking the genuine irruption of another figure's point of view, or as the presence of a voice being heard within the inner debate of a laminated subject.

A parallel might be established between such a partition between rhetorical and experienced heterogeneity in discourse and the functions of facial movements and expressions in human communication. According to Ekman and Friesen (1969) and Ekman (1984), speakers seem to be capable of making a secondary use, for semiotic functions, of the same behaviors that appear spontaneously when subjects experience a given emotion or affect, in the case of smiles or of eyebrow raising for instance. In a similar perspective, rhetorical perspective-taking displays might turn out to be semiotic behaviors based on the mimicry of spontaneous reactions.

References

Authier-Revuz, J. 1982. "Hétérogénéité montrée et hétérogénéité constitutive, élements pour une approche de l'autre dans le discours." *DRLAV* 26: 91–115.

Bakhtine, M. 1929/1963/1970. *La Poétique de Dostoïevsky.* Paris: Seuil.

Bakhtine, M. 1979/1984. *Esthétique de la Création Verbale*. Paris: Gallimard.

Bertrand, R. 2003. "Etre soi avec les mots d'autrui." In *Le Sujet*, J.-M. Merle (ed), 269–280. Paris: Ophrys.

Bertrand, R. and Espesser, R. 2002. "Voice diversity in conversation: a case study." Speech Prosody 1. (11–13 avril 2002). In *Proceedings of Speech Prosody 2002,* B. Bel and I. Marlien (eds), 171–174. Aix en Provence: Laboratoire Parole et Language, .

Bolinger, D. I. 1985. *Intonation and Its Parts*. London: Edward Arnold.

Clift, R. and Holt, E 2006. "Introduction." In *Reporting Talk: Reported Speech in Interaction*, R. Clift and E. Holt (eds), 1–15. Cambridge: Cambridge University Press.

Couper-Kuhlen, E. 1998. "Coherent voicing. On prosody in conversational reported speech." In *List N°1 Interaction and Linguistic Structures* (Konstanzer Online Publikations System: KOPS Datenbank).

Ducrot, O. 1984. "Esquisse d'une théorie polyphonique de l'énonciation." In *Le Dire et le Dit,* 171–233. Paris: Minuit.

Ekman, P. 1979. "About brows. Emotional and conversational signs." In *Human Ethology*, M. von Cranach, K. Foppa, W. Lepenies and D. Ploog (eds), 169–248. Cambridge, Cambridge University Press.

Ekman, P. 1984. "Expression and the nature of emotion." In *Approaches to Emotion*, K. Scherer and P. Ekman (eds), 319–343. Hillsdale, NJ: Lawrence Erlbaum.

Ekman, P. and Friesen, W. 1969. "The repertoire of nonverbal behavior." *Semiotica* 1 (1): 49–98.

Fonagy, I. 1983. *La Vive Voix*. Paris: Payot.

Goffman, E. 1981. "Footing." In *Forms of Talk*, E. Goffman (ed), 124–159. Philadelphia: University of Pennsylvania Press.

Goodwin, C. 1981. *Conversational Organization*. New York: Academic Press.

Goodwin, C. 1984. "Notes on the structure and the organization of participation." In *Structures of Social Action*, J. M. Atkinson and J. Heritage J. (eds), 225–246. Cambridge: Cambridge University Press.

Goodwin, C and Goodwin, M. H. 2004. "Participation." In *A Companion to Linguistic Anthropology*, A. Duranti (ed), 222–244. Oxford: Basil Blackwell.

Lee, V. and Beattie, G. 1998. "The rhetorical organization of verbal and nonverbal behavior in emotion talk." *Semiotica* 120 (1–2): 39–92.

Kendon, A. 1967. "Some functions of gaze direction in interaction." *Acta Psychologica* 26 (1): 1–47.

Maury-Rouan, C. 2002. "Do Listener's facial expressions influence speaker's discourse?" Paper delivered at Gesture: The Living Medium, 1st Conference of the ISGS. Austin, Texas.

Maury-Rouan, C. and Noël-Jorand, M-C. in press. "Dialogism and dialogue : physical vs. imaginary presence of addressee's influence over discourse parameters." *Journal of Pragmatics*.

Maury-Rouan, C, Vion, R. and Bertrand, R. 2007. "Voix de discours et positions du sujet. Dimensions énonciatives et prosodiques." *Cahiers de Praxématique, 49, Numéro Spécial*: *À la recherche des voix dans le Discours*, J. Barberis (ed), 133–158.

McClave, E. 1998. "Cognitive and interactional functions of head movements in conversation." In *Oralité et Gestualité*: *Communication Multimodale, Interaction*, S. Santi, I. Guaitella, C. Cave and G. Konopczinsky (eds), 365–370. Paris: L'Harmattan.

Nölke, H. 1993. *Le Regard du Locuteur – Pour une Linguistique des Traces Enonciatives*. Paris: Kimè.

Ohala, J. J. 1980. "The acoustic origin of the smile." *Journal of the Acoustic Society of America* 68: S.33.

Vincent, D. and Perrin, L. 1999. "On narrative vs. non-narrative functions of reported speech: a socio-pragmatic study." *Journal of Sociolinguistics* 3 (3): 291–313.

Vion, R. 1998. "De l'instabilité des positionnements énonciatifs dans le discours." In *Pragmatics: Selected Papers from the 6th International Conference of the IPRA, Vol. 2*, J. Verschueren (ed), 577–589.

Voloshinov, V. N. 1986. *Marxism and the Philosophy of Language*. Cambridge: Cambridge University Press.