A CONTRASTIVE INVESTIGATION OF DISCOURSE INTONATIONAL CHARACTERISTIC FEATURES OF SOFIA BULGARIAN AND HAMBURG GERMAN IN MAP TASK DIALOGUES

Evelina Grigorova, Sofia University, fk6a004@uni-hamburg.de
Vladimir Filipov, Sofia University, filipov@fcml.uni-sofia.bg
BistraAndreeva, Institute of Phonetics, University of the Saarland,andreeva@coli.uni-sb.de

ABSTRACT

Ten MAP Task dialogues for Sofia Bulgarian (SB) and six for Hamburg German (HG) are recorded and analyzed by means of X-Waves Software Package. The discourse intonation features focused on are denial and convergence. It has been observed that for German denial can be integrated into discourse-listing through intonation: *Ja*-acknowledge and *Nein-/Ne*-denial moves are both manifested by intonation rises. For Bulgarian, intonation rises in answering moves occur only in the acknowledge subtype: rises in denials (*Ne-*) are associated with uncertainty and surprise. The HG *Ne*- and SB *Ne*-moves are resynthesized by means of PSOLA, twelve stimuli being obtained for SB and sixteen for HG. Two appropriate contexts marked for discourse-listing and follow-up moves are excerpted from the MAP Task and are included in perceptual tests whereby native speakers are asked to determine the appropriateness of each stimulus in relation to each context. 'The results for Bulgarian contradict our preliminary observations.

Convergence is defined as the matching of corresponding movements in pitch ranges and signals sympathetic agreement with the other speaker's point of view. The check: answer move sequence can be viewed as instantiating convergence and exemplifies both lexical and Fo movement repetition, especially where ellipted moves are concerned. The two resynthesized sequences for HG and SB respectively are "Im Westen" and "Pravo nagore" as manifested in check and answering contexts. As above, native speakers are expected to determine the appropriateness of each stimulus in relation to each context. It has been observed that the differences between checks and answering moves for both HG and SB are phonetically manifested and are also established as being relevant by the perceptual tests, yet they cannot be accounted for phonologically by tone alignment: convergence seems to attenuate the phonological differentiation between checks and answering moves.

1.INTRODUCTION

It is commonly accepted that intonational meanings are not only language specific and context bound but are also determined by some highly elusive factors. As a result the literature abounds in a multiplicity of often contradictory theoretical frameworks, whereby the majority of scholars attempt to arrive at a solution of the problem by basing their argumentation solely on one major criterion. [13] puts forward the idea of polysemy analysis of the Fo contours; [10] build their theory on the semantic concept of 'mutual beliefs'. Such a vein of analysis tries to establish a one-to-one correspondence between a contour and a discourse function. It presupposes implicit knowledge of the inherent features of discourse, whose explicit representation, however, including its prosodic characteristics(cf. [12]), is considered to be the objective of discourse analysis. Thus, on one hand, there is a rigidly defined phonology of specific languages, and, on the other, descriptions of the prosodic features of discourse [7], [2]. The two approaches hardly share any common methodological grounds.

Recent years have witnessed the advent of cooperative tasks for the elicitation of spontaneous speech, e.g. the MAP-Task [5]. The intonational analyses of cooperative task data strongly corroborate the claim we have made, i.e. a direct relation between a contour and a discourse function, hence meaning [3] or a description of global Fo tendencies [6]. The latter work provides two characteristic features of the cooperative dialogues, termed 'convergence' and 'divergence', respectively, the former occurring "when participants match corresponding movements in pitch ranges of utterance to …. signal enthusiasm and sympathetic agreement with the other speaker's point of view' (ibid.: 276). 'Divergence' is the opposite of 'convergence'. This paper focuses on the signaling of agreement and negation by means of intonation in the respective language varieties as parts of units bigger than moves.

2. THE EXPERIMENT

2.1. The MAP-Task dialogues

With MAP Tasks the analysis of the specific functionally defined moves tries to correlate intonational form with discourse function. Checks, i.e. questions further specifying information that has

been known to a certain degree already, are some of the most frequent moves in the MAP Task, followed in turn by a Reply, i.e. an answer to any question, the latter can be a check too. Often the functioning of a check is linked to some additional meaning, the latter being in turn manifested in its intonational form. Some scholars posit subtypes of the check move [3], [4]. Our MAP-Task corpus for SB and HG displays a peculiarity, i.e. consecutive check and reply moves can manifest identical contours as viewed within an autosegmental framework. According to our observations such signals of unanimity appear at informationally weak slots in the dialogue, namely where due to the fact that the expected information has already been somehow provided and at times adequately explained, the effect is as if the check is posed automatically. It is also a characteristic feature of such slots in the dialogue that both checks and replies are often realized as syntactically ellipted, the two different moves being lexically identical. Copying of the contour of the interlocutor and the neutralization of features typical of the two distinctive moves is usually glossed as 'I am with you, I share your train of thought'. Adapting the contour both in terms of form and range to that of the interlocutor or its copying are attested phenomena both in different registers and in child language [11].

As far as negation and contradiction are typical of the MAP Task dialogues due to the presence of misplaced objects on the maps, it is worth noting that their intonational realization in discourse is language specific. Our corpus also displays different realizations of the negation 'Nein' or 'Nee'/'Ne' depending on their function in the dialogue. German 'Nein' is more easily incorporated in larger parts of the dialogue such as listing questions, the latter being pitched similarly to 'Ja' of agreement on a rising contour. Under similar conditions Bulgarian 'Ne' seems to preserve its intonational identity.

To validate our observations concerning the specific and universal characteristics of the MAP Task dialogues we have set four perceptual tests with resynthesized stimuli from the originals: 'Nee' (reply in listing), 'Im Westen' (check), 'Ne' (question expressing uncertainty),and 'Pravo nagore' (check) excerpted directly from our corpus.

2.2. Description of the stimuli

In the case of negation 'Nee'/'Ne' (*No*) the relative timing of the level tone and the rise (on 'n' and on 'e', respectively, in the original stimulus) as well as the Fo-values reached at the end of the stimulus vary. For both 'Im Westen'/'Pravo nagore' the Fo range is manipulated; the relative timing of Fo_{max} is not varied, since we do not aim at establishing the typical alignments for interrogation (cf. [1]), but at the possible native speaker's interpretation of a small Fo contour variation in different contexts, i.e. we try to 'project' convergence in our stimuli. For 'Im Westen' the Fo variations are as follows: S1 (stimulus 1) is the original that can be phonologically described as L+H* L-L%, the maximal Fo value (Fo_{max}) being 286 Hz

located on the accented syllable 'Wes'; in S2 the Fo_{max} is lower - 260 Hz; in S3 it is 240 Hz and in S4 it is 220 Hz. In S5 the Fo on the last syllable is 200 Hz (100 Hz in the original utterance), so that the contour becomes stylized (phonologically it is L+H* !H-L%); in S6 Fo_{max} is 260 Hz and the Fo value in the last reduced vowel 'e' falls from 125 Hz to 115 Hz in the last 'n'; in S7 only the Fo_{max} changes to 240 Hz and in S8 Fo_{max} remains 240 Hz, but the Fo value in the last syllable is still 150 Hz .

For the SB Check 'Pravo nagore' the original S1 contains a late Fo peak on the end of the post-accentual syllable ('vo') and a dynamic accent on the second stressable syllable 'go' with a steep Fo fall; S2 has a level and lower tone on the post-accentual syllable 'vo'; in S3 the tone in 'vo' remains the same as in S2, but the Fo at the beginning of the dynamically accented syllable 'go' drops from 171 Hz to 161 Hz; in S4 it drops even more (to 140Hz) – this is the only difference compared to the original; in S5 a phonological change occurs, i.e. the first accented syllable 'pra' has lower values than in the original, so that the pitch accent becomes L*+H.

The resynthesis was conducted by means of PSOLA at the Institute of Natural Language Processing, Stuttgart .

2.3. Conducting the tests

Each set of stimuli is presented in random order to each group of subjects: 16 citizens of Hamburg, age group 25-55, and 15 citizens of Sofia, age group 21-40. The stimuli are placed in three different contexts. For negation these are: (1) Reply with listing; (2) Reply; (3) Question expressing uncertainty. For checks these are: (1) Check; (2) Reply;(3) Correction. Subjects are expected to evaluate each stimulus in relation to its suitability to a given context on the following scale: 1-'very good', 2-'good', 3-'I don't know', 4-'bad',5-'very bad' for HG; for SB, the scale is reversed, since higher numbers are associated with better results by Bulgarians.

2.4. The results of the tests

Subjects' responses have been processed using ANOVA at the Institute of Phonetics, University of Saarland, and the Department of Psychology, ELTE, Budapest [14].

2.4.1. Test Nee (HG)
Significant for this set of stimuli is the context, not the stimuli (except in context 3), i.e. the acoustic differences between the stimuli are not statistically significant. In context 1 and context 2 after the post-hoc test the stimuli group together: for Context 1 responses move from 'good' to 'I don't know', 'good' being generally given to stimuli with a steeper Fo rise within the vowel; for Context 2 responses move from 'I don't know' to 'bad', the critical boundary for disapproval being the final Fo value (Fo >325 Hz).

Context 3 shows significance of the stimuli. The homogeneous groups according to Scheffe are

three but there are quite large intersections amongst them. The first three stimuli of Group I are rated as 'good' for final Fo > 325Hz; Group III is rated from 'I don't know' to 'bad' together with the flattening of the steepness of the Fo rise.

2.4.2. Test 'Ne' (Bulgarian)

As for Bulgarian negation both context and stimuli are significant. Grouping of the stimuli for Context 1 after the post-hoc test is as follows: (1) Group I includes stimuli with high final Fo and low initial Fo, the group being rated as 'bad'; (2) Group II presents mainly 'I don't know' responses for heterogeneous stimuli, e.g. low initial Fo rise (S6, S7, S8) and high initial Fo (S3); (3) Group III is rated as 'good' due to the sharp steepness of the Fo rise within the vowel (S1, S2) and lower initial Fo and final Fo.

Context 2 displays three groups: (1) Group I includes 'bad' responses for low initial Fo and high final Fo; (2) Group II includes 'I don't know' responses for low initial Fo and steep Fo rise within the vowel; (3)Group III includes 'good' responses for heterogeneous stimuli.

Context 3 contains the same stimuli and ratings. Group II includes stimuli with low Fob and 'good' responses, and Group III –'very good' responses for the stimuli with initial Fo not falling below a determined critical value, i.e. 155Hz with final Fo < 290Hz.

2.4.3. Test 'Im Westen' (German)

As for this test it is only context that is significant, the result being repeated for every one of the contexts.

For Context 1 according to Scheffe one group is formed spanning from 'I don't know' to 'bad'. The latter is due to the lower values of Fo_{max} in the stressed syllable. What is surprising is that the original Check stimulus is not recognized as such.

Context 2 presents also one group of 'I don't know' responses, whereby only the original is rated as 'bad'.

Context 3 shows no significance of the stimuli, yet only one group is formed. The responses vary from 'good' to 'I don't know'. The stimuli with higher Fo end (on '-en') are rated as 'good', the best rating being given to S5, the latter presenting a stylized contour – L+H*!H-L%.

2.4.4. Test 'Pravo nagore'

Both context and stimuli are significant in this test.

Context 1 presents three homogeneous groups, whereby I and II can be brought together on the basis of 'bad' responses (Ss 1, 2, 3,4), S5 being rated as 'I don't know' is in Group III. It is worth noting that none of the stimuli is recognized as Check.

In Context 2 there is a distinct dichotomy of the groups after the post-hoc Test. Group I's only member is a 'bad' response S5, whereas Groups II and III include Ss 2, 1, and Ss 3, 4 with increasing approval. The 'bad' rating of Group I may be due to the presence of a L*+H tone on the initial syllable. The approval given to Group II can be explained by

the lower beginning of the second stressable syllable 'go', hence the gradual Fo decrease in it and associates better with finality and deaccentuation. In the remaining cases a dynamic accent occurs on 'go', accompanied by a sharp Fo decrease within the syllable.

Context 3 presents three groups too: (1) Group I – 'bad' for S5 and S4. In the former case it may be due to the fact that the L*+H tone is not perceived as suitable for correction, whereas in the latter there is deaccentuation of the second stressable syllable '-go-'; (2) Group II consists of S3 rated as 'I don't know', the explanation being similar to that of S3, yet in the former the Fo characteristics of the prominent syllables 'pra' and 'go' prevent the complete deaccentuation of 'go'; (3)Group III includes 'good' responses due to the distinct prominence of the two stressable syllables 'pra' and 'go'(S1, S2).

3. DISCUSSION

3.1. Negation

The results of both tests for HG and SB show that generally HG subjects cannot discern distinctive differences in the proposed stimuli either for a 'Listing reply' or for 'Reply'. As for Context 3, 'a question with uncertainty', it is only the high final Fo stimuli that are accepted as 'good', the high final Fo value being probably perceived as the carrier of interrogation and uncertainty.

SB subjects consistently differentiate the stimuli for every context. What is common for all situations is that low Fob and high Fof stimuli are not accepted, the latter factor being probably perceived as unnatural. The result that the rising Fo contour of 'Ne' is perceived as 'good' in 'Listing reply' despite the lack of similar realizations in the corpus comes as a surprise. In this case SB subjects perceive the way HG subjects do though more categorically so: the results are 'good' in the cases of steep Fo rises within the vowel. Even more surprising is the result that approval is given to a heterogeneous group of stimuli in a Reply context, the result being due to the greater leeway in the realization of the different connotations in reply. For 'questions expressing uncertainty' 'good' is given to stimuli with higher ΔFo and both for initial Fo and final Fo threshold values are determined.

3.2. 'Convergence'

In this part of the experiment the SB subjects are again more consistent in their judgments of the stimuli. The HG subjects do not recognize any of the stimuli as Check, in spite of the fact that the original occurs in such a type of move. It may be that the L+H* tone is not informative enough for the recognition of interrogation despite the frequent occurrence of similar realizations in the MAP-Task corpus. A probable yet trivial explanation can be that for a question the subjects expect a high boundary tone – H%. Another tentative solution is that in the

prominent syllable Fo does not rise quickly enough or that Fo_{max} does not reach sufficiently high values. The latter explanation contradicts to some extent the results of Context 2 where it is with low Fo_{max} values that recognition of the reply does not take place. The alignment of Fo_{max} of all stimuli in the test is equivalent to 'late peak' in the sense of [8]. Its pragmatic meaning is 'addition of emphasis'. This makes the indeterminate result of Context 2 even harder to explain and yet gives grounds to believe that Correction will be recognized. 'Good' is given to stimuli having a higher Fo of the last syllable '-en' than the original, 'very good' being given to the stylized contour L+H*!H-L%.

The SB results have high statistical significance. Check is not recognized despite the presence of S5, in which the only tonal accent (on 'pra') is L*+H, a tentative explanation being the insufficient Fo rise in the post-accent syllable. S5 is given bad rating in Context 2 too which does not allow for the treatment of the L*+H accent as a prenuclear 'referring tone'. 'Very good' is given to the stimuli manifesting deaccentuation of the second dynamic accent in the stressable syllable '-go-'carried out by a gradual Fo fall within the latter. It is well known that this brings about good perception of finality. Exactly the opposite result occurs for Correction. This move is better recognized in the presence of a second dynamic accent on '-go-' which makes it sound more didactic.

4. Conclusions

The HG results show that there is a tendency to categorize the stimuli, yet no systematic rating is given of the possible variations. A possible explanation for the dispersion in the responses may be the big age span of the subjects. In support of the former are the results with natural stimuli arrived at in [9] which also show that the HG subjects look mostly for categorial specification, although the SB results show that the subjects look for interpretation even of small variations of the acoustic parameters of the stimuli. This is even more valid for negation.

ACKNOWLEDGEMENTS

5. REFERENCES:

[1] Andreeva, B. & W. Barry (1997), Intonation von checks inder Sofia-Varietaet des Bulgarischen. *Proceedings of FDSL2,Potsdam.*
[2] Couper-Kuhlen, E. & Selting, M. (eds) (1996), Prosody inConversation. Cambridge: Cambridge University Press.
[3] Grice, M. et al. (1995), The intonation of queries and checks across languages. *Proceedings of XIIIth ICPhS'95Stockholm.* Vol. 3, pp. 648-651.
[4] Grigorova, E. (1996), Rhythmische und intonatorische Besonderheiten des Bulgarischen in kooperativen Dialogen (unpublished Habischrift).
[5] Kowtko, J. Isard et al. (1992), Conversational games in dialogues. *HCRC Publications.* Edinburgh: University of Edinburgh.
[6] Li-chiung Yang (1995), Prosodic structure and discourse organization. *Proceedings of XIIIth ICPhS'95Stockholm.* Vol. 2, pp. 274-277.
[7] Local, J. et al. (1986), Towards a phonology of conversation: turn-taking in Tyneside English. *Journal of Linguistics22*, pp. 411-437.
[8] Kohler, K. (1991), A model of German intonation. *AIPUK25,* pp. 295-360.
[9] Misheva, A. & E. Grigorova (1997), Kontrastive Untersuchung des Einflusses der Muttersprache auf die Perzeption von suprasegmentalenCharakteristika der Rede am Beispiel Deutsch/ Bulgarisch. In: Simeonova, R.& H. Wodarz (Hrsg.) (1997) *Kontrastive Intonationsstudien Deutsch/Bulgarisch.* Hector:Frankfurt am Mein.
[10] Pierrehumbert, J. & J. Hirschberg (1990), The meaning ofintonational contours in the interpretation of discourse. In: Cohen, P.;Morgan, J. & M. Pollak (eds) (1990), *Intentions in Communication.* The MIT Press: Cambridge, Mass. pp. 271-311.
[11] Schaerlaekens, A.; Forrez, G. & L. van Bael (1990) Mimicry of intonation in 10- to 12-month-old children: a failure to replicate. *Clinical Linguistics and Phonetics,* vol. 4, No. 2, pp.139-144.
[12] Selting, M. (1995) Prosodie im Gespraech. Tubingen: Max Niemeyer Verlag.
[13] Taylor, J.R. (1989) Linguistic Categorization: prototypes in linguistic theory. Oxford: Clarendon Press.
[14] Vargha, A. (1999) MiniStat 3.1 verzió: Felhasználóikézönyv. Budapest: Pólya Kiadó.