Ropa, A. 1982. „Przedmiot i zadania fonetyki percepcyjnej". In: Maciejewski, J. (ed.). 1982. 49–56.
Sawashima, M., Hirose, H. 1983. "Laryngeal gestures in speech production". In: MacNeilage, P. (ed.). 1983. 11 – 38.
Singh, S. (ed.). 1981. *Measurement, Procedures of Speech, Hearing and Language*. Baltimore: University Park Press.
Sołtys-Chmielowicz, A. 1989. „Jeszcze o wymowie bezdźwięcznej". *Biuletyn Audiofonologii*. Tom I, zeszyt 1. 49 – 54.
Sołtys-Chmielowicz, A. 1998. *Wymowa dzieci przedszkolnych. Seria: Komunikacja językowa i jej zaburzenia 15*. Lublin: Wydawnictwo UMCS.
Stevens, K. N., Klatt, D. H. 1974. "Role of formant transitions in the voiced-voiceless distinction for stops". *JASA* 55. 653 – 659.
Stevens, K. N., Keyser, S. J., Kawasaki, H. 1986. "Toward a phonetic and phonological theory of redundant features". In: Perkell, J. S. et al. (eds). 1986. 426 – 463.
Stieber, Z. 1966. *Historyczna i współczesna fonologia języka polskiego*. Warszawa: PWN.
Styczek, I. 1979. *Logopedia*. Warszawa: PWN.

REALIZACJA KONTRASTU DŹWIĘCZNOŚĆ–BEZDŹWIĘCZNOŚĆ PRZEZ DZIECI NIESŁYSZĄCE

Streszczenie

W artykule podjęto temat realizacji kontrastu dźwięczność/bezdźwięczność w obrębie spółgłosek zwarto-wybuchowych w pozycji inicjalnej wyrazu przed samogłoską. Badaniami objęto 20-osobową grupę dzieci z głębokim uszkodzeniem narządu słuchu, rehabilitowanych z wykorzystaniem metody fonogestów, oraz 10-osobową grupę kontrolną dzieci słyszących. Podjęta analiza fonetyczno-akustyczna związana z pomiarem parametru VOT wskazuje na znaczne zaburzenia w realizacji badanego kontrastu w mowie dzieci niesłyszących. Podjęto dyskusję z dotychczasowymi polskimi ujęciami tego problemu w zakresie patologii mowy, określanymi jako tzw. 'mowa bezdźwięczna' oraz opracowano typologię realizacji dźwięczności w mowie zaburzonej. Pomiary parametru VOT uzyskane w grupie kontrolnej dzieci słyszących wyraźnie wskazują na istnienie fonetycznego kontrastu pomiędzy kategoriami głosek dźwięcznych i bezdźwięcznych (*voicing lead and voicing lag*), wykazują zróżnicowanie związane ze zmianą miejsca artykulacji, choć można wnioskować o ciągłym doskonaleniu artykulacji w tym zakresie.

NUCLEAR MELODY IN POLISH SEMI-SPONTANEOUS AND READ SPEECH: EVIDENCE FROM THE POLISH INTONATIONAL DATABASE *PoInt*

KATARZYNA FRANCUZIK, MACIEJ KARPIŃSKI, JANUSZ KLEŚTA, EMILIA SZALKOWSKA

ABSTRACT: Katarzyna Francuzik, Maciej Karpiński, Janusz Kleśta, Emilia Szalkowska. *Nuclear Melody in Polish Semi-Spontaneous and Read Speech: Evidence from the Polish Intonational Database Point*. Studia Phonetica Posnaniensia, volume 7, 2005, pp. 97-128. Adam Mickiewicz University Press. ISBN 83-232-1532-4. ISSN 0860-2085.

This paper presents a few findings concerning the nuclear melody based on the analysis of the recordings contained in The Polish Intonational Database *PoInt*. In its first sections, it provides a short description of the collection in question, mentions certain common problems connected with intonation labeling as well as outlines the relevant theoretical background. Then it focuses on selected relations between the category of a given utterance as well as the discourse function it realizes and its nuclear intonation.

Katarzyna Francuzik, Maciej Karpiński, Janusz Kleśta, Emilia Szalkowska, Instytut Językoznawstwa, Uniwersytet im. Adama Mickiewicza, ul. Międzychodzka 5, 60-371 Poznań, e-mail maciejk@amu.edu.pl

1. THE POLISH INTONATIONAL DATABASE *PoInt*

All the analyses presented in this paper were based on the study material contained in the Polish Intonational Database *PoInt* (conf. [Karpiński & Kleśta 2001]; [Karpiński 2002a]; [Karpiński, 2002b]), which is based on a corpus of contemporary Polish spoken by educated native users of the language. The database includes nearly 1200 samples, up to four intonational phrases each, representing both (quasi)spontaneous and read speech realized by male and female speakers, which were obtained with the aid of previously developed special monolog and dialog tasks as well as written texts taken from contemporary Polish literature. Each of the signals can be played in its natural form or in a version

containing the "hum" only[1], and is accompanied by a spectrogram, an intonogram, as well as orthographic and phonological transcriptions (IPA). The intonational and pragmalinguistic annotations available for each of the speech samples included in the database along with a user-friendly graphical interface and an effective searching device the application is equipped with make it to be a useful source of data for research in the domain of prosody and pragmatics as well as a practical aid in the area of teaching phonetics, phonology or even Polish as a foreign language. The annotated phrases can also be used as reference material in the construction of advanced speech synthesis systems, or form the basis for the preparation of test material utilized in the process of testing speech recognition systems, especially those making use of prosodic features for the purpose of signal interpretation.

2. THE PROBLEMS OF INTONATION LABELING

Fox [2000] remarks that "intonation has always been perceived as a problem" in the area of prosody research. As Cutler and Ladd [1983] point out, the two main traditions in this domain, viz. acoustic-phonetic and linguistic-phonological, or, in other words, bottom-up and top-down, did not notice each other for a long time, and even today the gap between the phonetic and the linguistic studies of intonational phenomena remains quite substantial. The evidence of discrepancies between these two approaches (and their intermediate variants) can be found in the symbolic representation systems for prosodic phenomena. The number of prosody labeling systems reflects the complexity of this issue. Not only a particular theoretical approach may give birth to another labeling procedure and a new set of symbols, but also very practical needs may call for a special, modified or entirely new labeling system. All the same, there are many common phenomena that must be considered and described by each "serious" theory of intonation.

One of the essential questions is whether intonation can be studied in separation, or, more precisely, whether it functions as an independent system, not related to the remaining prosodic properties of the speech signal. Consequently, is one allowed to study intonation, neglecting, for example, the rhythmic structure? According to our present knowledge, isolating intonation is only acceptable in a limited range of applications, while everywhere else it has to be analyzed along with other factors such as voice quality and volume or pace of speech (e.g., [Wennerstrom 2001:46]).

[1] The word "hum" is a commonly used term relating to the fundamental melody extracted from an original speech sample.

Another crucial aspect of intonation studies is the choice of the basic analysis unit and its relation to the linear structure of the utterance. Most researchers agree that spoken utterances consist of intonational units, termed "intonational phrases" or "intonational groups" (see: [Cruttenden 1994:35] for other terms and relevant discussion). Although on many occasions, especially in well-formed utterances, these units tend to correspond to or be aligned with syntactic phrases or some other formally distinguishable portions of text, in spontaneous speech this nearly one-to-one correspondence may be heavily disturbed.

Still another important issue is the difference between "perceived", "produced", and "objectively measured" intonation. One can expect that, as it is the case with some of the other language subsystems, there is no one-to-one mapping between the "conceived" (planned), produced and perceived intonational contour. This can be exemplified by the problem of phrase boundary perception, which was frequently faced by the authors of this article during the labeling procedure within the *PoInt* Database project. Different understanding of a given passage sometimes led the judges to hearing the phrasal boundary in different positions. Another example is related to the syllables pronounced in a creaky voice or even with full devoicing, which were still perceived as having an identifiable pitch height. Since most instrumental analyses are only partially relevant to human perception, it is, however, highly recommendable to take advantage of the help of experienced annotators, rather than rely on "linguistic" judgments based solely on instrumental analyses.[2]

Since the list of questions related to intonational labeling is very long, we have to concentrate on those of the highest practical importance. It seems that the choice of the description level and the decision concerning the basis of the labeling system are the issues which each student of intonation has to address before undertaking any serious work.

2.1. LEVELS OF DESCRIPTION

Linguistic studies of prosody have led to a wide discussion on the levels of annotation. While acoustic-phonetic research may be focused solely on the acoustic level, i.e. the signal itself, the linguistic perspective on intonation has enforced a more top-down oriented approach to prosody as being processed in the language comprehension and production processes along with other linguistic information. With the advent of pragmalinguistic views on prosody, a need

[2] At Prosody 2004 in Nara, P. Mertens presented new software that is able to label segmented text automatically on the phonetic-perceptive level [Mertens 2004]. It was not available in the course of our project.

arose for a system that would depict its role in language and its contribution to the meaning. It became clear that one could study intonation on a number of different "levels", ranging from a very technical analysis of the speech signal acoustic properties to relatively abstract pragmatic concepts (like "speech acts" or "discourse topics"), as they relate to intonation in everyday language usage.

One can easily realize that the two abovementioned "extreme" approaches to the study of intonation are nowadays avoided and most researchers make attempts at finding a "middle way". Purely physical-acoustic approach is quite limited, because it is well known that humans process speech signal in a specific way, different from the ways they process other sounds. On the other hand, an extreme "top-down" approach could turn out to be of little use and may be hard to base on experimental data. Consequently, the students of prosody are free to operate on a number of different analysis levels, somewhere between instrumental measurements and "armchair" theories, provided they are explicit about the level they use. Although the instrumental measurements and statistically-motivated data processing are claimed to be "objective", the status of this "objectivity" can be easily undermined by the fact that the algorithms used (and regarded as "objective") are mostly meant to create images or numeric representations of signal properties that are meant to meet our intuitional expectations. One must be fully aware that they only produce interpretations of the signal – very clear-cut and coherent ones but still not necessarily "objective" in a very wide sense. With a more "subjective" approach, however, new doubts come into consideration. One presupposes that there should exist a phonological-intonational system – an element of the language system – and phonological-intonational competence that enable humans to encode and decode intonational meaning. On the other hand, in the light of our present knowledge on spontaneous speech, this system should be extremely flexible, both for speech production and perception. There is virtually an unlimited number of possible intonational contours for, for example, a statement. Even if one decides to choose a finer category, like for example "a sad statement" or "a statement full of confidence", there are still many ways to intonationally express the qualities in question. One should also take the account of the speaker's and listener's meaning in this context (([Grice 1975], [Genova 2002]; see also the critique of Stevenson's theory by Grice). These and other facts lead to the attempts at finding a precise place for intonation in the system of language. It is therefore extremely important to define at which level our analysis operates, though even this may prove to be quite hard.

Hirst et al [2000] as well as Jassem [2002] offer clear and detailed discussions of the levels of analysis and description. Jassem proposes the following four levels, differing from those suggested by Hirst *et al* in minor details only:

Level 1: Pitch extraction (F_0 as a function of time)

Level 2: Smoothing the pitch curve and normalization as to abstract from personal variability

Level 3: Perceptual-phonetic level

Level 4: Phonological level

Although Jassem optimistically predicted that software for converting the annotational data between different levels would be shortly available, there are still numerous questions to answer about the details of level-conversion. As for today's state of art, it is still impossible to label intonation on the phonological level fully automatically. Human judges remain a must, but we still need more precisely defined methodology, because the degree of agreement among judges is usually not satisfactory. Another daunting possibility is that our image of the place and function of intonation in the language system is still quite distorted.

2.2. LEVELS AND DIRECTIONS

The character of intonation perception and its reflection in the method of prosodic labeling has been one of widely discussed issues in the field of prosody research. The main controversy in question is referred to as "levels vs. contours" [Cruttenden 1994] or "levels vs. configurations" [Fox 2000]. As Cruttenden [ibidem] points out, the level-based approach can be attributed mostly to the American tradition and the focus on intonational contours is typical of the "British School". While it can be argued that both these approaches are roughly descriptively equivalent [Fox, ibidem], it may be important to judge their "theoretical efficiency". One easily notices that the list of benefits and drawbacks is almost equally long for both approaches in question. In the case of level-based analysis, one needs to decide almost arbitrarily on the number of levels (see: [Bolinger 1951]) and assigning phonetically labeled contours to phonological categories may be harder, while with contour-based analysis, it can be easier to group labels into categories, but there is the risk of overlooking certain nuances of intonational meaning by, for example, neglecting the depth of falls or rises. T'Hart and his colleagues [1990] reasonably emphasize the perceptual significance of pitch movements, overtly claiming that there are no "levels". Still, it is technically convenient for many purposes to use the symbols of levels, even if in fact the description is focused on contours. A widely known and commonly applied system that originates from Pierrehumbert's [1980] seminal work uses only two levels as a sort of binary opposition. The occurrences of "intermediate levels" are regarded as results of certain phonetic phenomena (downstep, declination) and classified as phonologically irrelevant. In this way, one of the main problems with level-based approach is eliminated: There is no obligation to de-

clare any fixed number of levels. One should also note that with any higher number of levels, the question about the categorical perception of intonation arises, and it occurs surprisingly hard to be unequivocally answered [Post 2000: chapter 7].

3. NUCLEAR MELODY LABELING IN *PoInt*

The procedure of intonational labeling has been strictly defined and consequently employed for the entire material contained in the *PoInt* collection. While this is not enough to ensure "objectivity", the authors do hope that a relatively high level of coherence in annotation may be achieved by such means.

3.1. SYSTEM

There is a rather limited choice of intonation labeling systems dedicated to the Polish language. All of those known to the authors of the present text are focused on the theoretically important issue of representing the intonational-phonological system. However, they do not offer much in terms of practical advice for the labelers, leaving much space between the signal itself and the abstract labels depicting the meaningful aspects of the intonational contours. The comprehensive theory of intonemes proposed by Steffen-Batogowa [1996] is a very clear example of this approach, but a similar statement may well refer to Borkowska's and Skorek's work [2002]. Impressive and practical methods of intonation labeling (e.g. musical score) are used by Gubrynowicz and his collaborators for pathological speech [Gubrynowicz 2002]. Majority of Demenko's works (e.g. [1999]) are usually focused on the technological aspects of prosody annotation. However, in order to selectively label a large portion of read and spontaneous texts, the *PoInt* team needed a flexible, extensible system, easy-to-use and focused on the most important aspects of speech melody (i.e. "focal elements" of utterances).

Several factors led to the application of Jassem's system. While it was initially intended solely for well-formed utterances, it proved to be easily extensible and modifiable so that it could encompass more phenomena. It offered the option of focusing on the nuclear melody itself and the phonological level of description. Providing a basis for the intonational-phonological system, it did not exclude adding new units to the system. Moreover, the author claimed that the system was still open to modifications, and the labeled material provided a way to finding new support for his claims. Another factor was the role attributed by Jassem to the practice of annotation, which remained in accord with the approach of the entire team.

According to Jassem's suggestions [2002 as well as earlier oral communication], the following system was applied. Initially, five relative pitch heights were defined:

- extra low (xL);
- low (L);
- middle (M);
- high (H);
- extra high (xH).

A typical label consisted of two tone symbols, with the first one referring to the tone of the nuclear syllable and the second to the tone of the post-nuclear one (e.g., HL, LxL). The nuclear syllable usually occurs in the penultimate position for Polish, but in monosyllables it obviously takes the final position, which means that labels consisting of one tone symbol only were also possible (e.g. H, L).

Only a certain subset of all possible pitch movements is phonologically relevant. Namely, the following movements are "possible":

- raising: LH, LM, MH, HxH;
- falling: HL, HM, ML, LxL;
- flat: MM (accent is realized by means of other acoustic features such as duration or energy).

The inventory of tones for the nuclear syllable was more limited than for the post-nuclear one.

Since the system in its original form was meant for "well-formed" utterances only, it was extended for the purposes of *PoInt* labeling to achieve higher flexibility and cover certain phenomena that occurred to be of importance during the process of labeling. The extensions did not change the pivotal ideas of the system though and should be regarded as optional. Extensions were proposed after a preliminary analysis of the collected data.

Three types of level tones were marked: LL, MM and HH. Although the labelers could not systematically attribute clear and different "meanings" to them, they declared to perceive them as "different". Further analysis of the *PoInt* data may prove that some of these labels are superfluous on the phonological level, but neglecting them at the very beginning might have led to a loss of important information. The same refers to the second extension, i.e. adding an additional tone label for the final syllable when the nuclear syllable occurred earlier than in the penultimate position. (e.g. labels like HM-L). The final extension to the original system was actually a move down to the acoustic-phonetic level of annotation. In a substantial portion of the analyzed material, the final parts of the signals (e.g. the ultimate syllables of the intonational phrases) were produced in a way that made their "intonational perception" impossible. That was mostly due to one or more of the following phenomena: Final devoicing (*D*), extremely low

energy of the final segments (*E*) creaky voice in the final segments (*C*) or strongly shrunk syllable realization (*S*). The labels *D*, *E*, *C* and *S* are obviously not directly connected with the phonological or semi-phonological level of analysis, which is why they were put in square brackets.

3.2. THE LABELING PROCEDURE

Most labeling systems, with their specific theoretical backgrounds, call for the application of a specific labeling procedure. However, as was mentioned before, explicit and detailed descriptions of labeling procedures are not easy to find. One of the exceptions is the IViE labeling system ([Grabe, Post, Nolan 2001] as well as the materials available on the IViE website). Even though the authors call it "an extension to ToBI", it seems to add a new quality that makes it more useful and flexible. The IViE system offers the phonetic level of transcription. Although the authors justify its presence by the requirements of comparative studies, its meaning seems to be even deeper as it creates a desirable bridge between the pitch contour and the phonological level of its representation. The labeling procedure itself consists of clearly defined steps (finding prominences, labeling relative heights of the nuclear syllable and its neighbors, labeling on the phonological level). The labeling procedure used for PoInt was based on that of IViE, although it also added some new elements. Each portion of recordings was labeled by at least three judges. The labeling itself was preceded by discussion, team-labeling of a number of examples as well as "individual practice". It involved the following stages:

Step 1: Dividing a portion of signal into intonational phrases (IPs)

Even at that early stage, several serious problem had to be addressed. First of all, an implementable and clear definition of the intonational phrase had to be found. With a number of definitions in hand, the team decided to rely on the following formulations: (a) an IP boundary may be marked by a pause, but a pause itself cannot be regarded as an unequivocal marker of the boundary; (b) rapidly falling energy of the speech signal may be regarded as a marker of the boundary; (c) easily perceivable nuclear melody may mark the end of an IP in many "neutral" phrases; (d) the rhythmic structure of a signal may offer strong cues about the placement of the boundaries; (e) in many cases, the acoustic cues are not sufficient to determine the boundary position; in such situations, syntactic and semantic cues may prove to be helpful.

As mentioned above, the relations between the IP boundaries and the boundaries of "syntactic components" in spontaneous texts are not very clear.

What holds mostly for "well-formed" phrases, frequently fails in the analysis of spontaneous, emotional speech. In certain situations, the judges understood the same passage differently, which resulted in different phrasing and different perception of the nuclear syllable location. This may be considered as evidence of the weight of top-down processing in "human" labeling, although, on the other hand, one could also assume that it simply resulted from a difference at the "bottom end" of the perception procedure (i.e. possibly different perception of the signal at the early acoustic stage). More discussion on the speaker's and listener's perspectives on the prosodic phrasing can be found in, e.g., [Volskaya, Skrelin 2003].

When speech is well-formed grammatically, it is more probable to be well-formed phonetically, too. Therefore, the cases where there were no significant acoustic cues for phrasal boundaries were usually extremely hard to judge, because they – as a rule – lacked grammatical cues, too. Some stretches of speech were marked as "unfinished" phrases (*F). It took place when a given stretch of speech lacked perceivable nuclear melody and the most perceivable cue to its end was the beginning of the next intonational phrase.

Step 2: Finding nuclear syllables

Although the intonational structure of well-formed utterances offers space for one nuclear syllable only, in less controlled, spontaneous (unprepared) speech, a few candidates to play the role of the nucleus can sometimes be found. Some authors show that in most situations the informational structure of an utterance is related to the distribution of pitch accents (e.g. [Selkirk 1995]), but, as mentioned before, the choice of the candidate may be quite subjective. It may strongly depend on the contextual cues as well as the expectations of the listener, rather than the signal itself.

During the labeling procedure, two basic situations were distinguished:

1. The nuclear syllable occurs in its standard position, i.e. on the penultimate syllable of the last word in the IP (pre-penultimate in some words of foreign origin and ultimate in stressed monosyllables).

2. The nuclear syllable is intentionally shifted towards the beginning of the IP in order to mark the focus, express an emotional attitude or for some other reasons.

In this step, acoustic measurements were not taken into account. Three or four judges marked the nuclear syllable independently and discussed their choices, making use of close listening.

Step 3: Labeling tones

The labelers were supposed to use the inventory of tones and possible tonal configurations as listed in 3.1. After the stage of common labeling sessions, they

worked independently on the same portions of signals and met to discuss their results. The most important feature of the procedure was the method of "close listening". However, instrumental measurements, listening to a slowed-down version of the signal as well as studying the graphic representations of intonational contours were also used as secondary tools. When "close listening" produced incoherent or unclear results, the judges used *Praat* [Boersma, Wenink, ver. 3.8 and later] to create intonograms, spectrograms, as well as to listen to selected segments (usually syllables) of speech in isolation.

As mentioned earlier in this section, some segments of the analyzed signals could be hardly labeled intonationally because of their acoustic qualities (devoicing, low energy, creaky voice, syllable shrinkage or a combination of these). If the labelers did not completely agree about the character of the observed phenomenon, they used instrumental methods to determine its nature.

4. RESULTS

4.1. DATABASE PROFILE

As was mentioned before (cf. paragraph 1), the *PoInt* Database includes 1144 samples, up to four intonational phrases each. Consequently, the total number of labeled intonational phrases in the collection reaches almost two thousands.

Most of the nuclear melodies were labeled with two tone symbols (more than 1600 cases), almost 200 were marked with three symbols (where the nuclear syllable position was different from penultimate), and 134 labels consisting of one tone symbol only were used (where the nuclear syllable coincided with the final syllable in the phrase).

The phrases produced in the read texts were realized most uniformly – a vast majority of the melodies were labeled with two tone symbols, which may be due to the fact that the read speech was well organized and generally free from hesitations. Besides, the arbitrary choice of the texts obviously imposed a way of realization, including the choice of focus.

Almost all of the melodies labeled with three tone symbols in the read texts occurred in the child rhyme 'Entliczek...' (64 cases), probably owing to a specific, relatively strong rhythm pattern applied here by part of the speakers. As can be seen in the figures presented below, the read texts constitute the most numerous collection of samples – in fact they account for the majority of all the recordings in the database.[3] The figures also show that the two-letter symbols

[3] In the *PoInt* Corpus, however, semi-spontaneous speech is prevalent. The choice of the signals for the database was intended, among other aims, to give the user the opportunity to compare textually (and contextually) identical utterances realized by different speakers, which (here) was only possible with read speech.

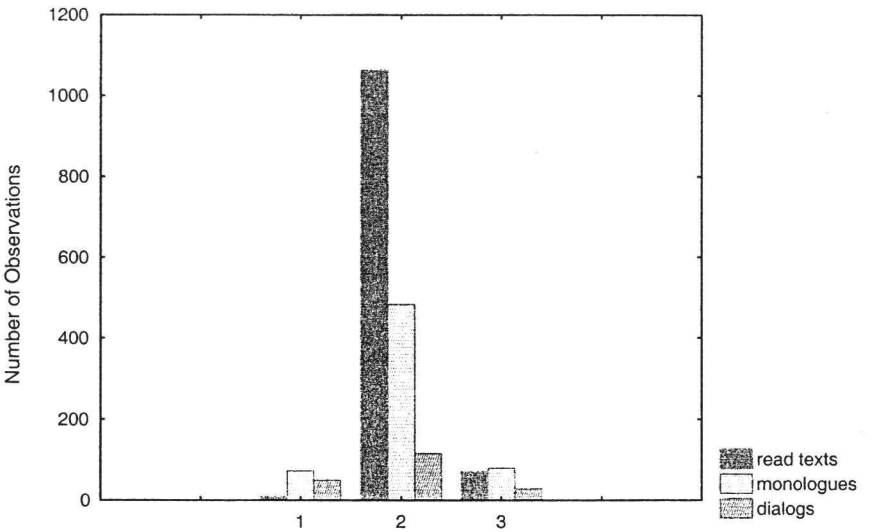

Fig. 1. The number of one-, two- and three-symbol labels within the read texts, monologs and dialogs

prevailed in all the three recording types, however the percentage of one- and three-letter symbols was relatively higher for the spontaneous speech realizations than for the read ones.

4.2. DATABASE INVENTORY – PITCH MOVEMENT DIRECTIONS

The labels applied by the material annotators fall into three main categories, viz. including one-, two-, and three-symbol labels respectively. For each of these categories two subcategories are distinguished.

One of the above subcategories includes the labels relating to the intonational phrases with a distinct melodic structure, which enabled unequivocal description using a combination of the five basic pitch level symbols: M, L, xL, H and xH. The resulting pitch movements within the abovementioned IPs can be thus classified as rising, falling, flat, monosyllabic (for one-symbol labels) or compound (for three-symbol labels). Even though it is possible to describe the compound melodies using a combination of the three basic pitch movement types, viz. falling, rising and flat (for example an ML-H label could be translated into a falling-rising melody), they are treated here collectively, mainly because of their relatively small proportion in the analyzed material as well as considerable within-group variation.

The other of the abovementioned subcategories includes the labels relating to the intonational phrases whose melodic structure was hard to determine. This was the case when an IP was unfinished / distorted or when a relevant portion of the signal was extremely low in energy, completely devoiced, produced in a creaky voice or strongly shrunk. Reliable determination of the pitch levels and the resultant pitch movements in the case of such phrases was virtually impossible.

As can be seen in the following table, the most frequently used melody patterns are, as could have been predicted, the falling ones. It is, however, worth noticing that melodies incorporating a creaky syllable rank as high as third with 363 occurrences altogether, which is why they have been extracted from the relevant subcategory in the following classification.

Table 1. *PoInt* Database Inventory. Melody Types in the *PoInt* Database

Annotation Category	Melody Type	Number of Occurrences	Comments
One-Symbol Labels	Monosyllabic	104	The pitch level of the final syllable in the IP distinctly identifiable.
	Creaky	9	The final syllable in the IP produced in a creaky voice.
	Other	21	The IP unfinished / distorted or the final syllable in the IP devoiced / strongly shrunk / extremely low in energy.
Two-Symbol Labels	Falling	592	The pitch levels of the two final syllables in the IP distinctly identifiable.
	Rising	455	
	Flat	240	
	Creaky	319	At least one of the two final syllables in the IP produced in a creaky voice.
	Other	57	The IP unfinished / distorted or at least one of the two final syllables in the IP devoiced / strongly shrunk / extremely low in energy.
Three-Symbol Labels	Compound	132	The pitch levels of the three relevant syllables in the IP distinctly identifiable.
	Creaky	35	At least one of the three relevant syllables in the IP produced in a creaky voice.
	Other	11	The IP unfinished / distorted or at least one of the three relevant syllables in the IP devoiced / strongly shrunk / extremely low in energy.

4.3. MALE AND FEMALE SPEAKERS – DIFFERENCES AND SIMILARITIES

Certain differences in the choice of melody types applied by male and female speakers have been observed. In general, men tend to use the falling melody patterns much more frequently than women while the rising melody patterns appear to be far more popular with females. It is, however, much more remarkable that women seem to produce creaky sounds nearly twice as often as male speakers (cf. fig 2 below).

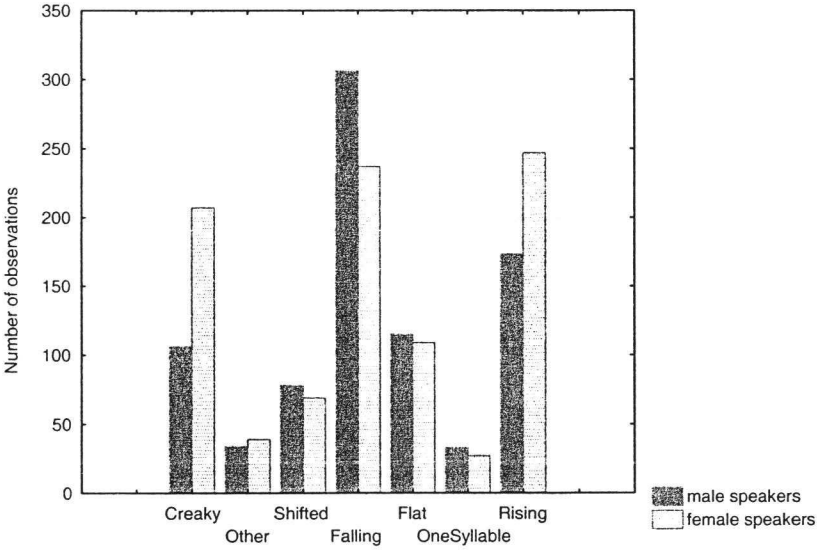

Fig. 2. Melody types realized by male and female speakers

Specific pitch movements applied by male and female speakers have also been studied. Even though women most frequently utilize the LH movement, while men prefer the ML and HL melodies, four out of the five most frequently used patterns (LH, MH, ML, HL) are identical for both sexes. The second most common pitch movement produced by women, however, is that labeled as LC, while the classification of the top five melodic patterns realized by men is complemented with that marked as LxL (cf. fig. 3 below).

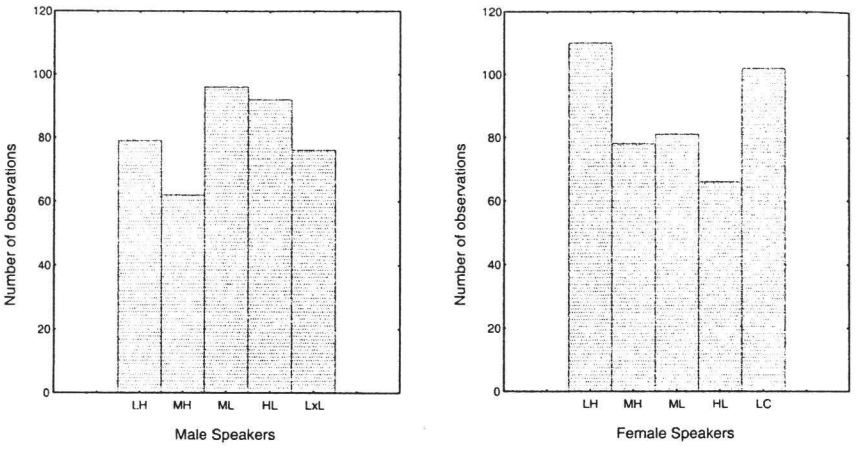

Fig. 3. The five pitch movements most commonly produced by male and female speakers

4.4. FINDINGS

4.4.1. READ TEXT REALIZATION VARIATIONS

4.4.1.1. ENTLICZEK PENTLICZEK

One of the tasks performed by the speakers participating in the realization of the *PoInt* Project consisted in reading a popular Polish child rhyme with a strong rhythmic pattern entitled „Entliczek Pentliczek"[4]. Three extracts from the rhyme were ultimately chosen for analysis and placed in the final version of the *PoInt* Database. The first two lines of the rhyme were selected with a view to analyzing the phenomenon of nuclear accent shift in the case of rhythm-driven prominence distribution, while the remaining two fragments, consisting of one line each, were chosen in order to investigate the relationship between the emotions expressed by the speakers and the intonational patterns they apply. The first of the above two lines was a rhetorical question containing an element of irony or compassion (depending on the reader's text interpretation), and the other one expressed strong protest accompanied by anger or, again depending on the speaker's interpretation, powerless opposition accompanied by tiredness or even resignation.

[4] "Entliczek" and "pentliczek" are nonsense words in Polish.

4.4.1.2. RHYTHMICAL WELL-FORMEDNESS RULES

Thirty percent of the intonational phrases connected with the production of the first line: "Entliczek pentliczek czerwony stoliczek" (Eng. "Hickory dickory dock, a little red table"), were realized with the lexical stresses on the word-initial syllables. This accentuation position, so unusual for the Polish language, can stem from the fact that the beginning of the text in question is often used as a child counting-out rhyme, where the utterance of the first syllable in each of the words is strongly emphasized by the accompanying pointing at consecutive group members. As all the words in the first line are trisyllables, the stressed syllables in this line become equally distant. It also appears that apart from repeating the same foot pattern (/ – –) in order to maintain the metrical well-formedness of the rhyme some of the speakers also strived for identical prominence of all the stressed syllables, which made it very difficult sometimes to reliably identify the IP nucleus.

It also happened at times that in the line: "No widzisz robaczku. I gdzie twój befsztyczek?" (Eng. "You see little worm. Where is your beefsteak now?"), the stressed syllables in the words relating to the given information items (Pol. *robaczku* and *befsztyczek* – Eng. *worm* and *beefsteak*) were as prominent as the stressed syllables in the words connected with the newly introduced information items (Pol. *widzisz* and *gdzie* – Eng. *see* and *where*). Such productions strongly undermine the validity of the nuclear accent theories which assume the presence of a single main prominence per IP only, either in the default position (i.e. on the last stressed syllable of the IP) or at a location shifted towards the beginning of the phrase – when it is necessary to put special emphasis on a newly introduced item of information.

The question thus arises whether it is possible, or reasonable, to identify intonational phrases with their nuclear melodies where intonation is strongly associated with the rhythm and the foot as a rhythmic unit. Concentrating on the intonation of smaller units, i.e. on foot intonation, and identifying pitch accents and patterns only, rather than looking for the nuclear syllables, seems to be much more appropriate in such cases (either as proposed by Pierrehumbert in her model [1980] or as Wennerstrom [2001] suggests in the context of spontaneous speech analysis).

4.4.1.3. NUCLEAR MELODIES REALIZED WITHIN A SINGLE SPEECH ACT

Almost all of the speakers read the second line of the rhyme: "*A na tym stoliczku czerwony koszyczek*" (Eng. "*And on this table [there is] a little red basket*") as a statement consisting of two intonational phrases (only one reader real-

ized it as a single IP). The first of the abovementioned IPs: /a na tim stɔlitʃku/, which constitutes the non-final part of the utterance, had a nuclear melody rising to the level of H or xH in 19 cases (43% of all the realizations under analysis) and a flat melody (HH, MM or LL) in another 16 cases (36%). A falling nuclear pattern was only observed in 6 productions of the IP in question (13,5%), while of the remaining two, one incorporated a creak on the post-nuclear syllable and the other was realized with a compound rising-falling melody. The second of the IPs (the final part of the statement) was, however, almost exclusively uttered with a nuclear fall to the level of L or xL or with a low flat nuclear melody LL. Domination of such patterns was also observed in spontaneous speech (cf. monolog data).

As has been partly confirmed by the above example, the nuclear melodies imposed on the non-final IP of an utterance related to a given speech act seem to be much more diverse than those realized on its final intonational phrase. This may result from the fact that the former are only responsible for signaling that the utterance has not reached its end yet (which can apparently be done using a range of different melodic patterns), while the latter are crucial for the correct decoding of the speech act type in question by the message recipient.

4.4.1.4. THE MEANING OF DIFFERENCES WITHIN A SINGLE NUCLEAR MELODY CATEGORY

The line: "*A ja już nie mogę, już dosyć, już basta*" (Eng. "*And I can't any more, I've had enough, basta*") was always read as three separate intonational phrases, and approximately a quarter of all the speakers consistently imposed the same falling melodic pattern on each of the IPs (e.g. HL + HL + HL). As could have been predicted, falling melodies generally dominated in the realizations of the utterance under analysis, with rising patterns accounting for about 5% of the productions only (7 out of 129). Even though the former could be assigned to a common category of nuclear falls (or transitions from a higher to a lower level, using the terminology of another theoretical approach), it appears that the size of the fall (or, in other words, the distance between the starting and the terminating pitch level) as well as the rate of the above change (the time needed to cover the distance) are also significant. In terms of emotions and attitudes, the HL melody in our example seemed to emphasize the expression of strong protest, while the LL or LxL melodies were rather associated with an overtone of tiredness or resignation. A rapid transition between the pitch levels of the nucleus and the subsequent syllable, resulting from a generally faster pace of speech within the realization of the utterance concerned, contributed, in turn, to the impression of stronger impatience and desperate willingness to change the current situation.

Surprising as it may seem, melodies with a greater pitch level span, which were identified with a stronger intensity of emotions in our example, were much more frequently applied by men, who made use of the HL melody approximately twice as often as women. The realizations of the latter, in turn, included twice as many examples of the LL, LxL and LC melodies as those of the former, with the ML and MC nuclear melodies accounting for about 30% of all productions in the case of both sexes (cf. fig. 4). Therefore, male speakers seem to have made use of a wider pitch range to express their emotions here than women, who mostly utilized only the bottom areas of their fundamental frequency, although it has been observed that in emotional spontaneous speech women tend to use the LH nuclear melody much more frequently than men, who prefer rising melodies of a smaller pitch span in such situations.

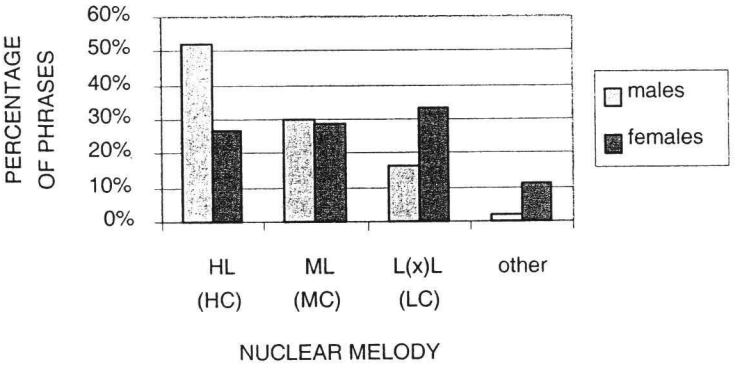

Fig. 4. The proportion of IPs with individual nuclear melody patterns in the productions of the rhyme line: "a ja już nie mogę, już dosyć, już basta" by both genders

4.4.1.5. NUCLEAR MELODY IN QUESTIONS CONTAINING INTERROGATIVES

The aim of this part of the study was to find out whether Polish speakers exhibit a strong tendency to realize questions containing an interrogative word with a single dominant nuclear melody type – with a fall, as it is the case in French, Portuguese and many other European languages, with a rise, which is predominantly applied in some African languages such as for example Telugu, or with a fall-dominated fall-rise nuclear melody type, commonly imposed on interrogative word questions of such languages as Norwegian or German [cf. Cruttenden 1996].

The nuclear melodies of the following two questions were put under analysis:

(1) "*I **gdzie** twój befsztyczek?*" (Eng. "*And **where** is your beefsteak now?*") and

(2) "*Babuniu, **jak** ten duży pies się nazywa?*" (Eng. "*Grandma, **what** is that big dog's name?*")

Both of the sentences had been read by the speakers, with the former being a part of the child rhyme described above and the latter constituting a fragment of contemporary Polish prose.

The first of the above questions was realized with a fall by 84% of the speakers taking part in the recordings (37 out of 44) and the second one was produced with a falling melody type by 63% of them. When the nuclear syllable in the above productions coincided with the last word accent of the IP, the nuclear melody was a combination of any two tone levels generating a falling pitch movement, with a preference, however, for narrow span falls, i.e. LL and LxL. Therefore, the melody was similar to that of declarative sentences in Polish, with the slight difference, however, that the pitch in the initial part of the IP was often a bit higher than it is the case in Polish declaratives, as a result of extra pitch height imposed on the question pronoun. In the case of such realizations, some listeners tended to identify this question word as the nuclear syllable of the intonational phrase, rather than the last accented syllable of the IP, which is the default position for the nuclear accent in Polish. (cf. Baranowska et al. 2003). When the nuclear syllable was perceived to coincide with the interrogative word, the melody patterns were then usually labeled as HM-L, which seems to reflect a continuous fall throughout the whole IP, although such labels as HL-L or HM-M were also included in the above annotations.

It seems that it may be the rule of language economy that is indirectly responsible for declarative-like realization of interrogative word questions in Polish: one cue per IP, viz. the interrogative word at the beginning of the wh- question, may be sufficient to distinguish this speech act type from a declarative, and that might be exactly why intonation is employed here to emphasize this very cue at the beginning of the phrase rather than to provide an additional one in the form of a rising melody at its end, which is, in turn, necessary in the case of most yes/no questions, often devoid of an interrogative word in contemporary Polish. It is not to say, however, that the kind of redundancy mentioned above does not occur in Polish wh- questions at all. The first and the second of the sentences under analysis were realized with a final rise by 16% and 37% of the speakers respectively. In the case of such productions, the LM and LH melodies were mostly favored (16% and 8% of all question word IPs respectively), with the MH and HxH patterns realized only once each. The justification for the presence of two separate question markers here (an interrogative at the beginning and a rising nuclear melody at the end of the phrase) might be the additional

discourse meaning conveyed by the IP. The extra rise, depending on its precise pattern, may be employed here to express one of such specific attitudes of the speaker towards the message recipient as interest, sympathy or irony [cf. Cruttenden for other languages].

4.4.2. NUCLEAR MELODY IN THE REALIZATION OF DIALOG MOVES

The idea of "dialog games" or "conversational games" is relatively old and can be traced back at least to late Wittgenstein ("language games"; [Wittgenstein 1953]). The term "dialog move" is, however, used with different meanings depending on the researcher. Many important issues related to dialogue unit taxonomies are insightfully discussed by Traum [2000].

The dialogue move labeling system used in PoInt was elaborated on the basis of the commonly known and widely tested HCRC Map Task scheme. The labeling methodology described in detail in Carletta et al [1995, 1996] proved extremely useful, especially that it was directly related to the seminal work of Kowtko [1996], who studied the intonational realization of certain dialogue moves.

It transpired, however, that due to the differences among dialogue types and their specific features, a more flexible, extended taxonomy was needed for the purpose of *PoInt* dialog labeling. Some ideas found in Carletta & Isard [2002], [Mann 2002] and [Traum 2002] as well as our own labeling experience allowed us to establish a new classification, related mostly to the intentions of the speakers, but also to the interpretation of their utterances by their conversational partners. The knowledge of the dialogue situation, of the aim of the conversation as well as of the speakers themselves (most of whom were familiar to the project team), assured a reasonable dose of security in putting forward hypotheses concerning the interlocutors' "intentional states" in the course of the dialog, while the actual roles played by their turns could be derived from the surrounding context. The classification is by no means intended to be an exhaustive one for all possible dialogue types. The following inventory of moves was proposed: *Statement*, *Instruction*, *Suggestion*, *Wh_question*, *Polar_question*, *Yes_answer*, *No_answer*, *Wh_answer*, *Acknowledgement*, *Confirmation*, *Self-repair*, *Correction*, *Explanation* and *Request_for_explanation*.

Since the phrases incorporated in the *PoInt* Database were chosen arbitrarily to show intonational variation in contemporary Polish, the findings presented below should be interpreted exclusively as considerations of possible intonational realization, and definitely not as generalized or normative claims.

Eighty-six stretches of speech were selected to be placed in the dialogue section of *PoInt* in accordance with the general rules accepted for the database, i.e. (a) one

signal should not contain more than four intonational phrases, (b) it should not exceed four seconds in duration, (c) and it should comprise a "meaningful linguistic unit" (at least one "finished thought"). IP boundaries as well as nuclear melodies were labeled. Some of the selected speech fragments consisted of a few short dialog turns, while others contained only one. Subsequently, the signals were divided into discourse units and labeled as dialog moves using the abovementioned taxonomy. It was noted that only a limited number of dialog moves were realized as single IPs, while many of them spanned over two or even three consecutive intonational phrases. In a number of cases, this resulted from disfluencies in speech (part of an utterance was repeated, repaired etc.), but it also happened in fluent speech. For multi-phrasal moves, a problem arose with the interpretation of multiple nuclear melodies. Brief descriptions of the realizations for the most common moves, along with analysis examples, are presented below.

A. Instruction (order)

Eight speech fragments labeled as *instructions* were derived from the map task recordings, three of which were realized as single IPs. If a move consisted of two intonational phrases, the first of them was the "primary instruction", while the second just added extra details (explanation, stress, narrowing-down etc.). Seven out of the eight moves contained at least one syllable with a high tone (H), which was arguably used to highlight the information central to the instruction. Even in this small sample, the variety of the nuclear melodies realized was astounding. Both full rises and falls were found, including occurrences of the extra low fall (LxL) as well as examples of the H and HH nuclear melodies.

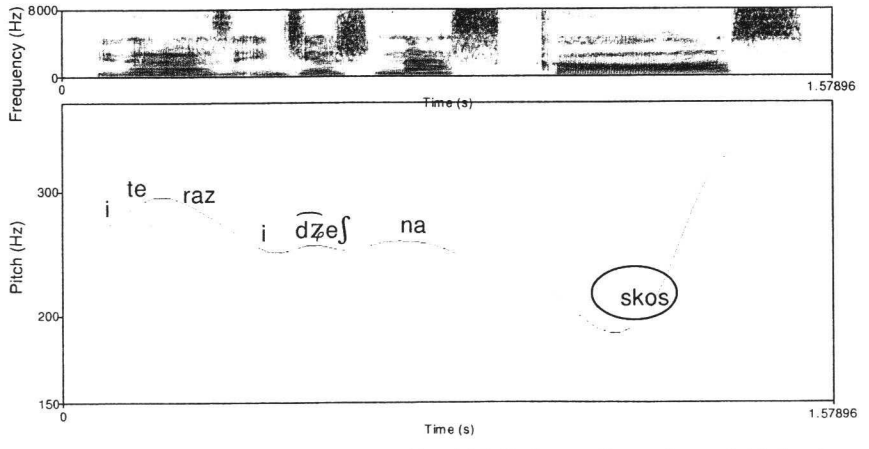

Fig. 5. An example of instruction realized with a high final tone (the nuclear syllable has been surrounded with an ellipse)

B. No/Yes_answer

Among the seven no-answers included in the material under analysis, five were realized using two intonational phrases and the remaining two stretched over a single IP only. All of these phrases were classified as non-rising, including the two cases where the dialog move consisted of a single IP with the nuclear accent on the ultimate syllable. Among all the occurrences of *Yes-answers* in our material, only one was realized with a melody which was classified as rising. Considering the observed proportion of non-rising melodies in the productions of both *Yes-* and *No-answers* in reply to polar questions, one could therefore formulate a tentative hypothesis that the above two move categories are normally produced with non-rising intonation.

C. Polar_question

Eighteen *Polar_questions* were included in the database. Six of them were realized using two intonational phrases and the remaining twelve stretched over a single IP. In three cases the nuclear syllable was shifted towards the beginning of the phrase and in another two reliable identification of the nuclear melody was impossible due to the pre-final syllable produced in a creaky voice. Among the remaining thirteen moves, most were realized with a high rise (four MH and two LH pitch contours) or with a high level melody (two examples of the HH pattern). Altogether, in 12 out of the 18 cases under analysis, the final tone of the final (or the only) IP was perceived as H.

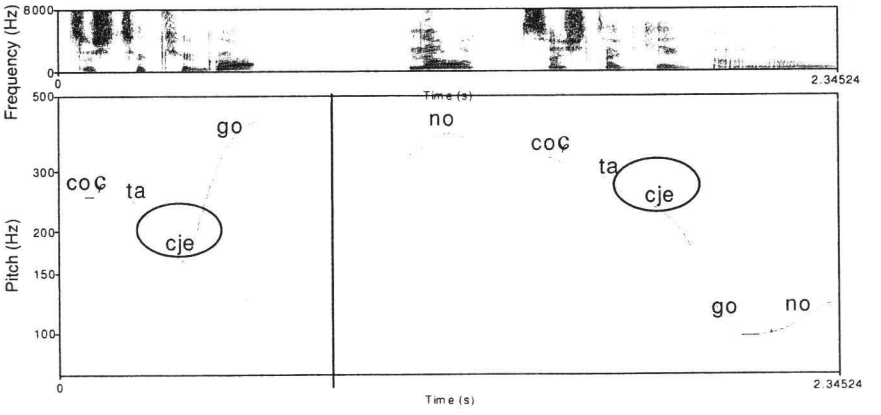

Fig. 6. A polar question with a rapidly rising (LH) melody and a y_answer with a rapidly falling pitch contour. The vertical line divides the text into two consecutive turns. Approximate translation: "Something like this? Well, something like this, yes.". The second turn is extended with the particle "no"(which was here translated as "yes"), produced with a slightly rising intonation

D. Wh-Question

Six examples of *Wh-questions* are available in the dialog section of the database. One of them consists of three intonational phrases, another one contains two and the remaining four stretch over a single IP. All the *Wh-questions* realized using one intonational phrase have a rising nuclear melody. The last IP in the most complex (three-IP) move was also produced with a rising melodic pattern, but the final phrase of the two-IP question, whose nucleus is shifted towards the beginning, has a falling melody (HM-L) throughout its entire duration.

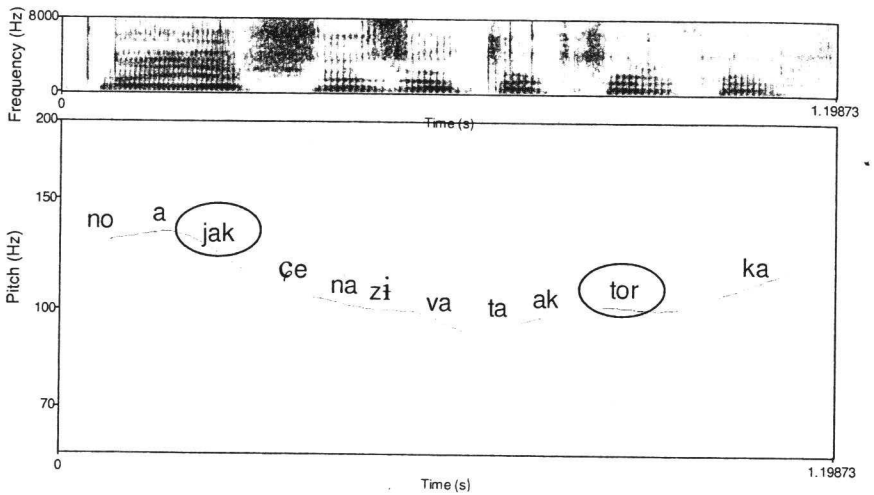

Fig. 7. A wh-question with a rising nuclear melody (approximate translation: "And what is that actress's name?"). Please note that the utterance starts relatively high and its entire intonation could be described as falling-rising. The gentle slope in the first half is probably associated with declination

E. Statement

Thirty-eight utterances in our material were labeled as *statements*. Five of the moves were realized using three intonational phrases, another fifteen contained two and the remaining eighteen stretched over a single IP. Shifts of the nuclear syllable towards the beginning of the phrase occurred in 9 cases. Although most of the contours here can be described as, at least, non-rising (which remains in accordance with the expectations based on theoretical knowledge), we have also found examples of statements with a clearly rising melodic pattern. A relatively large proportion of the nuclear melodies associated with the realization of statements resisted appropriate instrumental analysis due to the presence of creaky voice in their pronunciation. Sometimes, the nuclear melody had a creaky characteristic throughout its entire duration, so in six cases, quite interestingly, even the nuclear syllable was produced in a creak, for which reason its labeling in terms of height was hardly possible.

F. Detailed Answer (Wh-answer)

Nine occurrences of the wh-reply move were found in our database, three of which were realized using three intonational phrases, another three contained two and the remaining three stretched over a single IP. Except for one case, where the final phrase of the move realization finished in a high tone, the nuclear melodies here were never rising.

IPs in PoInt dialogs: Conclusion

This brief review of the *PoInt* Database dialog section was only supposed to give a general idea about the possible intonational contours that speakers utilize when realizing various dialogue move categories. Even if the number of phrases labeled here is too low for any generalizations, the variety of nuclear melody patterns used can be stunning, especially when we take account of the fact that the applied labeling system itself eliminates a huge amount of information by ignoring the prenuclear melody completely. It seems that a reasonable explanation of this situation may require not only considering the informational structure of the utterances under analysis but also taking their emotional load into account. A more comprehensive and exhaustive labeling system is currently in use with major portions of the *PoInt* corpus dialog section in order to enable tracking down the relations between prosody and dialog structure.

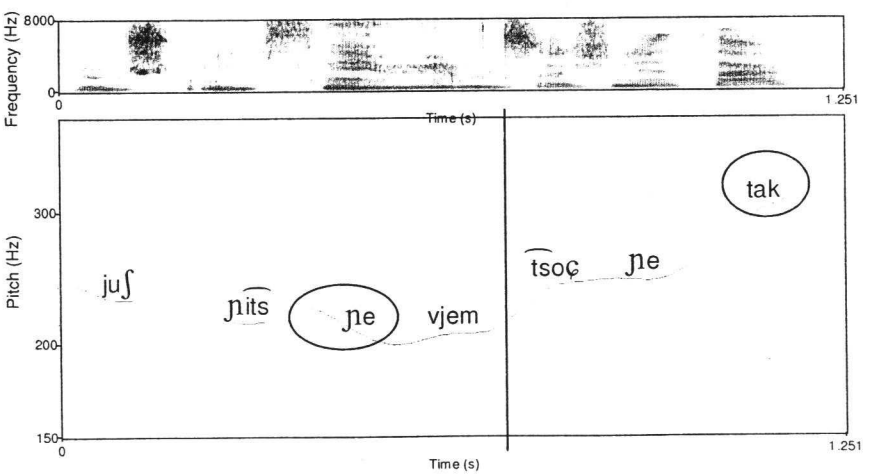

Fig. 8. A statement (with a slightly falling pitch contour) followed by a yn-question (with a rising melody ending in a high nuclear tone). Approximate translation: "I'm completely lost now. Is something wrong?" The vertical line divides the text into two consecutive moves

4.4.3. ATYPICAL REALIZATIONS OF STATEMENTS IN MONOLOGUES

4.4.3.1. NUCLEAR SYLLABLE LOCATION AND NUCLEAR MELODY CONTOUR IN STATEMENTS

Since over 97% of all *PoInt* Database utterances obtained as a result of monolog elicitation were declaratives, we decided to investigate in detail the location of the nuclear syllable and the nuclear melody contour in the final intonational phrase of statements. Our study material included 374 intonational phrases, 189 of which had been produced by female speakers and the remaining 185 by males. Figure 9 provides detailed information on the content of our data corpus in terms of the nuclear melody types realized (conf. section 3.1).

As can be seen in the below figure, four of the phrases under investigation (approximately 1%) were marked as "unfinished phrases" (symbol F) and were excluded from further analysis. All the remaining phrases were assigned to one of the following four categories, depending on the type of statement they were embedded in:

ST: neutral statement without any element of qualitative evaluation or personal opinion and without a perceptually identifiable distinct emotional load,

ST_EV: statement with an element of qualitative evaluation or personal opinion but without a perceptually identifiable distinct emotional load,

ST_EV_EM: statement with an element of qualitative evaluation or personal opinion and with a perceptually identifiable distinct emotional load,

ST_EM: statement without any element of qualitative evaluation or personal opinion but with a perceptually identifiable distinct emotional load (extremely rare).

Figure 10 shows the number of phrases assigned to each of the above categories. As can be seen, the ST-type and the ST_EV-type phrases, whose number in the collected data corpus was similar, accounted for over 91% of the whole material, while the size of the other two categories was substantially smaller.

All the nuclear melody types presented in Figure 9 were also categorized. They were divided into the following three major categories, depending on the symbol representing the tone of the phrase-final syllable:

HIGH: nuclear melody types ending in a syllable whose tone was marked as H or xH,

MEDIUM: nuclear melody types ending in a syllable whose tone was marked as M,

LOW: nuclear melody types ending in a syllable whose tone was marked as L or xL, as well as those whose final syllable was marked as C (for Creaky), D (for Devoiced), E (for extremely low Energy) or S (for strongly Shrunk).

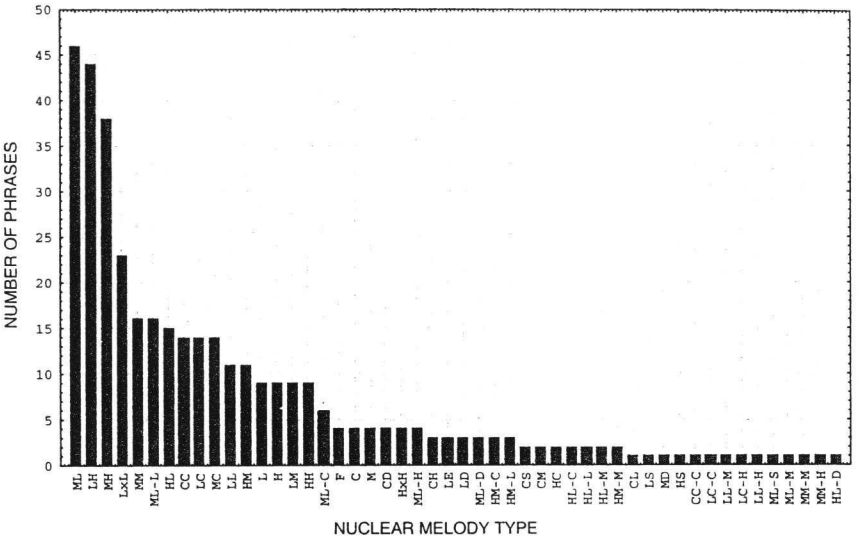

Fig. 9. The number of occurrences for all nuclear melody types realized in the study material

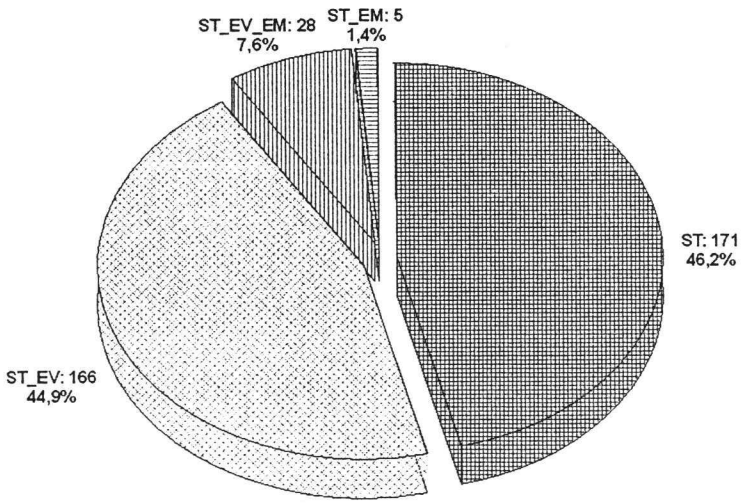

Fig. 10. The number and the proportion of phrases representing individual statement types in the collected study material

Figure 11 presents the number of phrases contained in our study material which fall into each of the above defined categories. As could have been predicted – given that the utterances under analysis are statements – the majority of intonational phrases in our data corpus (approximately 56%) were realized with a LOW-type nuclear melody; it is remarkable, however, that as many as 114 of them (over 30%) were realized with a HIGH-type nuclear melody, which can be quite surprising in view of traditional intonational descriptions for the Polish language.

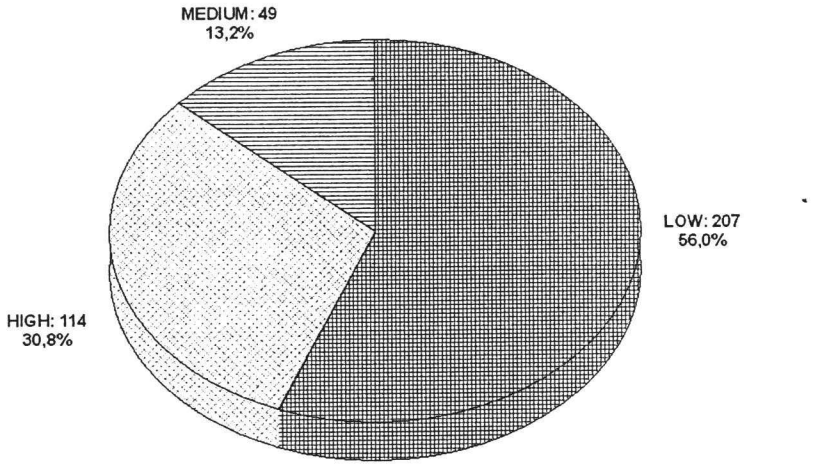

Fig. 11. The number and the proportion of phrases representing individual nuclear melody categories in the collected study material

The first stage of analyses concerned the location of the nuclear syllable in the phrases under investigation, with two situations possible here:

NON-SHIFTED NUCLEUS: the nuclear syllable coincides with the last stressed syllable in the phrase or

SHIFTED NUCLEUS: the nuclear syllable precedes the last stressed syllable in the phrase, which means that the onset of the nuclear melody is shifted towards the beginning of the phrase.

As transpires from the collected data, the nucleus was shifted in 53 cases only (approximately 14% of all phrases). It is remarkable, however, that in the male productions the nucleus shift took place almost twice as often as in the phrases produced by female speakers (see fig. 12).

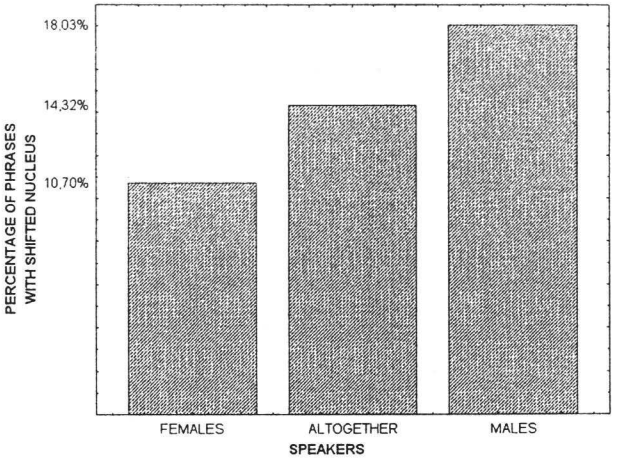

Fig. 12. The proportion of phrases with nucleus shift in the productions of both genders

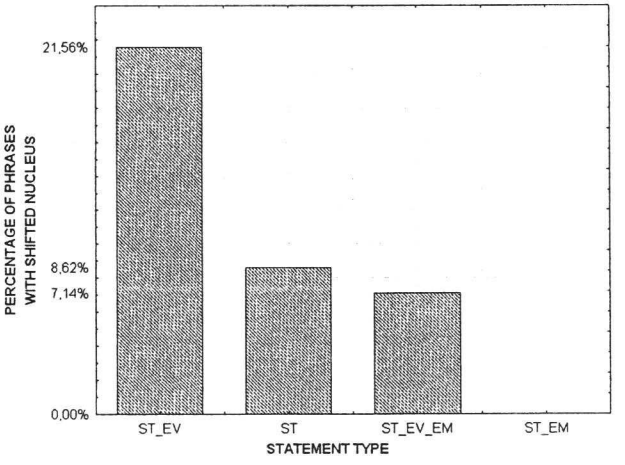

Fig. 13. The proportion of phrases with nucleus shift depending on the statement type

Subsequent analysis of statement types showed in turn that the nucleus was usually shifted in statements with an element of qualitative evaluation or personal opinion (nearly thrice as frequently as in neutral statements), however the addition of an emotional load seemed to reduce the tendency for nucleus shift regardless of the presence of the evaluative factor (conf. fig. 13).

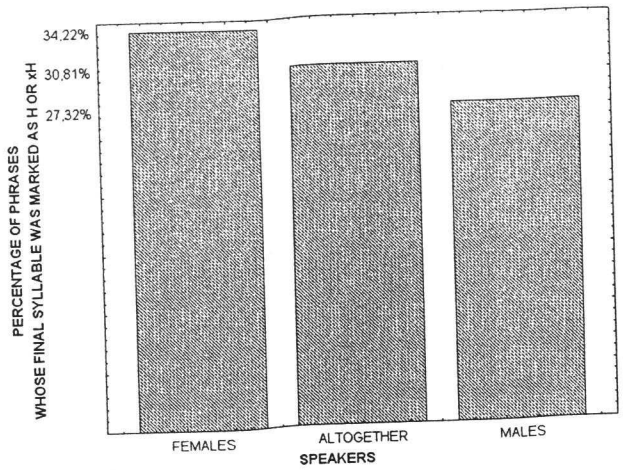

Fig. 14. The proportion of phrases with a HIGH-type nuclear melody in the productions of both genders

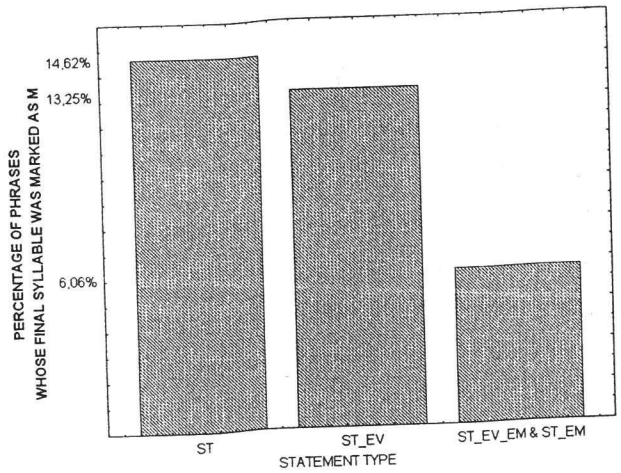

Fig. 15. The proportion of phrases with a MEDIUM-type nuclear melody depending on the presence of an emotional load in the statement

The second part of analyses concerned the application of individual nuclear melody categories depending on the speaker's sex and the type of statement. It was found that female speakers were more prone to impose HIGH-type intonational contours on their utterances than males, even if the difference in this respect was not as significant as one could have predicted basing on the descriptions available in the literature (conf. fig. 14).

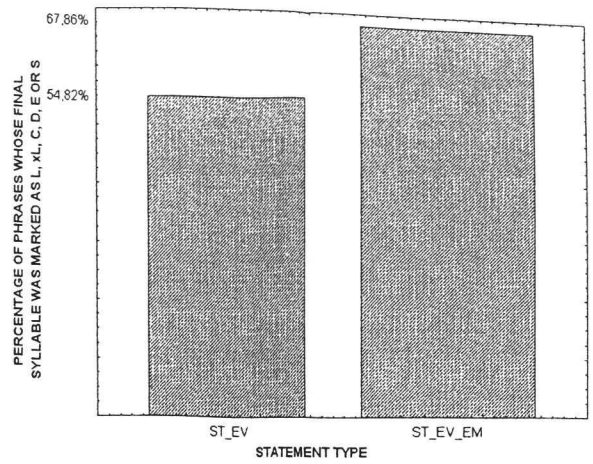

Fig. 16. The proportion of phrases with a LOW-type nuclear melody depending on the presence of an emotional load in the evaluative statement types

It was also found that the category including MEDIUM-type intonational contours, which was generally the least numerous one in our data set (approximately 13% of all phrases – conf. Figure 11), had hardly any representation among statements with an emotional load (see Figure 15), whose addition to a statement was in turn very often associated with an enhanced tendency for the imposition of a LOW-type nuclear melody (see Figure 16 below). The above-mentioned tendency may, however, partly result from the fact that the emotions expressed in the collected material were mostly of negative nature and were usually annotated pragmalinguistically as *dislike*, *aversion* or *disgust*, *sadness* or *depression*, *disregard* or *boredom*, *disappointment* or even *indignation*.

5. CONCLUSION

Although the content of the *PolInt* Database is strongly selective and arbitrary, a number of conclusions concerning the general tendencies as well as peculiarities in the area of Polish intonation can be arrived at on the basis of the collected material. First of all, the variety of possible (acceptable, effective etc.) intonational realizations for the majority of dialog moves and speech act categories is striking. The complexity of the relations between the text itself (or the mental representation of the text at various stages of its production) and its intonational realization leads to the obvious inference that a great number of factors contribute to the final suprasegmental structure of any utterance. It seems, how-

ever, that some of the above factors may play a dominant role in certain situations, while remaining much less significant in others. For example, rhythmic factors are definitely more important in the case of reciting poetry than they are when reading prose.

In the light of the above findings, a hypothesis can be put forward that the final suprasegmental structure of an utterance is determined by a dynamically changing configuration of the following factors:

- *information and discourse structure*: the way information is conveyed in the text, the arrangement of basic units as well as the rhetorical/discourse relations between them; these factors seem to be equally important in both prepared and spontaneous speech (perhaps with the exception of highly emotional speech), but the way they influence the intonational characteristic of oral production can strongly vary;
- *syntactic structure*: more or less related to the abovementioned factors (in spontaneous, informal, everyday communication these aspect seems to play a secondary role; please note that the syntactic structure may be sometimes ambiguous, and even if it is not, it is liable to misinterpretation by the reader);
- *text form* ("textual gender", style etc. – e.g. poetry, academic lecture, sermon);
- *emotional attitude* (to the content of the text and to the listener);
- *individual speaking style* – the features that are typical of a given speaker and can be traced down in most of his utterances, independently of the current textual gender;
- *external factors* (like noise or any sort of unexpected events affecting the speaker).

Without a reasonable amount of knowledge about the roles and weights of the abovementioned factors, a number of phenomena commonly occurring in spontaneous speech can hardly be explained. A given shape of a part of an intonational contour may result from a number of interacting factors. This, in turn, leads to a conclusion that our further research should be focused on explaining the impact of the factors in question as well as on their possible interactions. While such a task seems to be extremely complex and painstaking, we do hope that some tendencies found in the entire material can be explained on a simpler, less detailed level, appropriate for speech technology applications (e.g., dialog systems design).

REFERENCES

Baranowska, E., Francuzik, K., Karpiński, M., Kleśta, J. 2003. "The Identification and Meaning of the Nuclear Melody Placement in Polish Read Texts". *Interfaces Prosodique*. Nantes.

Batogowa-Steffen, M. 1996. *Struktura przebiegu melodii polskiego języka ogólnego*. Poznań: Wydawnictwo Sorus.

Boersma, P., Weenink, D. 2005. "Praat: doing phonetics by computer (Version 4.3)" [Computer program]. Retrieved from http://www.praat.org/

Bolinger, D. 1951. "Intonation: Levels versus Configurations". *Word* 7. 199–210.

Borkowska, K., Skorek, J. 2002: *Podstawowe jednostki intonacji języków rosyjskiego, białoruskiego i polskiego*. Zielona Góra: UZ.

Carletta, J. et al. 1996. *Dialogue Structure Coding Manual*. (project materials).

Carletta, J., Isard A. 2002. "Dialogue Structure Annotation Using a Pipeline of Simple Grammars". ISLE *Workshop on Dialogue Tagging for Multi-modal Human Computer Interaction*, Edinburgh, December 15–17.

Carletta, J., Isard, S., Doherty-Sneddon, G., Isard, A., Kowtko, J., Anderson, A. 1997. "The reliability of a dialogue structure coding scheme". *Computational Linguistics* 23(1). 13–31.

Cruttenden, A. 1994. *Intonation*. CUP.

Cutler, A., D. Ladd, R. (eds). 1983. *Prosody: models and measurements*. Heidelberg: Springer.

Demenko, G. 1999. *Analiza cech suprasegmentalnych języka polskiego na potrzeby technologii mowy*. Poznań: Wydawnictwo UAM.

Fox, A. 2000. *Prosodic Features and Prosodic Structures*. Oxford: OUP.

Genova, D. 2002. "Options for the explication of meaning in natural language". (a web-based publication).

Grabe, E., Nolan, F., Post, B. 2001. "Modelling intonational variation in English: The IViE system". In: S. Puppel, G. Demenko (eds). *Prosody 2000*. Poznań: Wydział Neofilologii UAM.

Gubrynowicz, R. 2002. "A Study of Speech Prosody of Subjects with Profound Hearing Loss Recorded at Child Age and 20 Years Later". *Speech Prosody* 2002.

Hirst, D., Di Cristo, A., Espesser, R. 2000. "Levels of representation and levels of analysis for the description of intonation systems". In: M. Horne (ed.). *Prosody and Experiment*. Kluwer Academic Press.

Jassem, W. 2002. "Classification and organization of data in intonation research". In: A. Braun, H. R. Masthoff (eds). *Phonetics and its Applications. Festschrift for Jens-Peter Koester*. Wiesbaden: Franz Steiner Verlag. 289 – 297.

Karpiński, M. 2002a. "The Corpus of Polish Intonational Database: Technical Specification". *Investigationes Linguisticae*, vol. VIII. Poznań: Institute of Linguistics, UAM.

Karpiński, M. 2002b. "Raport końcowy z projektu *Intonacyjna Baza Danych dla Języka Polskiego*". (Polish Intonational Database Project: Final Report). Submitted to KBN. Available from the author.

Karpiński, M., Kleśta, J., 2001. "The Project of Polish Intonation Database Intonational Database for the Polish Language". In: S. Puppel & G. Demenko (eds). *Prosody 2000*, Poznań, Faculty of Modern Languages and Literature, AMU.

Kowtko, J. C. 1996. *The function of intonation in Task-Oriented Dialogue*. Doctoral Thesis, University of Edinburgh.

Mann, W. 2002. "Dialogue Analysis for Diverse Situations". In: Bos, Foster & Matheson (eds). *Proceedings of the sixth workshop on the semantics and pragmatics of dialogue EDILOG 2002*, 4 – 6 September 2002, Edinburgh: UK.

Mertens, P. 2004. "The Prosogram: Semi-Automatic Transcription of Prosody Based on a Tonal Perception Model". *Proceedings of Prosody 2004 Conference*, Nara.

Pierrehumbert, J. 1980. *The phonetics and phonology of English intonation*. Doctoral dissertation, MIT.

Pierrehumbert, J., Hirschberg, J. 1990. "The meaning of intonational contours in the interpretation of discourse". In: P. Cohen, J. Morgan & M. Pollack (eds). *Intentions in communication*. Cambridge, Mass.: MIT Press. 271–311.

Post, B. 2000. *Tonal and Phrasal Structures in French Intonation*. Thesus: The Hague.

Selkirk, E. 1995. "Sentence Prosody: Intonation, Stress, and Phrasing". In: John A. Goldsmith (ed.). *The Handbook of Phonological Theory*. London: Blackwell. 550 – 569.
t'Hart, J., Collier, R., Cohen, A. 1990. *A Perceptual Study of Intonation. An Experimental-Phonetic Approach to Speech Melody*. Cambridge: CUP.
Traum, D. R. 2000. "20 Questions for Dialogue Act Taxonomies". *Journal of Semantics* 17(1). 7–30.
Volskaya, N., Skrelin, P. "Prosodic phrasing strategy from speaker's and listener's point of view". *Proceedings of the XVIIth International Congress of Linguistics*. Prague, July 24–29, 2003.
Wennerstrom, A. 2001. *The Music of Everyday Speech. Prosody and Discourse Analysis*. Oxford: Oxford University Press.
Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Basil Blackwell.

MELODIA RDZENNA W POLSKICH TEKSTACH QUASI-SPONTANICZNYCH I CZYTANYCH – CHARAKTERYSTYKA ZAWARTOŚCI INTONACYJNEJ BAZY DANYCH DLA JĘZYKA POLSKIEGO *PoInt*

Streszczenie

W artykule zaprezentowano wybrane wnioski dotyczące melodii rdzennej, oparte na analizie nagrań zgromadzonych w Intonacyjnej Bazie Danych dla Języka Polskiego *PoInt*. W pierwszej części tekstu przedstawiono wykorzystany korpus nagrań, wspomniano o pewnych problemach związanych z opisem intonacji oraz naszkicowano tło teoretyczne badań. Następnie omówiono wybrane relacje pomiędzy kategorią danej wypowiedzi, jej funkcją dyskursową i intonacją rdzenną.

REVIEW

Irena Sawicka, *An Outline of the Phonetic Typology of the Slavic Languages*. Toruń 2000, Wydawnictwo Uniwersytetu Mikołaja Kopernika, ss. 169, 22 mapy, 25 tabel, streszczenia w językach polskim, serbsko-chorwackim, rosyjskim i czeskim.

Tytuł recenzowanej książki wystarczająco uwypukla jej specyfikę. Nie brakuje w literaturze slawistycznej opracowań kontrastywnych z dziedziny fonologii i są też podejmowane próby sformułowania ogólnych zasad typologii fonologicznej języków słowiańskich. Znacznie mniej uwagi poświęca się w pracach slawistycznych o nieco szerszym zasięgu zagadnieniom fonetycznym, jeśli nie brać pod uwagę takiej klasycznej pracy, jak słynny *Grundriss der slavischen Phonetik* Olafa Brocha – książki cieszącej się zasłużoną sławą, ale opublikowanej już prawie sto lat temu. Wzmianka o pracy Brocha jest nieprzypadkowa: recenzowana tu książka Ireny Sawickiej przypomina ją i szerokością spojrzenia, i precyzją spostrzeżeń, i znakomitą ogólną orientacją w przedmiocie.

W recenzowanej książce aspekt fonologiczny jest również należycie uwzględniony, ale nacisk położony jest na aspekt fonetyczny, mianowicie na takie zjawiska, które tylko częściowo znajdują odbicie w opracowaniach fonologicznych, ponieważ są często niedystynktywne, ale za to warunkują specyfikę artykulacyjną i brzmieniową poszczególnych języków słowiańskich; zjawiska te są ponadto ważnym elementem wszelkich rozważań na temat tendencji rozwojowych tych języków, gdyż należą do tego, co określamy jako bazę artykulacyjną języka – zespół nawyków artykuacyjnych warunkujący mnóstwo szczegółowych zjawisk fonetycznych i wyznaczający ramy dla możliwych zmian w systemie fonetycznym.

Dyskusja nad typologią fonetyczną i fonologiczną języków słowiańskich obracała się dotychczas wokół bardzo ograniczonej liczby parametrów typologicznych, mogących stanowić tylko podstawę wybiórczej klasyfikacji, por. np. wysunięty przez Isaczenkę podział na języki mniej lub bardziej samogłoskowe czy spółgłoskowe. Autorka recenzowanej pracy wnosi do dyskusji wiele nowych parametrów, mających charakter wystarczająco ogólny, by można się było na ich podstawie pokusić o bardziej całościową klasyfikację fonetyki słowiańskiej (do możliwości typologii holistycznej, łączącej cechy fonetyczne z morfosyntaktycznymi, Autorka odnosi się sceptycznie). Te parametry to: akomodacyjny lub nieakomodacyjny typ wymowy, ogólny charakter struktury sylaby oraz prozodia słowa i frazy. Autorka przekonująco pokazuje elementy współzależności między wyodrębnianymi cechami, zwłaszcza między akomodacyjnym lub nieakomodacyjnym typem wymowy a strukturą sylaby i cechami prozodycznymi.

Rozdział I poświęcony jest znanemu już ze starszej literatury podziałowi na języki mniej lub bardziej spółgłoskowe. Autorka stara się uściślić kryteria podziału, zwracając uwagę, że ocena częstotliwości powinna się opierać nie tylko na udziale samogłosek i spółgłosek w inwentarzach fonemicznych, ale również na częstotliwości wystąpień tych fonemów w tekstach. Z drugiej strony