# Prosodically Aided Word Sense Disambiguation in Polish-English Speech Translation

Grzegorz Krynicki

Institute of Modern Languages and Literature,
Adam Mickiewicz University
al. Niepodległości 4, 61-874 Poznań
krynicki@wa.amu.edu.pl

ABSTRACT

This presentation reports on the influence that prosody of selected ambiguous Polish utterances may have on their interpretation and translation into English. A simple method for parametric description of the pitch curves coextensive with these utterances is discussed. On the basis of the pitch parameters obtained by the above method from a small speech corpus, the two-group classification of the ambiguous utterances was performed with respect to their interpretation and translation into English. The classification task was carried out by means of Statistical Discriminant Analysis. The classifier provided 82.5 to 97.5% correct classification rate depending on the disambiguated word. Perspectives for the application of the prosodically aided word sense disambiguation in Polish-English Speech Translation are suggested.

## 1. Introduction

Although prosody is not functionally and systematically related to the segmental level of Polish language, it has been observed that in casual speech some Polish ambiguous words show fairly consistent correlation between their prosodic features and their meaning ([1], p.131, c.f. [10], p. 61). This interdependence is most conspicuous for utterances whose pitch patterns may constitute simple and complete tone units. Among these words are exclamations (e.g. *aha*, *dosyć*), particles (e.g. *akurat*, *tak*) and adverbs (e.g. *dobrze*, *blisko*).

The implementation of this observation in the field of Polish-English Speech Translation requires that the choice of these words be additionally restricted by the constraint that at least two of their different meanings cannot be rendered by the same English equivalent. For example, it would be inappropriate to translate the Polish particle *akurat* as the English phrase *tell me another* when it was meant as an expression of satisfaction, as in the dialogue

*A*: O, przymierzałaś moją sukienkę. Jak leży?
(Oh, you've tried on my dress. Does it fit?)
*B*: Akurat! {=expression of satisfaction}
(Perfectly!)

and, conversely, we would not translate *akurat* that originally expressed disbelief as *perfectly*:

*A*: Słyszałeś, że był kiedyś w Stanach?
(Did you know that he's been in the States?)
*B*: Akurat! {=expression of disbelief}
(Tell me another!)

A similar phenomenon has also been reported for German discourse particles ([2], p.4) and for English utterance-initial particles ([3], p.128).

## 2. Empirical data

For the purpose of this study 5 polysemous expressions were selected: PROSZĘ ("Come in!", "Please, do."), AKURAT ("Tell me another!", "Perfectly!"), DOSYĆ ("Enough!", "So so."), NO NO ("Well, well!", "Don't be cheeky!"), DOBRZE ("All right.", "Correct."). **It was hypothesised that the strategies Polish native speakers employ to disambiguate these utterances in their speech are mainly and consistently restricted to the modification of fundamental frequency and the temporal arrangement of the pitch curve.** Intensity was not considered for technical reasons.

In order to verify this hypothesis, the words in question were presented in disambiguating contexts to a group of 40 native speakers of Polish: 20 male and 20 female university students of English, Polish and Economics. Each of them was asked to participate in an app. 3-min. recording session. During each session, 5 short dialogues were conducted between the experimenter and the subject on the basis of printed transcripts. Subject's part in every dialogue contained an ambiguous word in one of its two senses. It's meaning was determined by a disambiguating contexts. 20 informants out of 40 read the dialogues where the ambiguous words were presented in one set of disambiguating contexts, and the remaining 20 subjects contributed the dialogues with the same potentially ambiguous words but in different disambiguating contexts. In all, the corpus consisted of 200 (5×20 + 5×20) dialogues illustrating two senses of the ambiguous words. No speaker read two dialogues with the same potentially ambiguous word. The subjects were not informed of the purpose of the study until the elicitation procedure has been completed. The informants were instructed to be natural in their responses and relaxed.

When all the utterances had been recorded, their pitch tracks were extracted[1] and subjected to frequency-normalisation. The method for frequency-

[1] The pitch was computed by means of WinCECIL v2.2 ©Summer Institute of Linguistics 1994-97. All the tone contours were smoothed by the software-internal procedure. The parameters adopted for pitch extraction were the same for all contours: Voicing threshold = 40 Hz, minimal number of contiguous data points per string (frequency values given every 0,05 ms) = 6 items, percentage change of the string = 5%.

normalisation was basic (c.f. [4], p.5). On the basis of the sufficient number of frequency measurements contributed by a given speaker (more than 8000 frequency data points in each case, that is, approx. 200 data points per second, c.f. [5], p.173), his or her mean pitch and the voice range were computed. After recalculation of the $F_0$ parameter from normal to logarithmic scale, for all the utterances of a given person, the mean pitch of 0 and the range of the person's voice ± standard deviation was adopted.

The normalised pitch tracks were then parameterised. In the process of parameterisation, seven predictor variables were obtained for each pitch track. These predictor variables were later classified by means of Discriminant Analysis. It was postulated that if such a classifier can be trained to correctly disambiguate between the senses of the utterances it "hears" on the basis of their $F_0$ and temporal features, the hypothesis outlined above would be substantiated.

## 3. Discriminant Analysis

Discriminant Analysis is a multivariate statistical procedure designed to find such a combination of the predictor (independent) variables and their coefficients that will maximise the distinction between the groups. This combination is formulated as a Discriminant Function. In this study 2 groups were considered, each containing parameters calculated for the pitch pattern of the utterances in each of its meanings. The two groups were represented by the corresponding 2 grouping (dependent) variables. On the basis of the pitch parameters calculated for a pitch track, the Discriminant Analysis was used to predict which group the given pitch track belongs to. Discriminant Analysis was conducted in STATISTICA™ 5.0 PL[2].

Out of several constraints that the data must comply with for Discriminant Analysis to be performed ([6], p.22, [7], pp.125nn, [8], p.9), minor violations were observed for the requirement that the within-group variabilities must be approximately equal and the requirement that each of the groups must be normally distributed. It is believed, however, that minor violations of both conditions are not fatal to the results of the multivariate analyses ([6], p.22, c.f. [9], p.20).

The 7 predictor variables used in the analysis were:

1. VECTOR SIZE (LENGTH) - the number of data points in the pitch track from the first non-zero frequency measurement to the last non-zero frequency measurement inclusive. The zero frequency measurements that separated continuous frequency strings were counted in the computation of the size of the vector. However, neither the actual zero frequency values nor the interpolated values were included in the computation of the pitch track parameters;

[2] Statistica, version 5.5 A, Licence owner: Adam Mickiewicz University, Poznań; serial number: AXXP9088529404AR53

2. MAXIMUM F0 (MAX_F0) - maximum frequency value in the normalised pitch track;
3. TIME ARGUMENT OF MAX. F0 (TMAX_F0) - time point at which the normalised frequency string assumes its maximum value;
4. MINIMUM F0 (MIN_F0) - minimum frequency value in the normalised pitch track;
5. TIME ARGUMENT OF MIN. F0 (TMIN_F0) - time point at which the normalised frequency string assumes its minimum value;
6. MEAN VALUE OF PITCH TRACK (MEAN) - mean value of the pitch track produced by an individual speaker. Normalised with respect to the mean pitch of the speaker, as obtained from all the utterances of a given speaker;
7. STANDARD DEVIATION OF PITCH TRACK (STD_DEV) - standard deviation of the pitch track produced by an individual speaker. Normalised with respect to the variance of all the utterances produced by a given speaker.

## 4. Results of the classification

### 4.1. Percentages of correct classifications for each word

The classification was performed by means of **leave-one-out** method which consists in: training the classifier on all the training set observations except one test observation; recording the result of the classification for the test observation; iterating the above two steps until all the observations have been used exactly once as the test observation; averaging all the classification results to obtain the final classification result.

The final classification results for all the words are tabulated below.

| **Polish expression** | **English Translations** | | **Relative frequency of correctly classified pitch tracks (in %)** |
|---|---|---|---|
| PROSZĘ | Come in! | Please, do. | 87.5 |
| AKURAT | Tell me another! | Perfectly! | 95 |
| DOSYĆ | Enough! | So so. | 82.5 |
| NO NO | Don't be cheeky! | Well, well! | 82.5 |
| DOBRZE | All right. | Correct. | 92.5 |
| **Average of correct classification percentages** | | | 88.5 |

**Table 2. The results of the Discriminant Analysis classification of pitch patterns coextensive with ambiguous Polish expressions.**

These results lend a tentative support to the claim that prosody of some Polish words may be fairly strongly correlated to their interpretation and

rendition into English. Hence, it is proposed that prosodic information about these words should be taken into consideration in Polish-English Speech Translation systems when deciding about the choice of the best English equivalent for these words.

### 4.2. Relative importance of different prosodic features in the process of disambiguation

An important feature of the Discriminant Analysis is its ability to formulate standardised ***discriminant function coefficients***. From their magnitude conclusions can be drawn about how the individual pitch track parameters are being used by native speakers to discriminate among the meaning groups. This information, apart from the linguistic and psycholinguistic insights it may provide, should be particularly useful in the case when not all potentially significant predictor variables can be considered in the computation of the best English equivalent of a Polish utterance (e.g. for the sake of computational efficiency, which is crucial in real-time Speech Translation systems). The hierarchy of predictor variables with respect to their discriminability would allow the researcher to disregard the variables that have least discriminating power if the computations are too time consuming.

The table below shows the standardised coefficients of the discriminant function used to discriminate amongst the different levels of grouping for *proszę* utterance

| | |
|---|---|
| LENGTH | 1.831926 |
| TMAX_F0 | -0.50837 |
| MAX_F0 | -0.47303 |
| TMIN_F0 | -0.72158 |
| MIN_F0 | 0.370581 |
| MEAN | 0.178785 |
| STDDEV | 0.203784 |
| Cumulated value | 1 |

**Table 3. Standardised discriminant function coefficients for grouping (*proszę* utterance).**

Thus, the standardised Discriminant Function is formulated as follows

DF = 1.831926*LENGTH - 0.50837*T_MAX_F0 - 0.47303*MAX_F0 - 0.72158*T_MIN_F0 + 0.370581*MIN_F0 + 0.178785*MEAN + 0.203784*STD_DEV

DF values obtained by means of this linear equation allow the formulation of different discriminant variables ('roots') depending on the configuration of the pitch parameters included as variables of this equation.

The stepwise Discriminant Analysis with a backward selection of the variables shows **the relative importance of the pitch track parameters for**

**the disambiguation of the** *proszę* **utterances**. In the backward, stepwise Discriminant Analysis, the variables are removed from the model one by one from the ones of the lowest discriminating power to the ones of the greatest discriminating power. The sooner the variable is removed, the smaller discriminatory power it has. In Table 4, the first column contains the information about how many variables have been removed at a given step of the analysis. In the first step all the variables are included (0 removed), which also means the discriminability of the Discriminant Function DF is as high as the procedure allows (87.5 %).

| Step, nr of removed variables | name of the variable removed | p-level | nr of variables present | Wilks' Lambda | F | % of correct classifications |
|---|---|---|---|---|---|---|
| 0 | | 0.0001 | 7 | 0.4245 | 6.1988 | 87.5 |
| 1 | MEAN | 0.8536 | 6 | 0.4249 | 7.4440 | 87.5 |
| 2 | STDDEV | 0.6942 | 5 | 0.4269 | 9.1276 | 87.5 |
| 3 | MAX_F0 | 0.5025 | 4 | 0.4326 | 11.471 | 87.5 |
| 4 | MIN_F0 | 0.3304 | 3 | 0.4447 | 14.981 | 85.0 |
| 5 | TMIN_F0 | 0.0779 | 2 | 0.4854 | 19.611 | 82.5 |
| 6 | TMAX_F0 | 0.2532 | 1 | 0.5030 | 37.532 | 82.5 |

**Table 5. Wilks' Lambda vs. % of correct classifications for the pitch parameters in the stepwise Discriminant Analysis for the word *proszę*. LENGTH (not included in the table) turns out to have the least discriminating power.**

Similar analysis was conducted for the remaining 4 words: *akurat, dosyć*, *dobrze*, *no no*. In order to find a relative importance of the prosodic features for establishing one of the senses of any ambiguous word, a general measure of the discriminating power was constructed for all the 7 pitch parameters. The measure of relative discriminating power was calculated as the arithmetic mean of the Wilks' Lambdas obtained in the backward, stepwise Discriminant Analysis for a given pitch parameter across five different words. In this way we obtain the total relative discriminating power of the pitch parameters irrespective of the word they were derived from:

| | PROSZĘ | AKURAT | DOSYĆ | NO NO | DOBRZE | Mean λ | Rank |
|---|---|---|---|---|---|---|---|
| LENGTH | 1 | 0.23 | 0.78 | 0.61 | 1 | 0.72 | 1 |
| MEAN | 0.5 | 1 | 1 | 0.47 | 0.39 | 0.67 | 2 |
| STDDEV | 0.47 | 0.47 | 0.74 | 0.47 | 1 | 0.63 | 3 |
| MAX_F0 | 0.49 | 0.4 | 0.62 | 1 | 0.39 | 0.58 | 4 |
| TMAX_F0 | 0.49 | 0.23 | 0.87 | 0.45 | 0.74 | 0.56 | 5 |
| TMIN_F0 | 0.49 | 0.27 | 0.63 | 0.5 | 0.43 | 0.46 | 6 |
| MIN_F0 | 0.49 | 0.23 | 0.71 | 0.45 | 0.39 | 0.45 | 7 |

**Table 6. Discriminant power of all the 7 pitch parameters, expressed by Mean Wilks' Lambda across five words. The higher Mean λ, the more discriminating power a parameters has.**

From the above calculation we obtain a general measure of discriminability for different parameters across the ambiguous words. The *length* of the pitch track has the greatest discriminating power (Mean $\lambda = 0.67$). The least important for the disambiguation of the utterances we analysed is the *miminum frequency* of the pitch track (Mean $\lambda = 0.45$). So, for example, if we were to choose only two pitch parameters from the above list in order to improve the correct disambiguation rate in our Speech Translation system, we would choose the *length* of the pitch pattern that accompanies the ambiguous utterance to be disambiguated and the *mean* of all its frequency data points as these two parameters rank highest with respect to their discriminability.

## 5. Conclusions

It has been shown that prosody may help to disambiguate some Polish utterances and therefore may prove useful in Polish-English Speech Translation. The ranking of the prosodic features according to their discriminating power has been established and its applicability in the design of Speech Translation Systems has been acknowledged.

REFERENCES

[1] Krynicki G., (1999) *Suggested Improvements in the Linguistic Aspect of the Electronic Polish-English Dictionary*, in: "Speech and Language Technology. Volume 3", Poznań.

[2] Stede. M., Schmitz B., *Discourse Particles and Routine formulas in Spoken Language Translation*, in: 1999 Speech Translation Summit Proceedings, Barcelona. pp. 3-9.

[3] Byron D., Heeman P., (1997) *Discourse Marker Use in Task-Oriented Spoken Dialogue*, in: Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech), Rhodes, cited after [2].

[4] Jassem W., Demenko G., M. Krzyśko (1988) *Klasyfikacja podstawowych wzorców intonacyjnych z zastosowaniem funkcji dyskryminacyjnych*, IPPT PAN, Warszawa.

[5] Jassem, W., (1984) *Modern English Phonology*, PWN, Warszawa.

[6] Jassem, W.,(1998) *An Acoustical Linear-Predictive and Statistical Discriminant Analysis of Polish Fricatives and Affricates*, in: "Speech and Language Technology. Volume 2", Poznań.

[7] Chatfield Ch., Collins A., (1980) *Introduction to Multivariate Analysis*, Chapman & Hall, London.

[8] Krzyśko M., (1982) *Analiza Dyskryminacyjna*, WN-UAM, Poznań.

[9] Lew, R., (2000) *Sandhi Voicing of Polish Learners of English*, post-doctoral dissertation,URL: http://main.amu.edu.pl/~rlew/voicont.htm

[10] Demenko, G. (1999) Analiza cech suprasegmentalnych języka polskiego na potrzeby technologii mowy, WN-UAM, Poznań.