Parameterization of prosodic headedness

Uwe D. Reichel[1], *Katalin Mády*[2], *Štefan Beňuš*[3]

[1]Institute of Phonetics and Speech Processing, University of Munich
[2]Hungarian Academy of Sciences, Budapest, Hungary
[3]Constantine the Philosopher University, Nitra & II SAS Bratislava, Slovakia

reichelu@phonetik.uni-muenchen.de, mady@nytud.hu, sb513@nyu.edu

Abstract

Prosodic headedness generally refers to the location of relevant prosodic events at the left or right end of prosodic constituents. In a bottom-up procedure based on a computational F0 stylization we tested several measures to quantify headedness in parametrical and categorical terms for intonation in the accentual phrase (AP) domain. These measures refer to F0 level and range trends as well as to F0 contour patterns within APs. We tested the suitability of this framework for Hungarian and French known to be left- and right-headed, respectively, and applied it to Slovak whose headedness status is yet less clear. The prosodic differences of Hungarian and French were well captured by several of the proposed parameters, so that from their values for Slovak it can be concluded that Slovak tends to be a left-headed language.

Index Terms: headedness, accentual phrase, intonation stylization, Hungarian, French, Slovak

1. Introduction

Left- or right-headedness of a language refers to its characteristic to stress words always on the first or last syllable as well as to the tendency to locate relevant prosodic events at the left or right end of prosodic constituents. Headedness has been reported for many languages, among them French which is classified as strictly right-headed [1, 2], and Hungarian classified as strictly left-headed [3, 4, 5, 6, 7].

Both languages have fixed lexical stress defined as the potential location of phrase-level accents [8]. Stress is word initial in Hungarian [5] and word-final in French [9]. Both leftmost and rightmost location in Hungarian and French, respectively, remain when combining words to prosodic phrases, which is e.g. formalized in [10]'s right-heading principle for French. The headedness of Slovak is less clear. Like Hungarian it has word-initial stress, but on the phrasal level the accent location is only predominantly but not always phrase-initial. Among the exceptions are polar questions accented on the right.

The focus of this study is on the headedness characteristics of the prosodic constituent *accentual phrase* (*AP*). Following the definition of [11] an AP can obtain only one pitch accent and it shows some language-specific regularity in its pitch shape. In left- and right-headed languages the pitch accent has not only a prominence-lending function but due to its fixed position it serves also as a boundary marker [12, 11]. The other end of the AP may be marked by a non prominence-lending boundary tone, too [11]. Leaving aside the terminological confusion (see [13] for an exhaustive list of more or less AP equivalent names) for French there is established agreement on the AP to be a prosodic constituent on its own. In recent studies on accent groups (AG) evidence was also found both for Hungarian and Slovak that these AGs actually are APs. AGs are defined as rhythmic units that in left-headed languages consist of an accented syllable and all following unaccented syllables till the next accented one. Since the AGs in Hungarian and Slovak fulfill the AP criterion of pitch shape regularity [14, 15, 16] they can be considered as APs in the sense of [11]. This regularity refers to pitch shape within the AG for both languages [16] as well as to consistent deviation patterns from higher-level prosodic units in Hungarian [14, 15].

Due to the congruence of accent position and pitch contour it can be safely stated that headedness has a crucial influence on the pitch shape within APs. Thus AP contours differ between right-headed French on the one side and left-headed Hungarian on the other side and they should be indicative for the headedness of Slovak. This study proposes a framework to quantify these differences in a data-driven bottom-up way in parametric and categorical terms by means of computational intonation stylization.

2. Data and preprocessing

Data. For the Hungarian and Slovak data we randomly selected 150 intonational phrases (IP), respectively, containing about 440 manually segmented APs from corpora of collaborative dialogues [16, 15]. For both languages five female (mean age 28 for Hungarian, 36 for Slovak) and five male speakers (mean age 27 for Hungarian, 25 for Slovak) had been recorded. The French data was obtained from the Rhapsodie corpus [17, 18] containing dialog and monologue data that is segmented and annotated phonetically, prosodically, and syntactically. From this corpus we extracted a random sample of the same size as the Hungarian and Slovak data from the spontaneous-speech dialog part (each 15 female and male speakers, mean age 44 and 43, respectively). In the French annotation a different terminology is used for the prosodic units in question. We thus defined according to the specifications given in [18] the constituent type "*rhythmic group*" as accentual phrases and the next-higher level type "*intonational packages*" as intonational phrases. This equivalent treatment of rhythmic groups and accentual phrases is compliant to the AP terminology of [9] (*"rhythmic unit"*, and [19] (*groupe rhythmique*).

Preprocessing. Fundamental frequency (F0) was extracted by autocorrelation (Praat 5.3, sample rate 100 Hz). Voiceless utterance parts and F0 outliers were bridged by linear interpolation. The contour was then smoothed by Savitzky Golay filtering [20] using third order polynomials in 5 sample windows and transformed to semitones relative to a base value. This base value was set to the F0 median below the 5th percentile of an

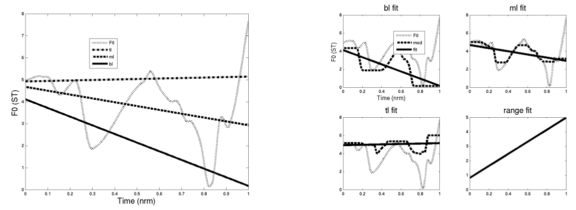

Figure 1: **A (left):** *Stylization of base-, mid- and topline based on F0 median sequences of different percentiles. Stylization of F0 range change.* **B (right):** *Base-, mid-, topline and linear range stylization results.*

utterance and served to normalize F0 with respect to its overall level.

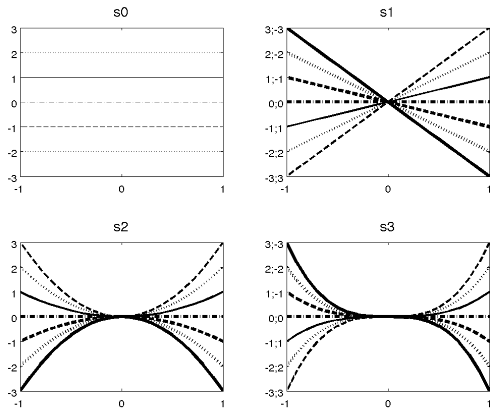

Figure 2: *Influence of each coefficient of the third order polynomial $t = \sum_i s_i \cdot t^i$ on the contour shape. All other coefficients set to 0. For compactness purpose on the y-axis both function and coefficient values are shown if they differ.*

3. Stylization

3.1. Level and range trends in accentual phrases

To capture the F0 register in terms of its level and range [21] we fitted a base-, a mid-, and a topline within the APs as described in more detail in [22]. The midline represents the F0 level, whereas the base- and topline provide the F0 range information. The robust fitting procedure that does not need to rely on fuzzy local peak and valley detection is illustrated in Figure 1 and consists of the following steps:

- A window with a length of 200 ms is shifted along the F0 vector with a step size of 10 ms;
- within each window three F0 medians are calculated: one for the baseline based on the values below the 10th percentile, one for the topline based on the values above the 90th percentile, and one for the midline based on all the values;
- linear polynomials are fitted through each of the three resulting median sequences.
- Furthermore, a linear polynomial is fitted through the pointwise distances between base- and topline.

The level trend within an AP is defined here as the midline slope. The range trend is defined as the slope of the regression through the pointwise distances between top- and baseline. The linear level and range stylizations are shown in the right half of Figure 1B.

Expectations: In left-headed Hungarian we expect a higher amount of falling level and converging range, and thus more negative slope values. For right-headed French we expect the opposite trend.

3.2. F0 patterns in accentual phrases

For parameterizing F0 patterns within AGs we adopted the CoPaSul approach of [23], that is, we fitted a baseline through the intonation phrase (IP) analogously to the method described in section 3.1 and subtracted this IP baseline from the F0 contour. Then we fitted a third-order polynomial to the residual contour within each AP. In order to compare the parameters across different AP lengths, the time was normalized to the interval -1 to 1. Figure 2 shows the F0 shape in dependence of systematic coefficient variation.

Expectations: Following the illustration in Figure 2, we expect s_1 and s_3 to be lower for left-headed Hungarian compared to right-headed French.

3.3. Categorical differences

In order to derive AP contour classes as in [23] the polynomial coefficient vectors were clustered by k-means [24]. As described in [23] in more detail the optimal number of clusters was derived by means of subtractive clustering [25], whose parameters were optimized by the Nelder-Mead [26] method. In [23] this way of determining initial cluster centers turned out to yield stable clustering results on disjunct data subsets. The resulting contour classes are shown in Figure 3. For each language we calculated the probability for each contour class, and between each language pair we measured the distance between the language-dependent class probability distributions P and Q in terms of their information radius $R(P||Q)$ [27], which is given as follows:

$$R(P||Q) \quad = \quad \frac{D(P||\frac{P+Q}{2}) + D(Q||\frac{P+Q}{2})}{2} \qquad (1)$$

This distance measure fulfilling the symmetry criterion is a symmetric version of the Kullback-Leibler divergence $D(p||q)$, which is:

$$D(P||Q) \quad = \quad \sum_{c \in C} P(c) \log_2 \frac{P(c)}{Q(c)} \qquad (2)$$

Thus $R(P||Q)$ quantifies the difference between the probabilities P and Q of the contour classes c in language pairs. The class occurrence variability is measured for each language in terms of the class unigram entropy $H(C)$:

$$H(C) = -\sum_{c \in C} P(c) \log_2 P(c) \qquad (3)$$

where C denotes the set of contour classes c derived by clustering.

Expectations: Referring to Figure 3 left-headed Hungarian should show a preference for left-headed classes, mainly c_2 and c_4, while in right-headed French the probabilities of the right-headed classes, mainly c_3 and c_6 should be higher.

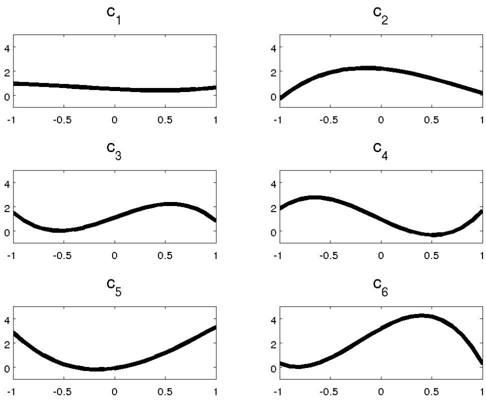

Figure 3: *F0 contour classes in accentual phrases.*

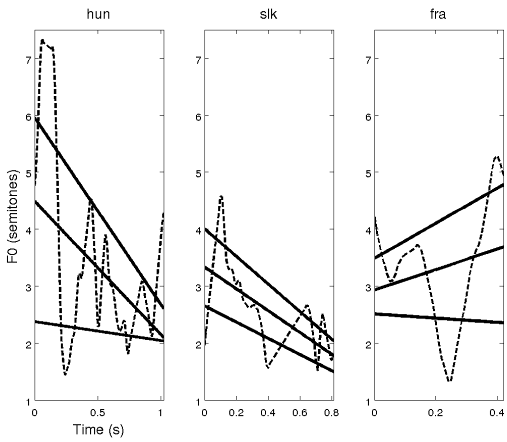

Figure 4: *Observed typical F0 trends for Hungarian, Slovak, and French. Hungarian, Slovak: falling level (midline slopes are -0.0234 and -0.0192, respectively), converging range (range slopes are -0.0295 and -0.0100). French: rising level (midline slope 0.0180), diverging range (range slope 0.0346).*

4. Results

4.1. Parameter comparisons

In Figures 4 and 5 observed typical cases for trend slope and polynomial contour shape are shown.

The language-dependent trend and shape coefficient values are presented in Figures 6 and 7, respectively. For all parameter comparisons the Kruskal-Wallis tests and Scheffé post-hoc test ($\alpha = 0.05$) have been applied.

With respect to level trend all three languages differ significantly from each other ($p < 0.001$, $\chi^2_2 = 141.54$; Figure 6). In line with headedness Hungarian shows a negative median midline slope -0.0345, while the median slope 0.0152 for French is positive. For Slovak like for Hungarian the slope -0.0121 is negative.

The same holds for the range ($p = 0.003$, $\chi^2_2 = 28.44$; Figure 6). Again, range becomes smaller towards the end of the AP for Hungarian indicated by the negative slope -0.0074, while in French range increases by the median slope 0.0031. Again the Slovak results (negative slope -0.0028) match to Hungarian more closely.

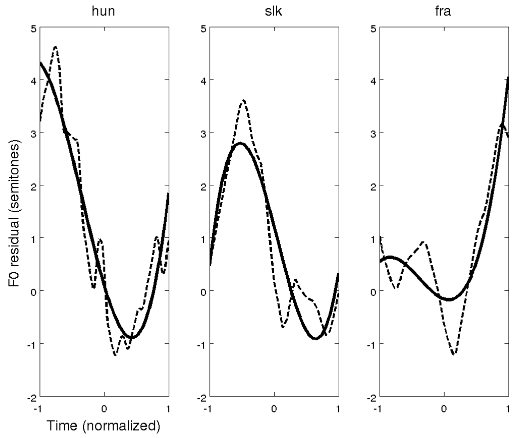

Figure 5: *Observed typical AP contour shape patterns for Hungarian, Slovak, and French. For Hungarian s_1=-4.1382, s_2=2.9836, s_3=2.9040. For Slovak s_1=-4.7202, s_2=-0.8223, s_3=4.6466. For French s_1=-0.3958, s_2=2.4574, s_3=2.1487*

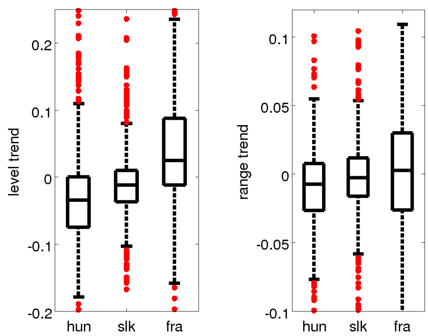

Figure 6: *Level (left) and range (right) trends within an AP.*

We found significant differences for all polynomial coefficients between the languages (Figure 5). The offset coefficient s_0 was highest for French, followed by Slovak, and Hungarian ($p < 0.001$, $\chi^2_2 = 41.26$) . In line with the level trend results, the linear coefficient s_1 was clearly highest for French, again followed by Slovak, and Hungarian ($p < 0.001$, $\chi^2_2 = 79.35$). For the quadratic coefficient s_2 again all languages differed significantly ($p < 0.001$, $\chi^2_2 = 85.10$), but the median value was only negative for Slovak showing its tendency for concave (rising-falling) contours. With respect to the third order coefficient s_3 French differed significantly from the other two languages ($p < 0.001$, $\chi^2_2 = 74.85$) yielding the smallest median value.

4.2. Class comparisons

As can be seen in Table 1 the AP contour class probability distributions are more similar for Hungarian and Slovak (0.13) than between these two languages and French (0.15 and 0.17, respectively). Table 2 gives a more detailed description about the language-related probability of each of the classes shown in Figure 3. Overall, with respect to F0 movement most neutral

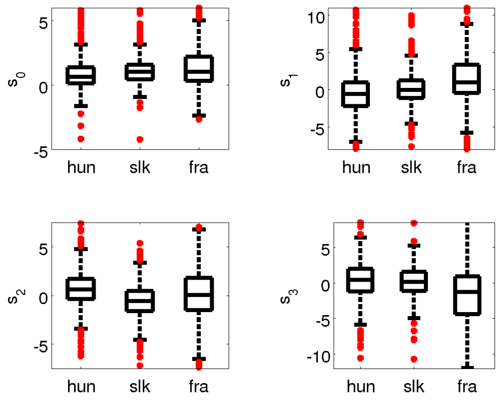

Figure 7: *Values for the offset s_0, linear s_1, quadratic s_2 and the third-order s_3 polynomial coefficients for each language.*

class c_1 occurs with the highest probability. We further categorized each class into *left-* and *right-headed* the following way: by the corresponding polynomial coefficient centroid vector we mapped normalized time to an F0 contour and calculated the mean values of the first and second half of this contour, respectively. The contour classes c_1, c_2, and c_4, for which the first half's mean was higher were classified as *left-* the others, i.e. c_3, c_5, and c_6 as *right-headed*. The last two columns of Table 2 show the tendencies of Hungarian and Slovak for left-weighted classes, and of French for right-headed classes. The contour classes probability distribution is most uniform in French which is reflected in the highest class unigram entropy value (2.55) followed by Hungarian (2.30) and Slovak (2.16).

Table 1: *Information radii between language-related AP contour class probability distributions.*

	slk	fra
hun	0.13	0.15
slk		0.17

Table 2: *AP contour class probabilities for each language. For $\sum_l$ and $\sum_r$ the probabilities of left-headed (c_1, c_2, c_4) and right-headed (c_3, c_5, c_6) classes are summed up, respectively.*

	c_1	c_2	c_3	c_4	c_5	c_6	$\sum_l$	$\sum_r$
hun	0.40	0.11	0.08	0.21	0.13	0.08	0.72	0.28
slk	0.42	0.25	0.14	0.08	0.05	0.05	0.76	0.24
fra	0.20	0.15	0.23	0.11	0.15	0.16	0.47	0.53

5. Discussion

5.1. Headedness feature validation

For the quantification of headedness within APs we proposed (1) linear trend measures, (2) polynomial coefficients for contour shape stylization, and (3) a contour clustering for a shape comparison in categorical terms. In the following the suitability of each of these measure groups will be discussed.

Level and range trends. The given results confirm our expectations that midline slope and the slope of the pointwise distance between top and baseline do indeed reflect headedness within APs. Hungarian APs have both negative level and range slopes while in French APs level and range rise.

Polynomial contour shape stylization. The results for the suitability of the polynomial coefficients are less clear-cut. In line with the findings on level trend, the first order coefficient s_1 behaved like the trend slope parameters.

The differences of the 2nd order coefficient s_2 revealed a qualitative shape difference already described for Hungarian and Slovak in [16]. The difference is illustrated in Figure 5 by observed typical contour examples. Whereas in Hungarian the F0 maximum tends to occur at the leftmost position of the AP, in Slovak the maximum is shifted a bit rightwards. French seems to behave mirror-inverted to Hungarian with the maximum F0 at the rightmost position. Thus Hungarian and French APs tend to have more convex (falling-rising) shapes as opposed to the concave (rising-falling) shape of Slovak.

The third order coefficient s_3 behaves opposite to the expectations formulated in section 3.2. As illustrated in Figure 2 s_3 like s_1 contributes to rising F0 from a low start level by positive values and contributes to falling F0 from a high start level by negative values. However, for French a negative s_3 median -1.26 was observed indicating falling F0. A plausible explanation for this seemingly counter-intuitive finding is, that the overall F0 movement is determined mainly by s_1, while s_3 superimposes movements with finer granularity. Among these events on a finer granularity scale are French AP initial high boundary tones described by [13, 11]. It is likely, that the polynomials capture these initial F0 movements with low s_3 values.

Thus taken together, among the polynomial coefficients, only s_1 but not s_2 and s_3 turned out to be suitable for headedness quantification.

Contour classes. For contour classes the expectations were met. The contour class probability observed for a language corresponds well to its headedness. Furthermore, the class probability distributions are more similar between the examined left-headed languages as compared to the right-headed language. The concave shape in Slovak [16] is again well reflected in the preference of contour class 2. In line with the larger coefficient ranges to be seen in the French boxplots in Figures 6 and 7, also the higher contour class unigram entropy (cf. section 4.2) indicates that French intonation is more variable than Hungarian and Slovak speech melody.

5.2. Headedness classification

After testing the suitability of the proposed measures, we applied the suitable trend and linear polynomial coefficient measures and the class probability similarities to determine the headedness of Slovak. In comparison to the two anchors Hungarian and French, Slovak behaves more like Hungarian with respect to all features: like Hungarian it shows negative level and range trends, and it has a more similar contour class distribution to Hungarian than to French, which is expressed by the information radius values 0.13 and 0.17, respectively.

In theory-driven approaches [28, 29] headedness judgments are often controversial and not yet solved (see [30] for an overview). Thus the proposed bottom-up approach might contribute easy to obtain data-driven evidences to the ongoing discussions.

Acknowledgements. The conference stay of the second author was funded by the Alexander von Humboldt Foundation.

6. References

[1] A. Di Cristo, "Intonation in French," in *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press, 1999, pp. 195–218.

[2] S.-A. Jun and C. Fougeron, "The accentual phrase and the prosodic structure of French," in *Proc. ICPhS*, Stockholm, Sweden, 1995, pp. 722–725.

[3] K. É. Kiss, *The Syntax of Hungarian*. Cambridge University Press, 2002.

[4] L. Hunyadi, *Hungarian sentence prosody and universal grammar: on the phonology-syntax interface*. Frankfurt am Main: Lang, 2002.

[5] L. Varga, *Intonation and Stress: Evidence from Hungarian*. Hampshire, New York: Palgrave Macmillan, 2002.

[6] D. Ladd, *Intonational Phonology*, 2nd ed. Cambridge: Cambridge University Press, 2008.

[7] K. Mády, A. Szalontai, A. Deme, and B. Surányi, "On the interdepencende of prosodic phrasing and prosodic prominence in Hungarian," in *Proc. 11th International Conference on the Structure of Hungarian*, Piliscsaba, Hungary, 2013.

[8] T. Vennemann, *Neuere Entwicklungen in der Phonologie*. Berlin: Mouton de Gruyter, 1986.

[9] A. Di Cristo and D. Hirst, "Rythme syllabique, rythme mélodique et représentation hierarchique de la prosodie du français," in *Travaux de l'Institut de Phonétique d'Aix*, 1993, p. 924.

[10] P. Garde, *L'Accent*. Paris: Press Universitaires de France, 1968.

[11] S.-A. Jun and J. Fletcher, "Methodology of studying intonation: From data collection to data analysis," in *Prosodic Typology II: The phonology of intonation and phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2014, pp. 520–539.

[12] J. Vaissière, "Rhythm, accentuation and final lengthening in French," in *Music, Language, Speech and Brain*, J. Sundberg, L. Nord, and R. Carlson, Eds. London: Macmillan Press, 1991, pp. 108–120.

[13] S.-A. Jun and C. Fougeron, "Realizations of accentual phrase in French," in *Probus*, 2002, vol. 14, pp. 147–172.

[14] K. Mády, U. Reichel, and Š. Beňuš, "Accentual phrase in languages with fixed word stress: A study on Hungarian and Slovak (Extended Abstract)," in *Workshop Advancing Prosodic Transcription for Spoken Language Science and Technology II, Phonetics and Phonology in Iberia 2013*, Lisbon, Portugal, 2013.

[15] ——, "Accentual phrases in Slovak and Hungarian," in *Proc. Speech Prosody*, Dublin, Ireland, 2014, pp. 752–756.

[16] Š. Beňuš, U. Reichel, and K. Mády, "Modelling accentual phrase intonation in Slovak and Hungarian," in *Complex Visibles Out There*. Olomouc, Czech Republic: Palacký University, 2014, vol. 4, pp. 677–689.

[17] "ANR Rhapsodie 07 Corp-030-01, Corpus prosodique de référence du franais parlé," http:www.projet-rhapsodie.fr, June 24th 2014, version 1.0.

[18] A. Lacheret, S. Kahane, J. Beliao, A. Dister, K. Gerdes, J.-P. Goldman, N. Obin, P. Pietrandrea, and A. Tchobanov, "Rhapsodie: A prosodic-syntactic treebank for spoken French," in *Proc. LREC*, Reykjavik, Iceland, 2014, pp. paper hal–00 968 959.

[19] E. Delais-Roussarie, "Pour une approche parallèle de la structure prosodique," Ph.D. dissertation, Universit Toulouse le Mirail, 1995.

[20] A. Savitzky and M. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[21] T. Rietveld and P. Vermillion, "Cues for Perceived Pitch Register," *Phonetica*, vol. 60, pp. 261–272, 2003.

[22] U. Reichel and K. Mády, "Comparing parameterizations of pitch register and its discontinuities at prosodic boundaries for Hungarian," in *Proc. Interspeech 2014*, Singapore, 2014, pp. 111–115.

[23] U. Reichel, "Linking bottom-up intonation stylization to discourse structure," *Computer, Speech, and Language*, vol. 28, pp. 1340–1365, 2014.

[24] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. of 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1967, pp. 281–297.

[25] S. Chiu, "Fuzzy model identification based on cluster estimation," *J. Intelligence & Fuzzy Systems*, vol. 2, no. 3, pp. 267–278, 1994.

[26] J. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308–313, 1965.

[27] H. Schütze and C. Manning, *Foundations of Statistical Natural Language Processing*. Cambridge, Mass.: MIT Press, 1999.

[28] M. Halle and J.-R. Vergnaud, *An Essay on Stress*. Cambridge MA: MIT Press, 1987.

[29] G. Cinque, "A null theory of phrase and compound stress," *Linguistic Inquiry*, vol. 24, pp. 239–297, 1993.

[30] D. Ladd, "Intonation," in *Language Typology and Language Universals: An International Handbook*, M. Haspelmath, E. König, W. Oesterreicher, and W. Raible, Eds., 2001, pp. 1380–1390.