Prosodic Chunking of German as a Foreign Language

Hansjörg Mixdorff, Hamurabi Gamboa Rosales

1 Department of Computer Science, Beuth University Berlin, Germany
2 Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Mexico

Abstract

This study concerns the perception of boundaries and accented syllables by native German subjects as compared to foreign non-speakers and learners of the language at different proficiency levels. To this effect six-syllable sequences excised from a context of three poly-syllabic words of German were presented to participants who had to select the syllables they perceived as accented, as well as the locations of word boundaries. Results show that German native subjects perform well at the word boundary task, but mark correctly less than two thirds of accented syllables. Chinese and Mexican non-learners still detect a considerable number of word boundaries and accented syllables. Learners of German show improvement at the task with growing experience though they often pick legal subword units that do not necessarily form a plausible sequence. Correlation analysis of factors for syllable and boundary selection performed for non-learners and German subjects – as expected – shows considerably different behaviours. Whereas the boundary location does not influence the Germans’ decision on the accent location, Chinese and Mexican non-learners show a preference to mark an accent when the syllable is followed by a word boundary. We also found that the acoustic properties of the syllables had a larger impact on the non-learners’ decisions since they could not operate on linguistic knowledge of German.

Index Terms: Prominence, accent and boundary perception, L2 learning

1. Introduction

It is a well-known fact that non-speakers and native speakers of a language perceive and process stimuli from that language quite differently. When a person studies a foreign language (L2) (s)he makes a transition between the two states as (s)he acquires a growing competence as to L2 linguistic structures and vocabulary. An important competence for communication is the ability to process and chunk the speech stream into meaningful units. The segmentation of an utterance into words depends heavily on prosodic cues (see, for instance, [1]) such as F0 and duration – features which might be employed quite differently in the L2 – as well as the growing L2 lexicon on the part of the learner.

Lexical stress is a property of each poly-syllabic word and in German – in contrast to certain other languages – its location in the word is rather flexible, i.e. not predefined by default [2]. Hence three-syllable words, for instance, can exhibit lexical stress either on the first, second or third syllable. There exists a small group of words which are segmentally identical, but differ as to the lexical stress location (set in bold face): compare, for instance, ‘um-fah-ren’ (to go around) vs. ‘um-fah-ren’ (to run over). In the context of an utterance the lexically stressed syllables become potential loci of accentuation, usually associated with prosodic cues such as F0 transitions and lengthening. Therefore the learner is required to memorize this feature for each word and decode it from the speech stream.

This study elaborates on an initial experiment reported in [3], which concerned the perception of boundaries and accented syllables by native German subjects as compared to Chinese non-speakers and learners of the language at different proficiency levels. We aimed to investigate how subjects performed on a task for which they either had to rely – to varying degrees – on their linguistic knowledge as well as the acoustic properties of the stimuli.

Furthermore, we explored the perceptual interrelationship between boundaries and accented syllables. This work was originally inspired by Gilbert et al. [5] who showed in a learning experiment that speech chunking in French is performed in rhythm groups. However, whereas French words exhibits a default prominence on the ultimate syllable of a word, next to the word boundary, the situation in German as explained above is quite different.

The current study expands the work in [3] three-fold: (1) We perform an acoustic prosodic analysis of the stimuli and examine the dependency of listener judgements on several acoustic parameters, such as F0, syllable duration and intensity, (2) incorporate results from a different language group, namely native speakers of Mexican Spanish with or without knowledge of the German language, (3) reevaluate our results with respect to subword units which can be legal words of German. As [3] to this date is still in press we will first present the experiment in greater detail.

2. Experiment Stimuli and Design

The stimuli employed in this study are all six-syllable sequences excised from the context of a sentence with the structure “Menschen können A, B, C sein”, (“People can be A, B, C,”) where A, B and C are possible characteristics of people, either adjectives or past participles. These real words possess either two or three syllables, with the lexical accent on the first, second or third syllable, respectively. For each of the conditions we selected 10 unique words none of which contained syllables of the others. From this set of words we constructed groups of three words A, B and C, where A and C were always three syllables long, and B either contained two or three syllables. Here is an example: “Menschen können le-se-nen, schlag-fe-tig, lang-wei-lig sein.” (“People can be erudite, quick-witted, boring.”) Figure 1 displays a stimulus example demarked by grey vertical lines inside the surrounding carrier phrase, syllable boundaries are indicated by dotted vertical lines. By combining three words we yielded groups with all possible positions of the lexical accents, including conditions off accent clash.

We intentionally omitted the conjunction ‘and’ before the third item in order not to supply a morphemic marker of the boundary. The stimuli were then constructed in such a way, that only the B word was completely preserved in the stimulus whereas only parts of word A and C were present.
A stimulus created from the sample sentence could hence be “...lesen, schlagfertig, lang...”, with the lexical accent syllables set in bold face. The stimuli were recorded by one male and one female native speaker of German and the target sequences excised using PRAAT [1] after performing peak scaling to 98%.

In order for the syllables to be accessible for people without knowledge of German, we used a pseudo-transcription based on German SAMPA in which all non-letter symbols where replaced by other letters or deleted (in the case of the lengthening symbol “.”): LE-ZEN-SHLAK-FER-TICH-LANG. The stimuli, a total of 106 tokens, were randomised and an MS-Word-based questionnaire was developed which contained a list of the stimuli with PLAY links to the audio files.

We checked the frequency of the target words and found only 13 of the 30 trisyllabic words in a list of the 10000 most frequent German words [6], and seven of the 20 disyllabic ones. However, 15 of these words are ranked in the last third of the list. Hence, we expected most of them to be unknown to students of German at the beginner’s and even at the intermediate level.

In the header of the questionnaire we inquired about some personal details such as age and gender, learning history of German, time spent in Germany as well as languages spoken at home. Then the experiment design and its aims were explained, namely, to determine the boundaries between words of German as well as the accented (“strong”) syllables of these words which had been cut out of context. Participants were advised to play stimuli a maximum number of four times and note down any words they might have perceived. They were also informed that each stimulus contained at least one accented syllable and at least two boundaries.

The pseudo-transcriptions of the syllables were listed with checkboxes that enabled the selection of each of the syllables/boundaries as shown in the excerpt of the questionnaire displayed in Figure 2 with two illustrative examples in grey and the first seven trials.

Subjects were then asked to work through the examples one by one and make their choices based on their perception. They were also asked to note down how often they played the stimuli and whether they had identified any real words in the sequence. The completed questionnaires were saved in Word format by the participants and later on converted to RTF format for further evaluation of results.

3. Measurements of Acoustic Parameters

All sentences from which the stimuli were excised were segmented manually on the syllable level inside the PRAAT TextGrid, yielding syllabic durations for all stimulus utterances. Means and standard deviations of syllable durations were 245ms/99ms for the male subject and 230ms/93ms for the female.

F0 contours were extracted at a step of 10ms using the PRAAT default pitch extraction settings and subjected to manual inspection and correction.

All utterances were subjected to Fujisaki model [7] parameter extraction [8], see example in Figure 1. The figure displays from the top to the bottom: The speech wave form,
the F0 contour (+ signs: extracted, solid line: model-based), the
text, the underlying phrase and accent commands. We
employed the accent command amplitudes for quantifying the
interval of F0 transitions occurring in each of the stimulus
syllables.

Intensity contours were extracted in PRAAT with default
settings, and mean intensities in dB, as well as maxima
employing parabolic interpolation were determined for each
syllable.

4. Results of Analysis

From [3] we had results from eight native German listeners,
eight Chinese (CN) learners after their first year, 15 Chinese
listeners after their second year, three Chinese students after
year 5 and eleven Chinese non-speakers of the language.

For the current paper we expanded the German group by two
more subjects, as well as added 21 Mexican native speakers of
Spanish (MX), 12 without knowledge of German, five after
year 1 and 4 after year 3 of German classes.

We determined how many of the intended items
(boundaries and accented syllables) had been selected by the
participants. This measure, however, does not reflect
additional selections that were erroneous, that is, insertions.

We therefore calculated an error score, in analogy to
ASR performance evaluations by defining the error as follows:

\[
\text{error} = 100 \times (\text{insertions} + \text{deletions})/\text{total number of tokens}
\]

If we assume that two prominences and two boundaries are
randomly selected on each trial, chance level for the curre n
trial would be at 176.7 for the prominences and 175 for the
maxima.

As can be expected (see Table 1), native speakers fare
well at identifying word boundaries (96.0% correct) whereas

Table 1: Percentage correct and insertion and deletion
errors.

<table>
<thead>
<tr>
<th>group</th>
<th>percentage correct bound. mean/s.d.</th>
<th>percentage correct prom. mean/s.d.</th>
<th>error bound. mean/s.d.</th>
<th>error prom. mean/s.d.</th>
<th>N Subj. (m/f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germans</td>
<td>96.0/5.6</td>
<td>63.5/15.0</td>
<td>6.4/8.4</td>
<td>56.7/17.3</td>
<td>4/6</td>
</tr>
<tr>
<td>CN non-learners</td>
<td>69.1/10.2</td>
<td>46.0/13.5</td>
<td>64.3/19.4</td>
<td>96.2/17.3</td>
<td>3/8</td>
</tr>
<tr>
<td>CN after year 1</td>
<td>80.1/13.1</td>
<td>38.6/22.2</td>
<td>38.1/28.4</td>
<td>81.0/9.9</td>
<td>4/4</td>
</tr>
<tr>
<td>CN after year 2</td>
<td>88.8/7.1</td>
<td>35.1/12.8</td>
<td>20.8/8.8</td>
<td>84.0/11.6</td>
<td>9/6</td>
</tr>
<tr>
<td>CN after year 5</td>
<td>96.2/2.4</td>
<td>78.9/21.3</td>
<td>4.4/2.7</td>
<td>27.0/24.1</td>
<td>0/3</td>
</tr>
<tr>
<td>MX non-learners</td>
<td>70.3/20.3</td>
<td>37.9/14.6</td>
<td>62.3/38.8</td>
<td>91.4/11.8</td>
<td>7/5</td>
</tr>
<tr>
<td>MX after year 1</td>
<td>75.2/15.5</td>
<td>30.9/10.0</td>
<td>39.4/17.3</td>
<td>101.5/13.2</td>
<td>3/2</td>
</tr>
<tr>
<td>MX after year 3</td>
<td>88.6/9.2</td>
<td>45.6/13.4</td>
<td>25.8/10.8</td>
<td>84.8/21.7</td>
<td>2/2</td>
</tr>
</tbody>
</table>

The results of correctness stated so far are based on the
rigorous assumption that all original two- and three syllable
words from which the stimuli were taken were identified
correctly. However, since the left and right words were most
certainly incomplete, these truncations sometimes produced
subword units that corresponded to legal German words. In
addition also the central word often lends itself to further
segmentation, compare, for instance, “liebervoll” (affectionate)
and “Liebe voll” (love full), though these sequences are not
necessarily plausible and require a different assignment of
stress. However, for learners at the initial stage these shorter
words are probably much easier to recognize than the rather
infrequent larger ones. In some cases, subword chunking,
however, did not produce meaningful results. The word
“verlegen” (embarrassed) was sometimes subdivided into
“ver” and “legen” where only the right part bears a meaning of
it itself (to lay). We also observed cases where the
compound accent syllable was missing in the stimulus,
compare “großzügig” – generous, and “vorschnellzügig”
(speedy)’, and subsequently the secondary stress of the
original word became the primary one of the resulting new

In order to assess the effect of subword chunking we
related the number of insertions on the boundary and accent
level to the number of additional boundaries and accents due
to the potential subword units. The results for the different
groups are listed in Table 2. Here we see a clear difference
between non-learners and learners of German in that many
more insertions of the learners create legal subwords. Another
interesting result is that the vast majority of all boundary
insertions by the Mexican learners can be attributed to these subword units. With respect to non-words created by additional word boundaries we see that the proportion of insertions explained is quite similar for all groups, except for the advanced Chinese students who are already almost perfect. However, the ratio of these insertions compared to the total number of additional boundaries due to subwords is only 24.0% for Chinese and 20.9% for Mexican learners after year 1, for instance, but 35.8% for the Chinese non-learners. This reflects the greater phonological competence of the learners.

Table 2: Percentage of insertions explained by subword units.

<table>
<thead>
<tr>
<th>group</th>
<th>% corr. bound. subword</th>
<th>% bound. creating non-words</th>
<th>% corr. prom. subword</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germans</td>
<td>70.3</td>
<td>16.7</td>
<td>58.7</td>
</tr>
<tr>
<td>CN non-learners</td>
<td>48.3</td>
<td>18.0</td>
<td>30.0</td>
</tr>
<tr>
<td>CN after year 1</td>
<td>72.4</td>
<td>17.3</td>
<td>46.2</td>
</tr>
<tr>
<td>CN after year 2</td>
<td>73.2</td>
<td>17.0</td>
<td>49.8</td>
</tr>
<tr>
<td>CN after year 5</td>
<td>66.7</td>
<td>33.3</td>
<td>93.0</td>
</tr>
<tr>
<td>MX non-learners</td>
<td>50.6</td>
<td>17.4</td>
<td>30.9</td>
</tr>
<tr>
<td>MX after year 1</td>
<td>80.8</td>
<td>16.9</td>
<td>25.2</td>
</tr>
<tr>
<td>MX after year 3</td>
<td>81.4</td>
<td>12.2</td>
<td>44.6</td>
</tr>
</tbody>
</table>

We examined which factors triggered the selection of an item and compared Germans with Chinese and Mexican non-learners because these two groups are maximally different with respect to the kind of information they can draw on to access the stimuli. To this effect we calculated the ratio for each item – either prominent syllable or boundary - to be selected by the three groups.

In the following analysis we do not yet consider any acoustic measurements, only the structure of the stimuli. Each syllable was classified regarding the following features: (1) lexical stress, (2) vowel length, (3) lexical stress on left/right neighbour, (4) boundary type left/right, (5) number of phones in onset/coda. Boundaries were classified regarding the following features: (1) word boundary, (2) lexical stress on syllable to the left/right. Table 5 shows the results of correlation analysis between the above-mentioned features and the ratio at which a syllable or boundary was selected. Comparison of figures shows that both groups show similar tendencies. The fact that vowel length is correlated with the ratio at which a syllable was marked as accented is probably as much due to the structural properties of German as much as due to vowel length being a prominence-lending acoustic feature. In the set of words we selected for this study 68% exhibit lexically stressed syllables with long vowels. An important difference between Germans and non-learners is the effect of adjacent word boundaries on the tendency for a syllable to be perceived as accented. We find negative correlation with the left syllable boundary and positive correlation with the right syllable boundary which is considerably stronger for the non-learners. This means a stronger preference for word-final syllables to be perceived as accented than for word-initial ones. In contrast, the effect of the neighbouring syllables’ stress status appears stronger in the German subjects, indicating a preference for avoiding stress clash. We also calculated a regression model based on the factors chosen and find that it explains 87% of the variance for the German listeners, but only 68% for the non-learners.

With respect to the boundaries, the status of the boundary as either being an inter- or intra-word boundary clearly guides the decision whereas the stress status of the adjacent syllables is irrelevant to the judgments of the Germans. In contrast, a stressed syllable to the left of the boundary still has a significant effect on the judgments of the non-learners. This interrelationship matches the result for the accented syllables.

In terms of the learning effect during the experiment we compared the correctness of the results on the first and the second half of stimuli, but did not find any significant differences. We also examined the relationship between the judgments of the subjects and the acoustic properties of the stimuli. First of all we looked at how the acoustic correlates varied with the status of a syllable regarding stress and its position with respect to a word boundary. Table 3 lists some of the results for syllabic duration, accent command amplitude $Aa$, as well as mean and maximum intensity in $dB$. As expected, stressed and pre-boundary syllables are longer and exhibit larger $F0$ excursions as reflected by $Aa$. The values calculated from the PRAAT intensity contours reflect only slight differences.

Table 3: Means and standard deviations of some prosodic parameters with respect to the status of the underlying syllable.

<table>
<thead>
<tr>
<th>Syllable Status</th>
<th>Syllable Duration [ms]</th>
<th>Accent command amplitude $Aa$</th>
<th>Mean intensity [dB]</th>
<th>Max Intensity [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stressed</td>
<td>313/82</td>
<td>.30/20</td>
<td>70.8/4.6</td>
<td>79.6/3.8</td>
</tr>
<tr>
<td>Unstressed</td>
<td>193/74</td>
<td>.28/20</td>
<td>70.2/4.7</td>
<td>78.1/3.9</td>
</tr>
<tr>
<td>Pre-boundary</td>
<td>301/92</td>
<td>.40/14</td>
<td>70.4/4.4</td>
<td>78.5/4.1</td>
</tr>
<tr>
<td>Post-boundary</td>
<td>210/86</td>
<td>.21/22</td>
<td>69.6/4.7</td>
<td>78.4/3.9</td>
</tr>
</tbody>
</table>

Table 4: Correlations (Pearson’s $r$) between the ratio of a syllable being selected as accented and its underlying prosodic features.

<table>
<thead>
<tr>
<th>group</th>
<th>syllable duration [ms]</th>
<th>accent command amplitude $Aa$</th>
<th>mean intensity [dB]</th>
<th>max intensity [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germans</td>
<td>.61**</td>
<td>.10*</td>
<td>.06 (n.s.)</td>
<td>.17**</td>
</tr>
<tr>
<td>CN non-learners</td>
<td>.54**</td>
<td>.29**</td>
<td>.08*</td>
<td>.18**</td>
</tr>
<tr>
<td>MX non-learners</td>
<td>.60**</td>
<td>.30**</td>
<td>.13**</td>
<td>.28**</td>
</tr>
</tbody>
</table>
Table 5: Factors facilitating the selection of accented syllables and boundaries for Germans as well as Chinese and Mexican non-learners.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Corr. r with ratio selected (Germans)</th>
<th>Corr. r with ratio selected (CN non-learners)</th>
<th>Corr. r with ratio selected (MX non-learners)</th>
<th>Variance explained</th>
<th>Corr. r with ratio selected (Germans)</th>
<th>Corr. r with ratio selected (CN non-learners)</th>
<th>Corr. r with ratio selected (MX non-learners)</th>
<th>Variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical stress</td>
<td>.76**</td>
<td>.46**</td>
<td>.56**</td>
<td>.87</td>
<td>.97**</td>
<td>.72**</td>
<td>.82**</td>
<td>.97</td>
</tr>
<tr>
<td>Long vowel</td>
<td>.52**</td>
<td>.38**</td>
<td>.40**</td>
<td>.68</td>
<td>.03(n.s)</td>
<td>.12**</td>
<td>.12**</td>
<td>.74</td>
</tr>
<tr>
<td>Stress left</td>
<td>-.34**</td>
<td>-.25**</td>
<td>-.29**</td>
<td>-0.01(n.s)</td>
<td>.01(n.s)</td>
<td>.00 (n.s.)</td>
<td>-.04 (n.s.)</td>
<td>.83</td>
</tr>
<tr>
<td>Stress right</td>
<td>-.31**</td>
<td>-.17**</td>
<td>-.18**</td>
<td>-0.13**</td>
<td>-0.12**</td>
<td>-.24**</td>
<td>.30**</td>
<td>-.82**</td>
</tr>
<tr>
<td>Boundary left</td>
<td>-.13**</td>
<td>-.30**</td>
<td>-.23**</td>
<td>.11*</td>
<td>.36**</td>
<td>.31**</td>
<td>.31**</td>
<td>.97**</td>
</tr>
<tr>
<td>Boundary right</td>
<td>.15**</td>
<td>.13**</td>
<td>.10**</td>
<td>.21**</td>
<td>.24**</td>
<td>.30**</td>
<td>.30**</td>
<td>.74</td>
</tr>
<tr>
<td>N onset</td>
<td>.01(n.s)</td>
<td>.00 (n.s.)</td>
<td>-.04 (n.s.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N coda</td>
<td>.04 (n.s.)</td>
<td>.02 (n.s.)</td>
<td>.01 (n.s.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subsequently we correlated the ratio selected for prominent syllables with their associated prosodic features. The result is displayed in Table 4. As can be seen, syllabic duration is the strongest cue for non-learners, followed by F0 and maximum intensity of the syllable. It is interesting that the Mexicans seem to respond more strongly to intensity than the Chinese. A regression model based on the three factors duration, Aa and max intensity explains 58.2% of the variance for the Chinese learners and 65.1% for the Mexicans.

5. Acknowledgements

Many thanks go to Hongwei Ding, Tongji University, Shanghai, China, for collecting the data of Chinese subjects. Thanks also to Angelika Hönnemann for help with questionnaires and some of the data analysis.

6. References


