How do we chunk and pause in non-native vs native speech?

Lucie Judkins1, Charlotte Alazard-Guiu1, Corine Astésano1

1Laboratory of NeuroPsychoLinguistics, University Toulouse 2, France
lucie.judkins@univ-tlse2.fr, corine.astesano@univ-tlse2.fr, charlotte.alazard@univ-tlse2.fr

Abstract

The present preliminary study looks at speech units and pauses’ quantity, duration and distribution in native (L1) vs non-native (L2) spontaneous speech. 6 English L1 – French L2 speakers (2 females) were recorded in their L1 and L2 on a conversation task. 3 speakers had an elementary L2 proficiency level, the other 3 were advanced. We looked at 1) the effect of speaking an L2 vs an L1 (within-speaker analysis) and 2) the effect of the proficiency level (between-speaker analysis) on the quantity, distribution and duration of silent and voiced pauses and the quantity and duration of Inter Pausal Units (IPU).

Our results show that the IPU tend to be more numerous and shorter in L2 but those differences do not reach significance. Pauses are more numerous in L2 regardless of the proficiency level and the difference between L1 and L2 is less important in the advanced group. These tendencies concern mainly the quantity and less so the duration and distribution. Additionally, we notice a greater diversity in the type of pauses in L2.

These observations shed light on the reality of interlanguage and the methodological issues involved in its description. Implications for L2 speech analysis and evaluation are discussed.

Index Terms: non-native prosody, fluency, pauses, IPU, interlanguage, Second Language Acquisition (SLA)

1. Introduction

Studies in Second Language Acquisition (SLA) concerned with the description and perception of non-native prosody are scarce [1] but findings on some aspects of non-native prosody are discussed under the scope of fluency studies [2] [3]. Such work is becoming more and more numerous as second language learning increases in accessibility and demand and with it, the need to accurately assess learners’ oral skills. Fluency measures comprise several temporal variables traditionally assembled into categories: speed measures (such as speech rate, syllable rate, phonation/time ratio etc.), breakdown measures (Mean Length of Run, pauses), and disfluencies (repetitions, false starts, repair). Our study focuses on breakdown measures which are temporal variables that have been identified as relevant indicators of fluency [2] and as being acoustic correlates of speech rhythm. Indeed, rhythm being a fundamental and preeminent level of structuration in oral language [4], studying such variables gives us an understanding of how speaking an L2 affects temporal organization of speech, reflecting the underlying cognitive processes involved [5]. Silent and voiced pauses have been studied as markers of such processes. Pauses located at speech unit boundaries have been associated with macro-planning processes such as syntactic organization, whereas pauses located inside speech units reflects more local micro-planning processes such as lexical retrieval [6] [7]. These concepts of macro and micro-planning processes have been developed in Levelt's speech production model ([8]).

With such perspective in mind, we used a within-speaker design which contrasts with most studies on fluency/non-native speech, that are done through between-groups comparison of L2 vs native speech and focus on the "distance" between the groups’ performances. To our knowledge, very few studies have focused on the language-dependent within-speaker variation (except from [3] and [9]). However, a between-speaker design is necessary to observe the development of temporal variables variations as learners gain proficiency. Consequently, our study combines both a within-speaker analysis focusing on the effect of the language variable (L1 vs L2) as well as a between-speaker analysis in order to test the proficiency level (elementary vs advanced) effect on chunking and pausing.

This is a preliminary study which is part of a larger project involving native and L2 speakers of both French and English. The aim of this project is to compare the L2 effect with the spoken language effect and see how these two variables interact on rhythm patterns [10].

2. Corpus B-FREN3 methodology

This study was carried out on speech samples extracted from the corpus B-FREN3 [11], a bilingual French-English/English-French corpus comprising three speech styles: reading aloud, film extract retelling, and conversation. In this section, we describe the methodology used to collect the data relevant to the present study i.e., the recordings of English L1-French L2 (elementary and advanced proficiency levels) speakers in a conversation task.

We hypothesized that in L2, speakers would chunk more and make shorter speech units as well as realizing more and longer pauses. We also expect these tendencies to be more important in the elementary level group than in the advanced group.

2.1 Speakers

The speech samples belong to 6 adult native speakers of English, aged between 28 and 41 (average = 33.2) at the time of recording. All are native English speakers of different varieties of English (1 Scottish, 1 Scottish-American, 1 New Zealander, 3 Americans; 2 females). Although this diversity could impact inter-individual variation, there is to this day no strong evidence that these varieties of English could in fact differ significantly in terms of temporal variables. On the opposite, Bolinger in [12] has pointed that British and American English share a common prosodic system, where differences appear to be mainly connected to pragmatic choices of the speakers.
All 6 participants speak French as an L2 at different proficiency levels such that we could split them into 2 groups of 3 speakers. One group of elementary (A2) L2 level speakers, and the second of advanced (B2) level speakers (A2 and B2 levels from the Common European Framework of Reference for Languages [13]). The L2 proficiency levels of the participants were first assessed by the experimenter (first author) on the basis of her expertise as a foreign language teacher, through an informal interview. They were re-assessed post-hoc by anonymous French as a Foreign Language teachers who were asked to evaluate the participants’ level on the basis of two recordings from the corpus (retelling and conversation task) and according to the CEFR level descriptors. Each participant was evaluated by 3 different judges and a final score was obtained by combining the 3 ratings into an average. The judges’ ratings confirmed the initial assessment.

2.2 Task & procedure

The participants were recorded while performing a casual conversation task with the experimenter, on a familiar topic such as their last vacation trip or the last book they read/movie they saw. The task was performed once in their L1 and once in their L2 (with a different topic) and the order of the tasks’ language was randomized. The recording sessions took place either at the experimenter or the participant’s house guaranteeing a casual and comfortable setting while assuring adequate acoustic conditions. Each task was recorded separately using a Linear PCM Tascam DR 40 Version 2 recorder. The recordings vary in length between 2 and 5 minutes. For the purpose of this study, the samples used are all between 118 and 123 seconds long, taken in the middle portion of the recordings, when the participant has had time to get used to the task and experimental setup.

2.3 Annotation

The 12 selected speech samples (2 per participant, 1 in English (L1), 1 in French (L2)) were manually annotated using Praat. Speech was segmented into Inter Pausal Units (IPU) which are defined as a speech run between silent pauses of at least 250ms [2] [14]. In total, each speech sample comprises on average 2 minutes of speech and 30 IPU. We used the same threshold for all silent pauses [15] [16] meaning they are de facto external of IPU, marking their boundaries. Voiced pauses are of two sorts: interjections and lengthenings. Interjections are word-external hesitation markers such as [œː] in French and [mː] in English [17]. As such, they are not defined by a duration threshold but they are easily identifiable on the basis of auditory and visual cues on the spectrograph [6] [18]. Lengthenings correspond to word-final syllables lengthened over 200ms, this threshold being generally considered as the pause perception minimum in the speech flow [7] [19]. Lengthenings are a subcategory of voiced pauses which are more common in French, given the high frequency of word-final open syllables [20]. They also occur in English, mainly on mono-syllabic words such as “and” or “the”. However, not all lengthenings are hesitation markers. Indeed, some can bear a semantic-pragmatic function [21] [22] and therefore not be connected to underlying speech planning cognitive processes. Consequently, these have been excluded from the study in the annotation process.

When a voiced pause was realized between two IPU (>250ms silence + voiced pause + >250ms silence) it was labelled as a “mixt pause”. Pauses were also labelled according to their distribution: IPU internal vs IPU external and within an IPU according to their position: initial, median, final, although this will not be commented in this paper. Table 1 shows the different types of pauses annotated and their characteristics.

<table>
<thead>
<tr>
<th>IPU-External Pauses</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory Pauses</td>
<td>&gt;250ms with breathing sound and visible pattern on oscillograph</td>
</tr>
<tr>
<td>Silent Non-respiratory Pauses</td>
<td>&gt;250ms</td>
</tr>
<tr>
<td>Mixt Pauses</td>
<td>Voiced Pauses between 2 silent pauses &gt;250ms</td>
</tr>
<tr>
<td>IPU-Internal Pauses</td>
<td></td>
</tr>
<tr>
<td>Voiced Pauses</td>
<td>No threshold, use of auditory and visual cues</td>
</tr>
<tr>
<td>Lengthenings</td>
<td>&gt;200ms</td>
</tr>
</tbody>
</table>

Table 1: Annotated pauses characteristics and thresholds

In our analysis, silent non-respiratory pauses and mixt pauses are combined under the term “external non-respiratory” pauses and internal voiced pauses and lengthenings are combined under the term “internal pauses”.

To measure speech chunking, we calculated the ratio of IPU quantity (multiplied by 100 for legibility purposes) relative to the total speech time (number of IPU * 100 / total IPU time) as commonly done [3]. The same formula was used to calculate the pause ratio expect this time the total speech time included IPU-external pauses (number of Pause * 100 / total IPU + total External Pause Time) [15] [16] [23].

2.4 Data processing

A Praat [24] script was used to extract all annotated items’ duration (IPU and all types of pauses). We looked at their quantity, raw duration and calculated the variation coefficient (VC) on the duration as an indicator of the dispersion following the formula:

\[ VC = \frac{\sigma}{\mu} \]  

Where \( \sigma \) is the population standard deviation and \( \mu \) the population mean.  

Statistics: We tested the effects of the language and proficiency level variables separately as the latter is only relevant on the L2 speech samples. We computed linear regression models on all the duration dependent variables (IPU, External Respiratory Pauses, External Non-Respiratory Pauses, Internal Pauses) using R Studio (version 2021.09.0) and use p-value = 0.05 as significance threshold. These tests were run only on the duration measures, which yielded substantial enough datasets. Statistic tests are not pertinent however on our chunking and pause ratios because of the reduced amount of data. We present our results on duration using medians (med.) which we find more representative of the data tendencies given the amount of variation (see variation coefficients presented in the next
section). The results on the quantity ratios are presented using means, more representative of a small sample size.

3. Results

3.1. Language effect (L1 vs L2)

This is a within-speaker analysis comparing speech samples in 6 speakers’ L1 (English) and L2 (French).

3.1.1 On IPU

In terms of quantity, the results on the chunking ratio show that speakers tend to chunk more (higher quantity of IPU) in L2 (chunking ratio mean = 44.37) than in L1 (chunking ratio mean = 38.46). The IPU duration (expressed in seconds) seems to be impacted by the language variable, however the effect does not reach significance (p-value = 0.291; F(1;394) = 1.117). IPU tend to be shorter in L2 (med. = 1.85) than in L1 (med. = 2.14) as predicted, with a difference of 292ms. In addition, our measure of the variation coefficient shows that the IPU durations vary a lot in L1 (61.68) as well as in L2 (61.98), which, as stated in the previous section, makes the median a better indicator than the mean. From these results, it seems like the language variable does not have a strong effect on the IPU quantity and size.

3.1.2 On pauses

Given our definition of the IPU, the external pauses (respiratory and non-respiratory) embody their boundaries. It is therefore logical that in accordance with the results exposed above on the IPU, the quantity of external pauses is not greatly impacted by the language variable either. Both respiratory pauses and non-respiratory pauses are a little more numerous in L2 but again the difference is small (mean respiratory pauses ratio L1 = 13.86, L2 = 15.08; mean non-respiratory pauses ratio L1 = 16.33, L2 = 19.7). The difference in duration is also not significant for both these types of pauses however the tendencies go opposite ways. Indeed, respiratory pauses tend to be longer in L1 (med. = 0.739) than in L2 (med. = 0.654; p-value = 0.219; F(1;195) = 1.517) which follows previous studies’ results [25], whereas non-respiratory pauses tend to be longer in L2 (med. = 0.691) than in L1 (med. = 0.525; p-value = 0.078; F(1;240) = 3.122). Noticeably, the variation in duration is much greater in the case of non-respiratory pauses (see Figure 3 comparing the variation coefficient of all items). This can be explained by the fact that these pauses comprise silent pauses as well as mixt pauses which are longer in essence as they are composed of a voiced section in between silences. The quantity of internal pauses is greatly impacted by the language variable. As shown in Figure 1 below, there are more numerous in L2 (mean = 27.18) than in L1 (mean = 10.2). However, the duration of internal pauses is not significantly different between L1 and L2 (med. L1 = 0.468s; med. L2 = 0.493s; p-value = 0.912; F(1;256) = 0.012) and the variation coefficient is under 50 (L1 = 36.31; L2 = 44.35), so internal pauses are fairly stable in terms of their duration in both L1 and L2.

3.2. Proficiency effect in L2 speech

This is a between speaker analysis comparing 3 anglophones A2 level English and 3 anglophones B2 level French.

3.2.1 On IPU

As shown in Figure 2 below, the A2 group chunks more than the B2 group. However the great between-speakers variability needs to be taken into account. The IPU duration is significantly impacted (p-value = 0.031; F(1;196) = 4.716) and as expected, more advanced learners (B2 level) make longer IPU (B2 med. = 2.371s.; A2 med. = 1.564s.). It is important to note the strong variability in IPU duration, as indicated by the variation coefficient: 66.23 in A2 and 57.72 in B2.

3.2.2 On pauses

As expected, respiratory pauses remain fairly stable in quantity (mean A2 = 15.59; mean B2 = 14.57) and duration (med. A2 = 0.683s; med. B2 = 0.629s) regardless of the proficiency level. Non-respiratory external pauses are more numerous (mean A2 = 22.91, mean B2 = 16.48) and longer in the A2 group (med. A2 = 0.830s., med. B2 = 0.508s.) but the difference is not significant (p-value = 0.190; F(1;128) = 1.735). The internal pauses don’t vary much in quantity (mean A2 = 27.53, mean B2 = 26.83) but their duration is significantly longer in the A2 group (med. A2 = 0.540s., med. B2 = 0.439s., p-value = 0.001; F(1;183) = 10.28).
3.3 Further remarks

Figure 3 below shows the variation coefficient of all dependent variables in both L1 and L2, and proficiency levels.

![Graph showing variation coefficient of all dependent variables]

This is illustrating the amount of variation in duration for each dependent variable. Clearly, the external non-respiratory pauses, connected to macro-planning processes vary a lot. The IPU duration also varies greatly and this variation does not seem to be imputable to the language nor the proficiency variable (as explained in sections 3.1 and 3.2).

As expected, the respiratory pauses vary much less as they reflect a basic physiological rhythm constraint. However, it is interesting to note that internal pauses, which are the ones impacted by the language variable in terms of quantity, also do not vary as much in duration overall, therefore reflecting a sort of regularity in the shape that take micro-planning related pauses.

Lastly, more detailed data on the types of pauses (nature and distribution) has not been presented here but we observed a greater diversity in L2, meaning that some types of pauses are not used in L1 but appear in L2, such as lengthening in medial and final position within IPU.

4. Conclusion & discussion

Our results have shown that the language variable has the greatest impact on the internal pauses quantity. This confirms our expectations as we hypothesized that speakers would pause more in L2 and supports the results of [23] as pauses are impacted most drastically within IPU. This indicates that it is mostly the micro-level of speech planning which is affected when speaking an L2, whereas external (non-respiratory) pauses, reflecting the planning process of the next chunk (macro-level), do not differ as much when speaking an L1 or an L2 (the difference in our results was not significant).

The quantity of IPU and external pauses behave similarly given our criteria for defining this speech unit, and even though they follow the expected tendency of being more numerous in L2, the differences are not significant, in contrast with other studies such as [26]. On a rhythmic perspective, we could infer that speakers tend to keep chunks' size fairly similar but fill them with more voiced pauses, such as the global chunk pattern is not altered.

The duration of IPU and all types of pauses is not significantly different from L1 to L2 although the tendencies also follow the literature: shorter IPU in L2, longer external non-respiratory pauses in L2, longer respiratory pauses in L1 ([23] [25], although [3] and [9] obtain different results).

On a methodological standpoint, it seems clear that pauses’ distribution must be taken into account in fluency and non-native prosody studies. Indeed, pauses do not vary equally depending on their functions (physiological, macro-planning, micro-planning) and associated positions in the speech flow. As mentioned by [25] the distinction of pauses’ types is often unclear or even absent in studies ([27]).

The effect of the proficiency level is significant on the duration of IPU which are longer in the advanced group (B2), and internal pauses which are longer in the elementary group (A2), as expected. The quantity of IPU and external non-respiratory pauses behave again hand in hand, following the expected tendency of being more numerous in the A2 group. Interestingly, the quantity of internal pauses is very comparable in the two groups and it is their duration which is affected by the proficiency level since they are significantly longer in the A2 group, in accordance with our hypothesis. This conclusion has to be nuanced though, as previous research on the impact of proficiency level on pauses ([3]) has shown that this can be influenced by the levels themselves. The difference in pause duration seems accurate to distinguish only between an elementary level and an intermediate or above (which is our case) but it is not as clear in the case of intermediate vs advanced levels. [28] have also shown that speech breakdown improvement (longer IPU and less internal pauses) is particularly strong in the first stages of SLA.

5. Perspectives

As mentioned in our introduction, this paper presents a preliminary study which is to be extended to the entirety of our corpus, allowing us to include speakers of French L1-English L2 as well. We aim to confront the findings on L2 and proficiency effects with the L1 effect. We are also planning further observations on speech rhythm correlates such as the size of rhythmic groups (accentual phrases) within IPU ([29]) and the regularity of metrical stress.

This work is carried out with the perspective of better understanding speech rhythm pattern variations across languages as L1 and L2 but also through the influence of the language itself (here, French and English). The larger aim of this project is to integrate knowledge about rhythm variation across languages and L1-L2 into L2 prosody teaching methods. We could for example imagine an L2 prosody training that focuses on where and how to chunk and pause in the target language, ideally taking into account the learners’ L1.

6. References
