PREDICTABILITY OF VOICING ASSIMILATION IN SPEECH PRODUCTION

Mária Gósy
Research Institute for Linguistics, Hungarian Academy of Sciences, and Eötvös Loránd University, Budapest, Hungary

ABSTRACT

During speaking, the mental lexicon is accessed (i) to select the necessary words, and (ii) to retrieve their phonological and syntactic patterns. However, the nature of real-time activation of words and phonological rules is largely unknown. This paper deals with the temporal properties of voicing assimilation (i) in language acquisition, (ii) in spontaneous speech, and (iii) in repetition tasks. Results show that (i) by the age of 4 Hungarian-speaking children acquire this phonological rule without mistakes, (ii) in spontaneous speech if the pause between two words does not exceed 55 ms, voicing assimilation takes place; however, if it exceeds 324 ms, voicing assimilation does not apply, (iii) in repetition tasks subjects retrieved voicing assimilation successfully in about 71% of all possible cases.

1. INTRODUCTION

In spontaneous speech, speakers transform their thoughts into various linguistic forms. The mechanism that makes this possible works at very high speed. During speaking the mental lexicon gets activated in order to let the necessary words be found that are appropriate for the actual topic of verbal communication. The same mental lexicon is responsible for retrieving these words' phonological and syntactic patterns. However, not much is known about the nature of real-time activation of words and phonological rules. Preparing words in speech production is a fast and accurate process, however, the underlying process is very complex.

Hungarian has a rule of voicing assimilation whereby obstruent clusters come to share the voiced/voiceless specification of their rightmost member. Voicing assimilation is postlexical it applies across word boundaries as well as within words. E.g.:

Ismered a fiút, boldog. ('You know the boy, he is happy.')
[fiju:**d b**oldog] [t] → [d]
A víz hideg. ('The water is cold.')
[vi:**s h**ideg] [z] → [s]

During speaking no one pays conscious attention to fulfilling the phonological rules of voicing assimilation, the speaker is not aware of this kind of operation. This can easily be accepted when assimilation applies within a word. However, many questions arise when the same rules apply across word boundaries in spontaneous speech. The first question concerns language acquisition. When are the children able to retrieve the actual pattern of voicing assimilation during speaking? The second question concerns adults' spontaneous speech where pauses between two words frequently block the realization of voicing assimilation. What is the reason behind this phenomenon? Is voicing assimilation predictable or, failures of this phonological rule are mere coincidences?

Taking Levelt's model of speech production [3] into consideration, what should be analysed here is the level of phonological coding which sometimes seems not to be activated properly. The activation of mental lexicon consists of four levels of processing: the activation of lexical concepts, the selection of lemmas, the morphological and phonological encoding of a word in its prosodic context, and the word's phonetic encoding. The present hypothesis is that the factor of time defines the successful operation of phonological coding across word boundaries. If the pause is long, the phonological process cannot take place since the following word is not yet activated. This means that the preceding word maintains its own basic phonological pattern stored in the mental lexicon because the delay of the activation of the following word restricts the operation of the phonological coding. There is an investigation concerning the timing of activation of the phonological level [4]. Results show that the speaker is able to retrieve the gender of a noun 40 ms earlier than he is able to define its phonological pattern. This paper aims at investigating the nature of voicing assimilation in relation to the factor of time.

2. ACQUISITION OF VOICING ASSIMILATION

The first meaningful words in language acquisition are picked up from the environment that provides morphologically, phonologically and phonetically ready structures to the child. However, at the beginning the child is not able to imitate the same forms due to his developing articulation gestures. When the child begins to create two-word and later multi-word sentences, word order is entirely dictated by semantics and there is no actual 'phonetic interrelations' among the word forms. Frequently, there is some silent pause between the words of such utterances – 50-400 ms according to our data – representing characteristic temporal patterns of the child's lexical access during speech production. Lexical access is a process of activation of single words in this phase of language acquisition so, the question of phonological coding does not arise.

By the age of 2;5 this simple process gets more complex when increasingly complicated semantic and morphological and syntactic structures appear in the child's speech production. The child quickly acquires not only the meaning of many words but also a relatively great set of frequently heard (and used) rules.

This level of language acquisition provides an opportunity for the child to start applying phonological rules across word boundaries as well.

Five children's (three girls, two boys) systematically tape-recorded spontaneous speech (altogether 10.5 hours) was analysed between the ages of 2 and 4. At the age of 2 all of them could pronounce two-word utterances and most of the voiced/voiceless consonant pairs correctly, so the realization of voicing assimilation could be analyzed (by means of CSL50). Three overlapping phases were differentiated during the period investigated. In the first phase (between 2 and 2;11) the children realized voicing assimilation within words but they were unable to perform it across word boundaries. They recalled the items of the intended utterance from their mental lexicon as independent words, with only a serial connection between them. The second phase begins at around 2 years and 10 months and lasts until the age of 3. This period is characterized by the overgeneralization of voicing assimilation. The children apply the rules not only to the appropriate consonant clusters but also to pre-sonorant obstruents. E.g.:

Within words: *Krisztián* ('boy's name')
[gristija:n]
nagyuraknak ('for noblemen')
[naJuragnak]

Across word boundaries: *ügyes vagyok* ('I am clever')
[yJe3 vaJok]
nem is megyek ('I do not go')
[nem i3 meJek]

The successful acquisition of the voicing assimilation rule presumes the child's language awareness concerning the nature of voiced and voiceless consonants, and in the first phase this leads to overgeneralization. (Note: there are some Hungarian dialects where voicing assimilation concerns all voiced consonants.) Frequently, in this phase the same utterance shows both the expected and the overgeneralized assimilation rules. E.g.:

egy pár zokni kell neki ('he needs a pair of socks')
(i) adult's production: [e**c** **p**a:r zo**kn**i kel: neki]
(ii) child's production: [e**c** **p**a:r zo**gn**i kel: neki]

The third phase (from the age of 3) shows an adult-like use of voicing assimilation in the children's spontaneous speech. After the age of 4 no trace of overgeneralization of voicing assimilation could be found with any of the children.

3. VOICING ASSIMILATION IN ADULTS' SPONTANEOUS SPEECH

3.1. Method and material

5-hour samples of 4 Hungarian-speaking adults' (two women, two men) spontaneous speech was analyzed by means of CSL50. Before the acoustic analysis, the recorded texts were transcribed using IPA symbols by 3 students of the Department of Hungarian Linguistics. Summarizing and controlling the transcriptions, all possible consonant clusters for voicing assimilation were defined whether phonological coding had been activated or not. Duration of pauses between words was measured as well.

A repetition experiment was conducted involving 5 young adults (3 females, 2 males, ages between 19 and 22) where the subjects' own speech perception influenced their speech production. In these experiments, a 3-minute text was administered through headphones to the subjects who had to repeat what they (had) heard – or, expected to hear – immediately. Both the subjects' repetition and the original text were tape recorded. The text was tape recorded by a female voice and contained 253 words and 32 places for voicing assimilation; 65.6% of them resulted in a voiced consonant while 34.4% in a voiceless consonant. All possible consonants for voicing assimilation occurred in the text. The speech tempo of the heard text was 10.6 sounds/s in general. The subjects had no problem with "synchronization". This method was first introduced by Chistovich and her colleagues [1] to show the nature of the listener's 'inner speech' in order to support the motor theory of speech perception.

4. RESULTS

4.1. Analysis of spontaneous speech

Data of spontaneous speech show that there were 6.1 places for voicing assimilation per minute in general across word boundaries. This is a relatively frequent occurrence of this type of assimilation in Hungarian. Speakers retrieved voicing assimilation successfully in 71% of all possible cases – independently of having had pauses between the words. In 62.4% of all cases of voicing assimilation there was no pause between the words, which resulted in a perfect activation of phonological coding. If there was no pause between the words there was no mistake in the students' transcriptions either. However, in 37.6% of all cases there were pauses of various durations between the words resulting in a lack of voicing assimilation in 77.2% of these cases. In other words, if there is a pause between the words, its duration influences the activation of phonological coding. If the pause between the words does not exceed 55 ms, voicing assimilation takes place. If the pause exceeds 324 ms, voicing assimilation does not apply. Between the two given values – 55 ms and 314 ms – the speech production mechanism is either able to activate voicing assimilation rules together with selecting the words, or fails to do so. If the pause is longer than 314 ms, there is no more chance for the mechanism to "recover" the voicing assimilation. The mean value of pauses when voicing assimilation took place is 149.22 ms while the mean value of those pauses that blocked the realization of voicing assimilation is 472.76 ms (Figure 1). (There is a significant difference between them: $p<0.001$.) The value of the longest pause between two words – where voicing assimilation should have been performed – was 1080 ms but the assimilation did not take place. The transcribers' own language awareness concerning voicing assimilation influenced their perception a great deal. All of them made specific mistakes: they coded the voicing assimilation across word boundaries in many cases when there was actually a lack of this phonological coding in the heard text. The students' working perception processes mirrored the heard sentences in their expected, normative forms fulfilling the necessary phonological rules. Since voicing

assimilation seems to be a strong phonological rule in speech production and also in speech perception – children's data

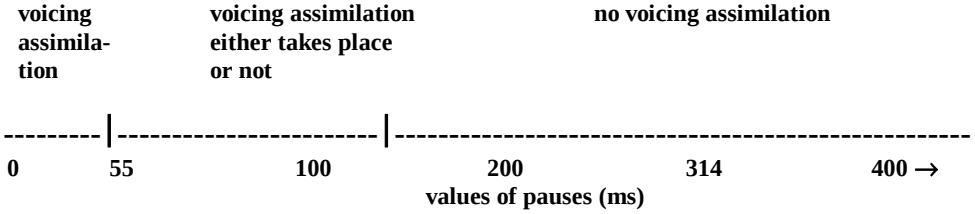

Figure 1. Realization of voicing assimilation across boundaries when there is a pause between the words

confirmed no mistakes after the age of 4 – listeners might have expected consonants in the sound combination instead of the really produced one. Analysis of the students' mistakes shows that there is a strong correlation between the speakers' pauses between the words and the real transcriptions of the voicing assimilations (r=0.76717). The longer the pause between the words the less the number of misperceived voicing assimilations (cf. Figure 2). If the pause exceeded 600 ms, the students realized the lack of voicing assimilation in all cases, however, if the pause was less than 600 ms they could not transcribe the text properly. The portion of the transcriptions' mistakes is 22.01% of all cases with the three students on average. However, if the pause between the two words was less than 100 ms the portion of the transcriptions' mistakes is 53.9% of all cases with the three students. None of the students made the opposite mistake, i.e. they did not code the lack of voicing assimilation if that had been produced. The expectation of the appropriate speech production defined the listeners' speech perception.

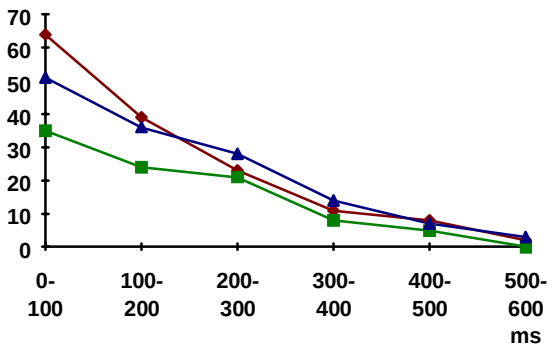

Figure 2. The interrelations of the pauses between the words in spontaneous speech and the transcribers' misperception in percentage

The lack of voicing assimilation can be classified into four categories according to the motivation of the occurrence of pauses with longer durations [2]. These pauses are found (i) before or after a conjunction, (ii) in cases of problems with lexical access, (iii) in the case of false syntactic structure and (iv) in cases of greater semantic shift between the two parts of the utterance. The conjunction frequently introduces a totally new string of thoughts, so the speaker needs time for macro- and microplanning which will block the phonological coding. E.g.: "*I have been told ... but I considered that...*". Problems with lexical access might have multiple reasons. Searching takes time and this leads to the lack of voicing assimilation. E.g.: "*There were two ... secondary grammar schools*." (The speaker wanted to be very proper in defining the actual type of school.) Spontaneous speech frequently contains references to known information that had been told previously. However, the structures of these references might not be very proper and a pause can occur while speaking. E.g.: "*of course this thing ... though I cannot give any evidence.*" The intended sentence was presumably as follows: "*Of course this thing is impossible though I cannot give any evidence to support my opinion*" (the speaker's lack of evidence had been a known information to the listener). The pauses of these cases are very useful for the listener since they provide time for him to rethink what he had heard in order to conceptualize the actual text.

2.2. Analysis of speech repetition

In this material it was only in 22.5% of all cases that speakers had pauses between the words (the same value for spontaneous speech was 37.6%). However, voicing assimilation took place only in 11.1% of all these cases, which is less than the corresponding value in spontaneous speech (that value was 22.8%). These data show that (i) pauses are rarer in repetition than in spontaneous speech, but (ii) in spontaneous speech the speaker is able to "recover" voicing assimilation in spite of pauses more successfully than in a repetition task (Figure 3).

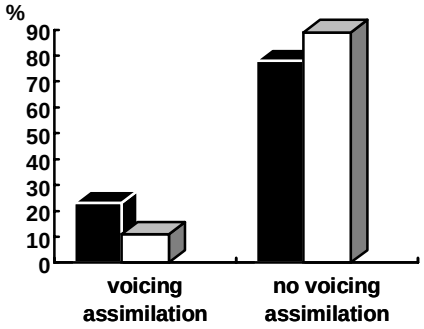

Figure 3. Realization of voicing assimilation across words with pauses in spontaneous speech and speech repetition (black columns=spontaneous speech, white columns=repetitions)

Analysis of the repeated speech material confirmed again that retrieval of voicing assimilation depends on certain temporal patterns. However, pauses occurring during repetition have got a different nature from those of spontaneous speech. These pauses result from mis-synchronization of the heard text, i.e. the speaker's production mechanism is in an expecting phase, generally without any idea what comes next. (If – on the basis of semantic/syntactic context – they are able to figure out the following items, there is no pause or just an extremely short pause between the words.) The duration of pauses was much shorter in repetition than in spontaneous speech. The "synchronization delay" is characteristic of the speakers, however, there are no great differences among the subjects. Similarly enough, all speakers started repeating the heard text with a delay of 580 ms on average. The shortest measured delay was 400 ms while the longest one was 940 ms. People who are good at "shadowing" can repeat back speech with a delay of little more than 250-275 ms.

Voicing assimilation took place when the pause between two words did not exceed 15 ms. If the pause is longer than 15 ms, the word will be activated from the mental lexicon in its stored form without any postlexical phonological alteration. There is no transient phase concerning the durations of pauses. Table 1 shows the duration of pauses resulting in lack of voicing assimilation across speakers. There is no great difference in the lowest values of pauses across speakers, however, there are great differences concerning the highest values of pauses among them. These data confirm that despite the differences of the durations of pauses across speakers, the realization barrier is almost the same for all of them.

Subjects	Mean duration (ms)	Range (ms)
S1(female)	30	17-240
S2 (female)	46	15- 62
S3 (male)	48	18- 70
S4 (female)	151	19-480
S5 (male)	227	20-630

Table 1. Duration of pauses blocking voicing assimilation

Realization of voicing assimilation depends again on pauses that show specific occurrences in the repetition task. As a result, voicing assimilation takes place more frequently in the middle of the utterances than at their beginning or close to their end. Speakers could follow more easily the syntactic structures than semantic ones so, voicing assimilation was performed more successfully in those places where syntactic structure could have been predicted in advance.

5. DISCUSSION

The following conclusion can be drawn for the temporal organization of certain levels of speech production mechanism: postlexical phonological coding should be effective/activated for some time before selecting and finding the next word in the mental lexicon. This period of time is about 300 ms if the production mechanism is working with transformations from thoughts to speaking, while it is only 15 ms if there is no background thinking. Accepting Levelt's hypothesis [3] the speaker's time possibility for microplanning is about 300 ms in order to have a successful phonological coding as well. If microplanning exceeds this period of time the postlexical phonological processes, like voicing assimilation, will not be accomplished.

Results demonstrated that voicing assimilation as a postlexical phonological process is (i) a factor of development, (ii) dependent on temporal factors that define the appropriate operations of microplanning, and (iii) highly predictable when precompiled syntactic structures are available.

Findings of these experiments also touch upon the question of the organization of the mental lexicon. Since Hungarian is an agglitinative language with a very rich morphology, a question might arise concerning the stored forms of the lexical units. On the basis of the experimental data, the double storing of some lexical units is assumed. Morphemes having a final voiced or voiceless obstruent that might be the first member of a consonant cluster might have two forms stored, one with the usual voiced or voiceless final consonant and another one with the counterpart of the "original" obstruent. E.g.: *kalap* 'hat' vs. |kalab| or *téged* 'you' vs. |téget|. This strategy of storing might apply also the suffixes like *–k* 'Plural' vs. |-g|. During phonological coding the appropriate form will be activated from the mental lexicon. This hypothesis needs further investigation together with the nature of phonological coding.

REFERENCES

[1] Chistovich, L. A. and Kozhevnikov, L.A. 1965. *Rech, jazik i vosprijatie*. Leningrad: Akademia Nauk.

[2] Gósy, M. 1998. A beszédtervezés és a beszédkivitelezés paradoxona./On the paradox of speech planning and speech production. *Magyar Nyelvőr,* 122, 3-15.

[3] Levelt, M. 1989. *Speaking. From Intention to Articulation*. Cambridge, Massachusetts: MIT Press.

[4] Turennout, van M., Hagoort, P. and Brown, C.M. 1998. Brain activity during speaking: from syntax to phonology in 40 milliseconds. *Science*, 280, 572-574.