PROCESSING OF DISFLUENCIES AS A FUNCTION OF ERROR TYPE AND AGE

Bóna, Judit[1] – Gósy, Mária[2] – Markó, Alexandra[1]

Eötvös Loránd University[1], Budapest – Research Institute for Linguistics, Hungarian Academy of Sciences[2], Budapest, Hungary

bona.judit@szentignac.hu - gosy@nytud.hu - markoxa@ludens.elte.hu

ABSTRACT

The effects that speakers' disfluencies have on the listener are rather complex. Speech perception is an incredibly fast process, given that while the mechanism interprets the incoming waveform as a series of linguistic segments and suprasegmentals, it is also continuously ready to receive and correct the incoming erroneous messages. The goal of the present experiment was to describe the correction process and determine its efficiency. Various types of disfluency were tested with nine-year-old children, young adults, and elders. The results show that the time span of the corrective process depends upon the type of disfluency, the context, and the listener's age. The higher operational level the production error involves, the more time is required for correcting it and the corrections are poorer than at lower operational levels.

Keywords: disfluencies, correction strategies, age, reaction time, correction success.

1. INTRODUCTION

The way in which speech processing is carried out over time can be modeled on a continuum between two contrasting principles. According to the first principle, speech information occurring first in time is always processed first. According to the second principle, even though the physical variable of time is defined as a before/after sequential order, the information occurring first in time is not always processed first [4]. The fact that processing primacy is not given to temporally early information but to that which is perceptually salient, can be noted when correcting diverse disfluencies. Disfluencies block the flow of acoustic-phonetic information, on the one hand, but they provide time for the listener to process speech, on the other [7]. The correction of disfluencies by the listener might require a much more time-independent procedure [1]. We suppose that there is a continuous backtracking in auditory memory that should immediately be activated in the case of some violation in the incoming speech. The perceptual mechanism of the listener reacts to disfluencies in a very peculiar manner [5].

Listeners are able to correct disfluencies that are errors (like ordering errors, false word activation, grammatical errors) both consciously and unconsciously, however, there are both observed and unobserved errors that cannot be corrected. This "failure" will easily lead to misperceptions and finally to misunderstandings. In addition, there are disfluencies to which the listeners are more sensitive than to others [3, 6]. There is little information, however, about the way and timing of corrective operations and about the interrelationships between the nature of disfluency and its correction success. The correction of the disfluencies seems to be an extraordinarily fast process. In the case of some incoming problem, it is immediately called into operation in parallel with speech processing. We assume that the correction process is ready all the time but inactive until the first alarm indicating that something is wrong. Our hypothesis is that the success and tempo of the correction process heavily depends on the nature of disfluency (type and occurrence) and on the listener's age.

There are three questions that need answers in these respects: (i) Does the occurrence of disfluency in the speech production process have any effect on the success and timing of the correction strategy? (ii) Does the type of disfluency have any effect on the timing and the success of correction? (iii) Does age of the listener influence the correction strategy in any way? A series of experiments have been conducted to seek answers to these questions.

2. MATERIAL, METHOD, SUBJECTS

Nine types of disfluency were selected from the Hungarian Disfluency Corpus [3]: grammatical error (gram.), lexical contamination (contl.) – when two lexemes are fused (e.g., *söci* from *sör* 'beer' and *foci* 'football') – and semantic contamination (conts.) – when two phrases are combined (e. g. *abban szereti jól magát* from *abban szeret lenni* 'he likes to be that way' and *abban érzi jól magát* 'he feels well that way') –, false word activation (word), perseveration (per.), anticipation (ant.), metathesis (met.), articulation error (art.) and error combinations. Each error type was represented by 5 productions that had not originally been corrected by the speaker. Thus, the experimental material consisted of 45 randomly presented utterances containing the above mentioned disfluencies. The isolated sentence-like utterances consisted of 6 words (18 syllables), on average.

A series of experiments were carried out with 60 subjects who fell into three age groups; twenty 9-year-old schoolchildren, twenty elders (66-76 years of age) while the third group included twenty university students and graduates (22-30 years). Half of them in the two younger groups were females, but females outnumbered males in the older group. No participants were excluded from the experimental groups. All subjects were monolingual Hungarian speakers and had normal, age-appropriate hearing. None of them had any speech defect. The subjects' task was to provide a corrected version of the original utterance immediately after hearing. Some listeners repeated only those words or word combinations they corrected while others repeated the whole corrected utterance. The tests were carried out individually and were tape recorded in a quiet room.

The recorded material was digitized and analyzed using Praat (version 4.4) software. Dependent variables included both reaction time (RT) values and response accuracy. Reaction time duration for all speech samples was defined from the end of the last sound of the test utterance to the onset of the first sound of the corrected segment. There was no problem in defining the appropriate acoustic markers for the last or the first sounds of the words concerned.

To test statistical significance, various methods were used (depending on the actual analysis) like match-paired t-test, analysis of variance (ANOVA), and Pearson's correlations (at a 99% confidence level, SPSS version 13.0).

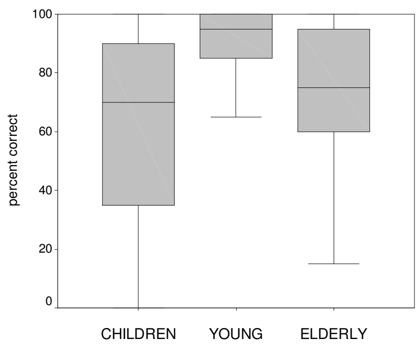
3. RESULTS

All listeners were able to recognize and repeat the original utterances. The two-way ANOVA showed no significant interaction between the factors of 'age' and 'error type'. The listeners were not confused by the presence of disfluencies and the length of the original contexts. Considering the experimental situation, the erroneous messages were not unexpected by the listeners. They were aware of the fact that the utterances would contain errors (listeners are not always prepared to hearing errors in everyday communication situations). So, they might have activated their correction strategies beforehand which is not what typically happens in normal communication situations. On the one hand, then, their correction processes are supposed to have been consciously active from the very beginning, but on the other hand, neither the communicative situation nor the longer contexts could have helped them in correcting the erroneous and relatively short experimental utterances.

3.1. Analysis of correct responses

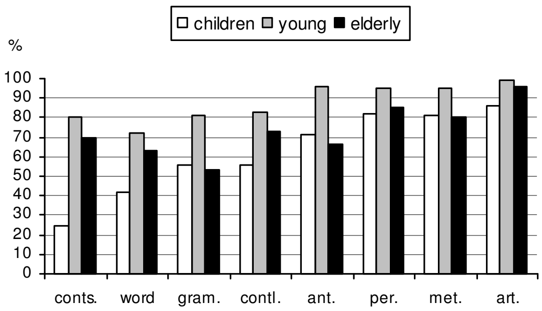
The 'age' factor was significant ($F(2, 132) = 12.663$, $p<0.001$). As expected, young adults were the most successful in the correction of the disfluencies. The older subjects gave fewer correct responses while the children produced the fewest correct answers. The Tukey post hoc test showed no significant differences between the children and the older subjects, while the young adults were significantly better than both groups. The average correct response was 85.6% (of all responses) for the young subjects, 71.9% for the old subjects, and 61.9% for the children. Figure 1 shows the average values of the correct responses and the ranges in each group.

Figure 1: Means and ranges of correct perception of the utterances containing disfluencies.



We assumed that the type of the disfluency would influence the listeners' correct responses. Statistical analysis showed that disfluency type was indeed significant ($F(8, 126) = 4.144$, $p<0.001$). Tukey post hoc testing revealed the following significant differences: between semantic contamination and perseveration as well as articulation error; between false word activation and perseveration as well as articulation error; and between grammatical and articulation error. Figure 2 shows the average correct responses in the three experimental groups. The results suggest that the higher operational levels of speech production (conceptualizing and formulating) at which the production error occurs (like false word activations and semantic contaminations), the worse the correction success is.

Figure 2: Correct perception of utterances with disfluencies according to disfluency types.



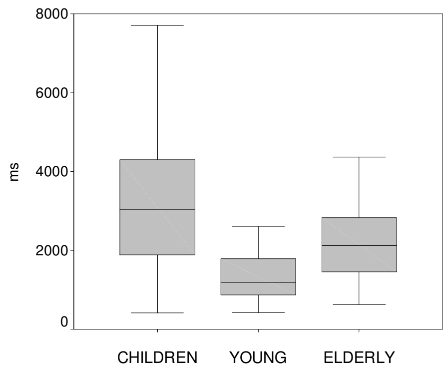
The "simple" articulation errors were corrected almost perfectly by the young people and relatively well by the other two groups as well. The most difficult errors seemed to be the speaker's false word and phrase activations. These errors are connected with the lexical access (how to find the required word in the mental lexicon) and also with the actual lexemes stored there (i.e., word knowledge). We hypothesized that lack of an item in the mental lexicon or failure to activate it might have prevented the correction of these types of disfluency. None of the children seemed to know the word 'disagreeable' and only one of them knew the word 'reconstruct'. The correct processing of ordering errors depended on their type (anticipation, perseveration, or metathesis) with the children and the elderly subjects. The correction of the anticipation errors was much more difficult than that of the perseverations and metatheses. This means that the reordering of segments requires a more holistic kind of processing, which is more difficult for children and elderly subjects.

The utterances containing disfluencies that concern more than one speech planning level (error combinations) were corrected in 89% of all responses by the young adults while only in 61% and 58% by the elderly people and the children, respectively. These findings suggest that young people are able to process these kinds of errors as well as other error types, but children and older people have difficulties in correcting them.

3.2. Analysis of reaction time

As expected, young adults produced the shortest reaction time for correcting the disfluencies appropriately, while the longest time was measured with the children. The 'age' factor was again significant ($F(2, 131) = 25.611$, $p<0.001$). Tukey post hoc testing revealed that all groups performed significantly different from one another (Figure 3).

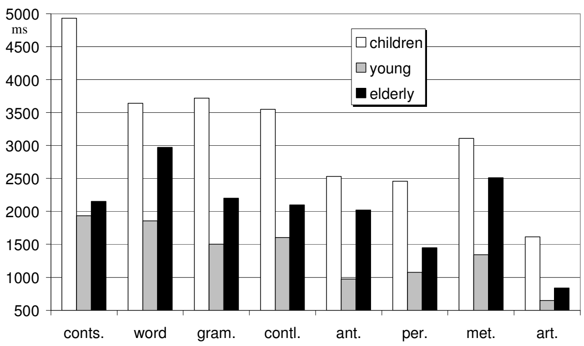
Figure 3: Means and ranges of reaction time values in the correction process.



The average RT was 1410.4 ms for the young subjects and 2111.9 ms for the old subjects while it was 3204.3 ms for the children. The adult listeners of this experiment needed 1761.15 ms (on average) to correct the utterances heard. This duration is equivalent to about 4 two-syllable-long Hungarian words! The timing of children's correction processing is equal to the duration of about 8 words. The question then arises whether the good performance in the corrections of disfluencies could be related to shorter processing time? The correlation analyses revealed a moderate negative correlation (r = -0.477, p<0.0001). This means that the larger the number of positive corrections the shorter the reaction time is. If we look at the data of each group separately, we find that there is a mild correlation between positive corrections and reaction time values in the case of the children and the elderly (r = -0.312) while there is a stronger one in the case of the young subjects (r = -0.696). This finding can be interpreted as follows: even though children and older adults allow themselves more time, this does not provide them with sufficient help to guarantee success in the correction processes. There are enormous differences in the reaction time values between the experimental groups. The long reaction time values might result in further misperceptions and misunderstanding of the utterances which is frequently the case with children and older people.

Analyses were carried out concerning the error types and the RT values for their corrections (Figure 4). The tendency is essentially similar across all groups. The longest reaction time was associated with correcting false word activations and semantic contaminations, while the correction of the articulation errors required the shortest time.

Figure 4: Reaction time values according to type of disfluency.



The reaction time values shorter than 100 ms mean that the corrections were processed in parallel with the semantic and syntactic processing of the utterances. On the other hand, occurrence of a segment in the wrong place (as in the case of anticipations and perseverations) did not present a serious problem for the young subjects. Nevertheless to correct metatheses (i.e., spoonerisms), young adults needed more time, about 350 ms on average. The children and the elderly had trouble both with anticipations and particularly spoonerisms; however, they seem to have had fewer problems with perseverations. Anticipations are regarded as 'good errors' in the literature [2] since the planning process does not stop short in these cases (as it does in the case of perseverations). For the elderly to correct the latter was a faster and more successful process. From this we can conclude that speech information occurring first in time is always processed first both with the old subjects and the children; but this is not the case with the young subjects.

As could be seen earlier, the correction of the grammatical errors was less successful and required more time than previously believed. This raises the question of whether the listener does not, in general, tend to analyze the utterance properly according to its grammatical structure or, if the semantic processing might suppress the need for detailed grammatical analysis. This might explain failure to correct some of the morphological and syntactic errors.

There were no significant differences between genders, either in successfully correcting disfluencies or in reaction times.

4. CONCLUSIONS

The current experiment was designed to explore the nature of correction processes in erroneous utterances by adult and child listeners. The strategies noted in disfluency corrections support the conclusion that (i) late information can be processed earlier in time particularly by the young subjects, (ii) segmental and suprasegmental speech processing and the necessary corrections can occur in parallel, and (iii) the data provide convincing evidence that the corrections and their timing depend on the type of disfluency and on the place where it occurs in the speech planning mechanism. The brain processes responsible for speech production have massive parallel capacity [2]. The listener's correction strategy has to deal with continuous input of both perfect and imperfect speech, and should be able to maximize its chances to successfully find the appropriate lexemes, suffixes, structures, etc. to restore the speaker's original speaking intention within an appropriate time structure in the case of errors.

5. REFERENCES

[1] Hartsuiker, R. J. – Kolk, H. J. 2001. Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology* 42, 113-157.

[2] Howell, P. 2003. Is a perceptual monitor needed to explain how speech errors are repaired? In Eklund, R. (ed.), *Proceedings of DISS '03*. Göteborg: Göteborg University, 31-34.

[3] Hungarian Disfluency Corpus. 2004. In: Gósy, M. (ed.), *Beszédkutatás*. Budapest: MTA, 19-187.

[4] Mattys, S. L. 1997. The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin and Review* 4, 310-329.

[5] Postma, A. 2000. Detection of errors during speech production: A review of speech monitoring models. *Cognition* 77, 97-131.

[6] Pouplier, M. – Goldstein, L. 2002. Asymmetries in speech errors and their implications for understanding. *ZAS Papers in Linguistics* 28, 73-82.

[7] Susca, M. – Healey, Ch. E. 2002. Listener perceptions along a fluency-disfluency continuum: A phenomenological analysis. *Journal of Fluency Disorders* 27, 135-161.