UDK 811.511.141'342.2
159.943.3:81'34
Izvorni znanstveni rad

Mária Gósy
Research Institute for Linguistics, HAS and Eötvös Loránd University, Budapest
Hungary

DISFLUENCIES AND SELF-MONITORING

SUMMARY

One of the major questions in speech science has been how are language production and comprehension related. Both processes require rapid and accurate retrieval of words and appropriate grammatical structures from the mental lexicon. Theories propose that word retrieval involves the selection of at least two levels of lexical information, semantic representations and word-forms. Spontaneous speech is characterized by various phonetic processes such as co-articulation, the variability of the phonetic form of words, and by various types of disfluency phenomena. As speakers we come across difficulties during speech planning while as listeners we have to cope with other peoples' speech difficulties resulting in erroneous utterances. Speech disfluencies are generally defined as phenomena that interrupt the flow of speech and do not add propositional contents to an utterance. There are various forms of disfluencies that occur and might slightly differ across languages.

This talk will discuss the results of a series of experiments that have been carried out with Hungarian speakers/listeners focusing on speech disfluencies. A corpus containing more than 5 000 perceived disfluency items was analyzed while recorded spontaneous speech material of about 8 hours was analysed. Several specific experiments were also carried out to investigate the problem with the participation of young people, children and elderly persons.

The following questions were posed. (i) How do the various types and occurrences of disfluency highlight the difficulties the speaker faces during speaking? (ii) What kinds of interrelations exist between the phonetic and phonologocial operations resulting in errors during speech planning? (iii) Do pauses refer to specific operations in the mental lexicon predicting the phonetic output? The temporal analysis of word retrieval was carried out in a "tip-of-the-tongue" elicitation experiment while pauses signaling the speaker's word-finding difficulties were measured also in spontaneous speech. (iv) How does the *speaker's self-monitoring work? What kinds of strategies lie behind the diverse outcomes of the self-monitoring process? There are disfluencies that the speaker fails to notice, on the one hand, and there are disfluencies that are noticed by the speaker but still remain uncorrected, on the other. There are yet other disfluencies that are corrected during speaking without the action being noticed by the speaker. (v) Are there any differences among the types and occurrences of disfluency depending on the age of the speaker? (vi) What kinds of effects do speakers' disfluencies have on listeners? Speech perception is an extremely 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 incoming erroneous messages. How are speakers able to monitor their own speech while listening at the same time to another's speech (using a shadowing technique) or while being disturbed by background speech noise (Lombard-effect)? What kind of strategy helps speakers to comprehend erroneous speech? Our findings enable a hypothesis to be formulated explaining the differences in speech disfluencies (both in production and perception) depending on their type, the speaker's age, the actual context and the speech planning level where they occur.*

Key words: *disfluency, self-correction (phonetics), speech production, phonetics, Hungarian language*

INTRODUCTION

For quite some time now, the relationship between language production and comprehension has been a hot issue in speech science. Both processes require rapid and accurate retrieval of words and grammatical structures from the mental lexicon. Most theories propose that word retrieval involves the selection of at least two levels of lexical information, semantic representation and word-forms. Spontaneous speech is characterized by several phonetic processes, such as co-articulation and variability of the phonetic form of words, as well as various types of disfluency phenomena. As speakers we experience difficulty in speech planning, and as listeners we have to cope with disfluencies in other people's speech. Speech disfluencies are generally defined as phenomena that interrupt the flow of speech and do not add propositional content to an utterance. There are various forms of disfluencies, and these might show some slight differences across languages. An enormous amount of research has been reported on since the classical works of Frieda Goldman-Eisler (1968) and Victoria Fromkin (1973), although it should be noted that speech production difficulties drew the attention of scholars much earlier. The Arab linguist Al-Ki-sa'i collected and published speech errors as early as the 8th century. Levelt's work (1983, 1989) has provided a significant impetus to research on disfluencies in spontaneous speech.

When we produce an utterance corresponding to some thought we wish to convey, we cannot just open our mental storage unit and find a message there ready for use. In speaking spontaneously we produce novel utterances, most of which have never been said by ourselves or others in exactly the same form. However, our thoughts do not always emerge in correct grammatical, phonological and phonetic forms. Various errors in the speech production process provide valuable insights into the units of speech production and into the mental planning process of fluent speech. The question is how successfully speakers transform their thoughts into well-formed grammatical structures.

Spontaneous speech varies not only in the amount and frequency of disfluencies it contains but also in the types that the given speaker produces. The question arises whether any language specific tendencies can be discovered in the incidence of various types of disfluencies in fluent speech. Research in English and German corpora showed similar tendencies in the incidence of most of the disfluencies that were studied in the two languages (Hieke, 1981). What is interesting is the considerable difference found in the amount of phonological repair in English (14.82% of all disfluencies) as compared to German (5.26% of all disfluencies). Another important difference was found in the number of syntactic errors (10.53% in German against 18.52% in English). In analyzing self-repair in Croatian it was found that the most frequent types were morphological and syntactic errors, wrong word retrievals and errors in executing the articulatory program, while phonological errors were infrequent (Horga, 1997). Disfluencies are supposed to be universal in the sense that they are the consequences of speech production processes; however, it is possible that in analyzing this phenomenon, in addition to speakers' behaviour language structure should also be taken into consideration.

Listeners can perceive speakers' disfluencies and speakers are also able to perceive their own disfluencies when speaking. Self-repair of speech errors demonstrates that speakers possess a monitoring device with which they can check the correctness of the speech flow. There is substantial evidence that speakers monitor their speech when speaking. This allows them to correct their speech errors. This self-monitoring works not only on the basis of acoustic feedback (i. e., hearing one's own speech), but also during speech planning prior to its motoric execution. Errors thus may be detected before they are actually articulated. The theory of speech monitoring was first proposed by Levelt (1983). According to this, self-repair in speech typically proceeds in three phases. The process starts with monitoring one's own speech and then the flow of speech is interrupted when trouble is detected. The second phase is characterized by hesitation, pausing, lengthening; these are called editing terms. The third phase consists of making the repair proper. Self-monitoring is probably based on parsing one's own inner or overt speech though there is little known about the division of labor between the two.

Speakers monitor their own speech through two routes, an external and an internal monitoring route. This monitoring process involves checking the ongoing speech planning and execution by means of detecting errors. The external route (via overt speech) is based on the speech perception and comprehension mechanism. The internal route (via covert speech) accesses primarily the phonological representations of the planned utterance. In other words, speakers monitor what they will say and what they have just said. They must keep records of their utterances over several seconds or even minutes to be able to continue talking. The "covert repair hypothesis" (Postma & Kolk, 1993) claims that disfluencies reflect the covert, prearticulatory repair of speech programming errors. Experimental data have confirmed that including an inner loop through the speech comprehension system generates predictions that fit the empirical data (Hartsuiker & Kolk, 2001). In the covert repair hypothesis the internal error detection possibility has been extended with an internal correction counterpart. Basically, the covert repair hypothesis contends that disfluencies reflect the interfering side-effects of covert, prearticulatory repair of speech programming errors in ongoing speech. Internally detecting and correcting an error obstructs the concurrent articulation in such a manner that a disfluent speech event will result. Furthermore, it is shown how, by combining a small number of typical overt self-repair features such as interrupting after error detection, retracing an utterance, and marking the correction with editing terms, one can parsimoniously account for the specific forms disfluencies are known to take.

In Hungarian, a pause occurring within a word seems to be a characteristic type of speech error (Gósy, 2005). It may indicate a semantic problem in selecting a constituent within a compound word or after a prefix or before a suffix, or uncertainty about a selected morpheme. The pause interrupting a meaningful unit in fluent speech is a marker of error detection by the speaker, providing evidence for internal speech monitoring. Pauses accounted for 10.38% of all detected speech errors in an off-line corpus of speech recorded from 18 speakers (Gósy, 2007). The number of syllables ranged from 2 to 8 syllables. Syllabification in Hungarian is so strongly encoded that the speaker corrects the syllable structure if the interrupting pause appears at the wrong speech sound, i.e. not at a syllable boundary. For example: *szós-* (a pause of 579 ms) *statisztikáját* ("his word statistics (acc.)"). The speaker uttered the first syllable *szó* together with the first consonant of the next syllable of the compound word. Realizing that he stopped within a syllable before continuing the articulation after the pause, the speaker repeats the initial consonant in order to recover the correct syllable. An attempt to restore correct syllable structure is seen in all cases where a pause interrupts the articulation of a word, which can be explained by assuming that phonemes are retrieved from the lexicon with their syllabic positions within the word specified.

Pauses within a word in Hungarian occur mainly at morpheme boundaries (before a suffix, after a prefix, at the boundary of two constituents in a compound), but practically, they can appear at any syllable boundary within a stem. The duration of these pauses seems to be dependent on the word retrieval process. The mean duration of pauses (signaling lexical retrieval difficulties) within a word between the prefix and the stem is 343.2 ms (e. g. *meg* /pause/ *látták* "/they prefix/ saw"). The mean value of pauses within a compound is 290.83 ms (e. g. *hát* /pause/ *térrel* "/with/ background"), between a stem and a suffix is 170.66 ms (e. g. *kutatás* /pause/ *nak* "for research") while within a stem it is 107.33 ms (e. g. *har* /pause/ *madik* "third"). The occurrence of the pause provides evidence for the functioning of the internal monitoring route for error detection. The longest pause duration occurred between the prefix and the stem, which shows either that the speaker had a problem in word finding, or in the editing phase. A similar difficulty in speech planning occurs with compounds, when the first constituent of the lexeme has already been produced, but the second constituent is "skipped": apparently, the speaker is in a kind of tip-of-the--tongue condition.

What happens if the speakers' external route of self-monitoring is strongly disturbed? Does internal self-monitoring take over the function, as far as it can? How exactly do speakers access the speech planning process through the two monitoring routes? Our hypothesis is that the internal route provides information on phonological planning when the external route does not work properly. Thus the question arises whether speakers follow the same strategy/strategies when they have one route available as when both routes are open.

This paper will discuss the results of a series of experiments, focused on the monitoring processes of speech disfluencies, carried out with Hungarian speakers/listeners.

SUBJECTS, METHODS AND MATERIALS

Experiment 1. A corpus containing more than 5 000 perceived disfluency items (Gósy, 2004) and recorded spontaneous narrative speech material from 18 adult speakers of about 6.5 hours were analyzed. Disfluencies were defined and categorized using the same criteria. Comparisons were made concerning the types and occurrences as well as the ratios of self-repairs between the two corpora.

Experiment 2. Twenty Hungarian monolingual subjects (10 females and 10 males, between the ages of 24 and 35, mean age 28 years) participated in the experiment. All of them had normal hearing and no speech defects. They were individually tested in a silent chamber. A short story of 3.5 s told by a native Hungarian female served as the input material. This text was manipulated by the random insertion of various disfluencies. Participants were asked to repeat the text they heard as quickly as they could (shadowing). They heard the text through headphones at a comfortable intensity level and were aware of the fact that their speech was going to be recorded. Because of the listening task and of the headphones it was impossible for them to monitor their own voices.

There were two types of shadowers: close shadowers with a latency of about 250-300 ms (8 subjects), and distant shadowers with a latency of 500 ms or more (12 subjects). Ten questions were asked immediately after the completion of the shadowing task in order to check whether the semantic and syntactic levels of the speech perception mechanism had been involved. The subjects were not informed previously about this part of the task. Questions concerned (i) the details of the story and (ii) the main idea of the text. The average of correct answers was 22% of all responses. Seven subjects could not answer any questions, and one of them was able to answer correctly six questions. These results show that the subjects were unable to use the upper levels of their speech decoding processes during shadowing (cf. Marslen-Wilson, 1985). Close shadowers were able to produce on-line speech analysis before they were fully aware of what words they were reproducing. In other words, they were relatively unaffected by post-perceptual processes. Some speakers also showed sensitivity to variations in the frequency of particular phoneme sequences in the language (Nye & Fowler, 2003). The shadowers' disfluencies and self-repairs were analyzed and compared with the repairing process of the original speaker's disfluencies.

Experiment 3. There is another situation in which the speaker is forced to use both routes for monitoring because of a surrounding noise. In this case we may assume that the incidence and types of speech errors will be different from those found in typical speech conditions. Samples of two-minute spontaneous speech were recorded under two conditions: "no noise" and "noise present". The subjects were asked to speak about their work and hobbies. They were not warned that noise conditions were going to be changed in the course of the experiment. However, they were given clear instructions to continue speaking irrespective of what was happening around them. One minute after they started talking, a loudspeaker transmitting background noise was switched on. The average speech-and-noise ratio was set at 30 dB. Previously recorded human conversational babble (with the participation of 5 subjects) was used as background noise for several reasons, including its naturalness, people's familiarity with it, and its spectral qualities (see Fig. 1).

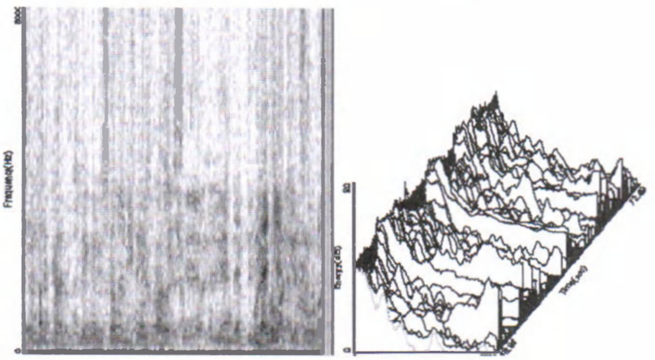

Figure 1. Spectrogram and intensity of background noise used in the experiment

Slika 1. Spektrogram i intenzitet pozadinske buke koja je korištena u eksperimentu

The subjects' spontaneous speech was carefully recorded in both conditions so that the material would be suitable for further acoustic phonetic analysis. The digital recordings were submitted to analysis (by Praat 4.04) using a 22 kHz sampling rate. Statistical analysis was carried out using the SPSS program for Windows (version 8.0). Eight females and seven males, altogether 15 subjects took part in the experiment (ages between 20 and 30). They had normal hearing and none of them showed any speech disability. Their recorded speech was analyzed for the incidence and types of disfluencies as well as self--repair. To test statistical significance, various methods were used, such as match-paired t-test, analysis of variance (ANOVA), as appropriate (using SPSS 8.0).

RESULTS OF THE EXPERIMENTS

Speech errors collected from on-line and off-line corpora are different (Fromkin, 1973). The speech errors collected by means of an on-line corpus are based on the listener' speech perception and comprehension, while those collected by means of an off-line corpus are based on listening to someone's recorded speech. The off-line corpus contains all the disfluencies that a speaker produced during speaking, and the listener can listen to the speech samples as many times as needed. In this way, speech errors from an on-line corpus show what is important for the listener, while those from an off-line corpus show the speaker's actual disfluencies. We assumed that there will be differences in the ratio of disfluencies depending on the corpora analyzed, but we wanted to obtain information on their exact ratios. Figure 2 summarizes the speech errors found in an on-line and an off-line Hungarian corpus.

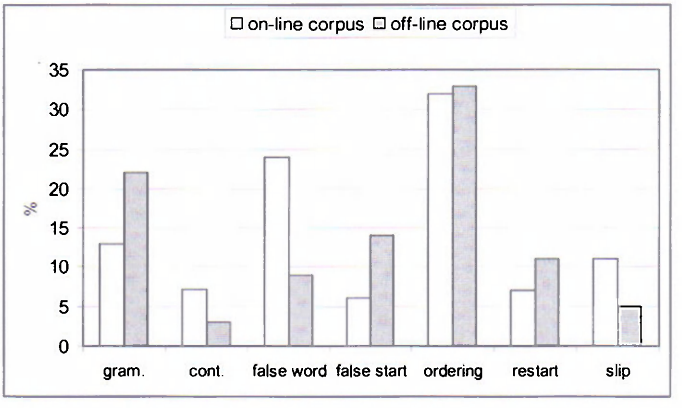

Figure 2. The ratio of perceived and actually produced disfluencies in spontaneous speech

Slika 2. Odnos percipiranih i stvarno proizvedenih disfluentnosti u spontanom govoru

Listeners are bound to notice false words, slips and contaminations: these disfluencies are very conspicuous, although not very frequent in spontaneous speech production. Listeners will also recognize ordering errors (anticipations, perseverations and metatheses), which are more frequent. On the other hand, they will tolerate a number of grammatical errors, false starts and restarts; indeed, such disfluencies often get past the listener's monitor. This means that the listener is sensitive to those errors that make his/her comprehension difficult but does not notice those that either provide an opportunity for slower speech processing but do not cause difficulty for comprehension (like incomplete words) or can easily be repaired. These data represent the monitoring results of the listener or, in other words, the outcomes of external monitoring in the case of the on-line data, and the functioning of both routes in the case of the off-line data.

The shadowing task was intended to eliminate the external monitoring route so that the internal monitoring route can be observed. It was assumed that in this way the use of top-down production strategies will be reduced. The question is how the listener is able to process (i) his/her own possible speech errors during shadowing and (ii) the inserted artificial disfluencies in the recorded speech material. Table 1 shows how the listeners coped with this double challenge.

Table 1. How listeners cope with speech errors during shadowing
Tablica 1. Kako se slušači nose s govornim pogreškama prilikom simultanog ponavljanja tuđeg govora

Original	Shadowers' speech production
correct	Correct (no error occurs)
	incorrect: false but real words, incorrectly articulated words and nonwords occur
incorrect	correct: self-monitoring alarms, repair process takes place
	incorrect: self-monitoring does not alarm, no repair occurs

Shadowers produced speech errors of which, apparently, they were not aware, even where the original speech sample did not contain any disfluency (Fig. 3).

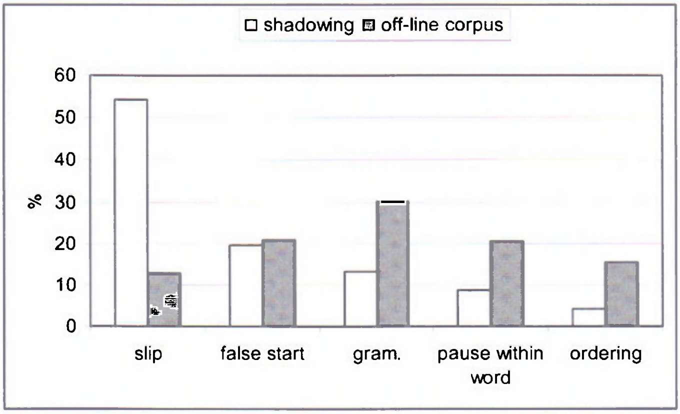

Figure 3. Ratio of error type disfluencies in shadowing and in the off-line corpus
Slika 3. Postotak različitih vrsta disfluentnosti koje se javljaju tijekom simultanog ponavljanja tuđeg govora i u off-line korpusu

There was an unusually large number of slips in the shadowers' speech production compared with data from the off-line corpus. There was no difference in the incidence of false starts, while all the other types of disfluencies showed considerable differences in the two data sources. Thus it can be concluded that the special task of shadowing renders the speakers' internal monitoring ineffective inasmuch they become unable to repair false phonological plans before executing.

Due to semantic and morphological factors, shadowers produced some existing but inappropriate words. Phonetically, they differed from the original words in one or a few speech sounds. E. g. original *kapták* ("/they/ got") was shadowed as *megkapták* ("/they/ [prefix]+got"); original *lakó* ("resident"), was shadowed as *rakó* ("putting"), original *vonulásukat* ("/their/ procession+Acc.") was shadowed as *vallásukat* ("/their/ religion+Acc."), original *lefejeztette* ("/he had him/ beheaded") was shadowed as *lefejezte* ("/he/ beheaded him"). Shadowers also produced a great many unrepaired nonwords that were phonetically similar to those of the original words, such as *szpáhik* ~ *spáhik*, *megeskette* ~ *megekskettl*, *tiltakoztak* ~ *ittélkoztak*, *harcba* ~ *harban*. All these speech errors show evidence that the internal route of monitoring was disrupted. Hartsuiker and his colleagues claim that lexical errors appear to be detected predominantly by the internal channel while phonological errors are detected by both channels with equal accuracy (2001, 2005). Our data seem to support this claim, since there were more undetected phonological errors than undetected lexical errors in the shadowed speech. Repair of errors of the latter type would have required external monitoring, but this was unavailable to the participants since they were engaged in continuous shadowing.

The next experiment was designed to explore the types and incidence of disfluencies in both silent and noisy conditions. All the disfluencies of the uncertainty type, characteristic of spontaneous speech, were found in the narratives of the present experiment, independently of silent or noisy conditions (filled pauses, fillers, repetitions, etc., cf. Fig. 4). However, their incidence significantly increased when noise was present during speaking (one-way ANOVA: $F(1, 29) = 12{,}104$, $p < 0.002$).

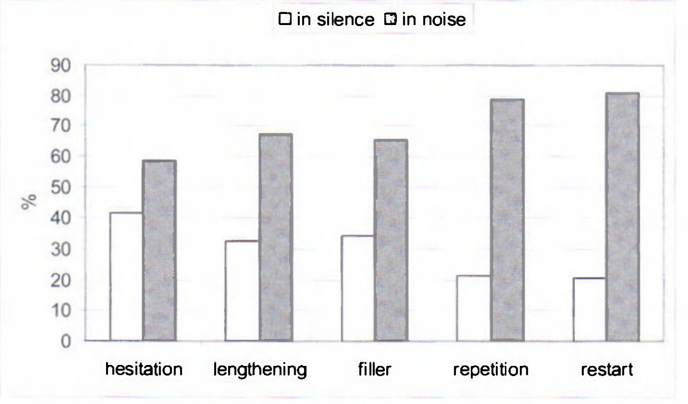

Figure 4. Uncertainty type disfluencies in both conditions
Slika 4. Postotak pojavljivanja odsječaka disfluentnosti iz kategorije nesigurnosti u tihom okruženju i u buci

All the possible speech errors described in the literature could not be traced in the narratives independently of silent or noisy conditions, but they had a higher incidence when noise was present, similarly to disfluencies of the uncertainty-type. What is interesting here is that grammatical and ordering errors occurred only under noisy conditions (Fig. 5). The ratio of all disfluencies in the experimental material was 66.2% in noise and 33.8% under silent conditions. Of all the disfluencies in the silent condition, 28.4% were categorized as belonging to the uncertainty type, while in the noisy condition 50.2% were of this type. These data support the view that speakers experience more difficulty in transforming their thoughts into grammatical forms when noise is present.

There were four types of speech errors in the narratives produced in noise and three types under silent conditions. 5.2% of all speech errors were found in the silent condition and 16.0% when there was a noisy background. This means that speakers in the latter condition committed three times more speech errors than they do normally. Twice as many grammatical errors and slips were found in the narratives spoken in noise. The ratio of false starts and anticipations increased substantially under noisy conditions. These data suggest that retrieval from the mental lexicon is disturbed, on the one hand (false starts), and the controlling of the order of speech sounds and words (anticipations) becomes problematic, on the other hand.

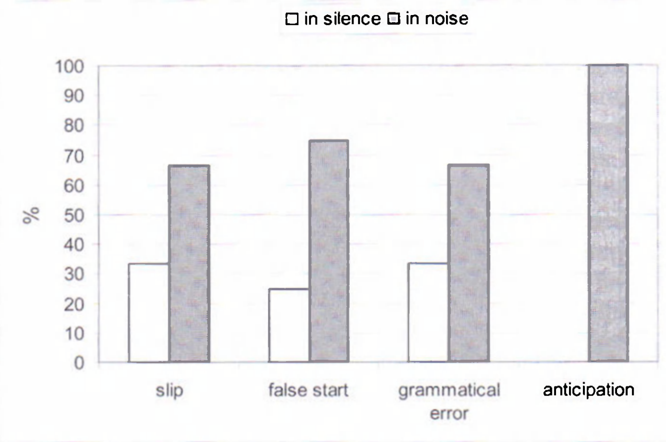

Figure 5. Error type disfluencies under both conditions
Slika 5. Postotak pojavljivanja disfluentnosti iz kategorije govornih pogrešaka u tihom okruženju i u buci

The subjects produced 1.41 disfluencies per person on average under silent and 2.31 disfluencies per person under noisy conditions. The difference is statistically significant (one-way ANOVA: $F(1, 29)$ 12,104, $p < 0.002$). The most frequent type of uncertainty disfluency under silent conditions was hesitation (41.4%) while restarts (80.9%) and repetitions (78.8%) accounted for most disfluencies under noisy conditions. Repetitions (21.2%) and restarts (20.6%) were the most infrequent types under silent conditions and hesitations (58.6%) and fillers (65.5%) occurred relatively seldom under noisy conditions. These data suggest that speakers' strategies during the speech planning process are different when noise is present. The speakers restart the words and repeat them when there is a mismatch between planning and execution in order to save their planning mechanism from intruding (disturbing) noise. When there is no noise in the background, speakers do not need this kind of "escape strategy", therefore they will tend to hesitate, use fillers and lengthen some speech sounds instead of repeating or restarting words.

Subjects show substantial individual differences in the types and incidence of disfluencies in their narratives. There was only one speaker who had fewer disfluencies in noise, while all the others produced more under noisy conditions (Fig. 6).

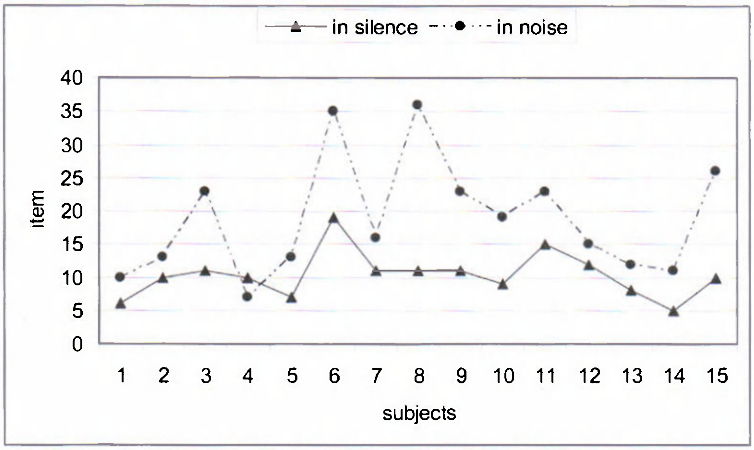

Figure 6. Number of disfluencies across speakers in silence and in noise
Slika 6. Broj disfluentnosti kod pojedinog govornika u tihom okruženju i u buci

The number of disfluencies depending on the experimental conditions varies widely across speakers. The number of disfluencies showed a sharp increase with nine subjects. The highest number of disfluencies was produced by a male speaker (36 items). This means that the subject's speech in noise contained some interruption at every 1.6 seconds while only at every 5.4 seconds in silence.

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 erroneous messages. In the shadowing experiment, 68.4% of all speech errors of the slip type (phonetically distorted existing Hungarian words) were shadowed correctly. This can be explained either by assuming that syntactic and semantic information is accessed, resulting in successful lexical retrieval despite the erroneous phonological structure of the word, or by assuming that the internal phonological monitoring process made the repair of speech errors possible without accessing higher perception levels. However, 31.6% of the distorted words were repeated unchanged. This shows that the monitoring processes did not alarm the "listener" about the error of the incoming speech flow or the "speaker" (who is the same person in this case) about the erroneous sound sequence s/he was going to articulate. We can conclude that roughly one-third of the perceived speech errors cannot be repaired for some reason. This figure is supported by another experiment where the listeners' reparation success was analyzed by means of immediate correcting of the erroneous utterances (Bóna et al., 2007). The average correct response was 85.6% (of all responses) and the results showed that the corrections were better when the original error occurred at lower operational levels of the speech production mechanism (e. g. slips as opposed to false word activations).

Analysis of speech error correction in the on-line corpus showed that almost half of all errors (51.8% of all) were repaired (Fig. 7). The tendency of correction rates for perseverations, anticipations and metatheses were similar to those found in a Dutch on-line corpus (Nooteboom, 2005).

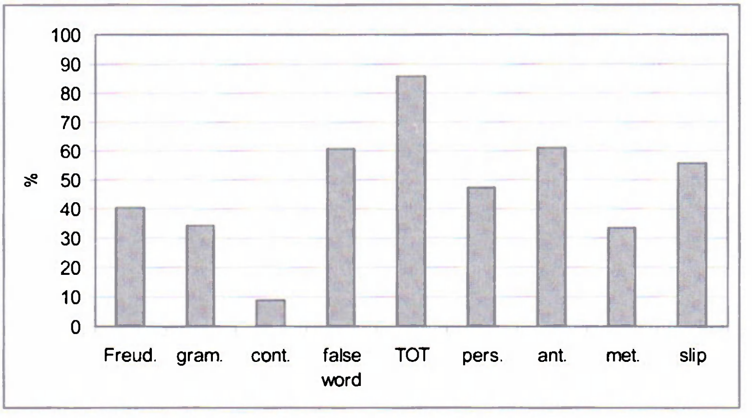

Figure 7. Corrections of various error types in the on-line corpus
(Freud. = Freudian slips, gram. = grammatical errors, cont. = contaminations, TOT = tip-of-the-tongue phenomena, pers. = perseveration, ant. = anticipation, met. = metathesis, slip = slip of the tongue)

Slika 7. Ispravljanje različitih vrsta pogrešaka u on-line korpusu
(Freud. = Freudovske omaške, gram. = gramatičke pogreške, cont. = kontaminacijske pogreške, TOT = fenomen *na vrhu jezika*, pers. = perseveracijske pogreške, ant. = anticipacijske pogreške, met. = metateza, slip = lapsus linguae (*slip of the tongue*))

It was assumed that repairing of one's own speech errors provides evidence mainly for the internal monitoring operations. If this is the case, repairing of one's own speech errors should be more successful when both monitoring routes are available. Nooteboom claims that in his corpus of Dutch speech errors roughly 50% of all errors were detected and corrected by the speakers (2005). Comparisons were made between self-corrections in shadowed speech, in speech under noisy conditions and in normal spontaneous speech (Fig. 8). (In the case of shadowing, in addition to corrected and uncorrected speech errors there were also some omitted errors.)

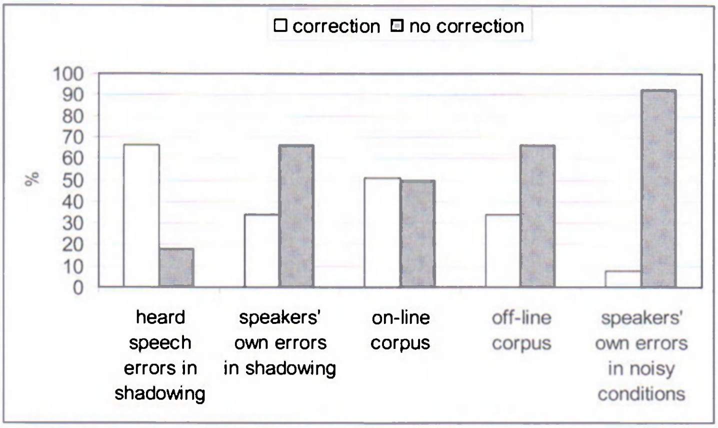

Figure 8. Self-correction ratios across various situations
Slika 8. Pojava samo-ispravljanja u različitim situacijama

In the shadowing task the subjects corrected 66% of all errors they heard in recorded speech, while they were able to correct only 34% of their own speech errors. The internal route is thus less effective in self-repair than the external route in repairing others' speech errors. The on-line corpus provides evidence that listeners are sensitive to both corrected and uncorrected speech errors, while the analysis of the off-line corpus supports that speakers do not correct the majority of their own speech errors for some reason (66.3%). This does not mean, however, that the speaker does not notice the error. This shows only that reparation does not take place. Finally, corrections of the speakers' own errors in noisy conditions were analyzed. In this condition speakers corrected a very small number of their own errors (below 10%). The accuracy of error detection seems to deteriorate under the influence of noise, resulting in an extremely poor performance of the monitoring strategies. The same results were obtained with Dutch speakers in self-repair of phonological and lexical errors (Postma & Noordanus, 1996). It should be noted here that speakers were much better at correcting other speakers' errors than the errors they themselves committed. Our statistical analysis showed a significant difference in the number of corrected speech errors in diverse communication situations (one-way ANOVA: $F(4, 249) = 8{,}316$, $p < 0.000$).

CONCLUSION

The data in our study suggest that some parts of the speech planning mechanism are more involved in self-repair than others, dependent on the given communication situation. The ratio of disfluencies increased under noisy conditions and during shadowing; however, their number as well as their types varied across speakers. The "covert repair hypothesis" holds that the speaker is able to inspect his/her speech program prior to its motoric execution, and if so, s/he must be aware of impending speech errors before they are actually articulated. Although prearticulatory repair of speech programming is possible, it is questionable how effective these operations are under various speech conditions. There is another factor that must be considered in explaining the self--monitoring processes, namely how we classify speech errors. Some errors may have a phonological source, but they may also be due to a lexical bias. For example, errors that yield existing words are considered phonological errors, while false words that allow of a plausible lexical explanation in the context are called lexical speech errors. However, there are cases when it is impossible to identify an error based solely on surface analysis. There is no acceptable solution for the treatment of erroneous suffixes, which is relatively frequent in an agglutinative language like Hungarian.

Most theories of self-monitoring agree that speakers detect speech errors through at least two channels: overt speech (the external channel) and internal speech (the internal channel). Less is known, however, about their relative contributions. Levelt assumes that the same self-monitoring mechanisms are used in both monitoring routes (1989). Nooteboom claims that self-correction of overt speech errors is obviously perception-based, but he concedes that the speaker may also have some means of accessing the intended phonological form of erroneously produced words (2005). Experimental data from diverse speech materials (speech under noisy conditions and aphasic speech) seem to support the view that the internal channel is more effective than the external channel (Hartsuiker et al., 2005). The authors referred to hypothesize that the division of labor between the two channels is controlled from the top down, but there are also bottom-up influences. The question is how many and which errors are detected by each channel? Covert repairs are not exclusively initiated by internal monitoring and overt repairs by external monitoring. Therefore, listening to one's own errors is different from listening to other speakers' speech errors. The timing analysis of the repairs shows that the internal channel contributes to a significant proportion of overt self-repairs (Oomen & Postma, 2001). Our present data, based on a shadowing experiment, appear to support that the operations of the two monitoring channels may be qualitatively different.

REFERENCES

Bóna, J., Gósy, M., Markó, A. (2007). Processing of disfluencies as a function of error type and age. *Proc. of the XVI ICPhS.*, 1333-1336.

Fromkin, V. A. (ed.) (1973). *Speech Errors as Linguistic Evidence.* The Hague: Mouton.

Fromkin, V. A. (ed.) (1980). *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand.* New York: Academic Press.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous speech*. New York: Academic Press.

Gósy, M. (ed.) (2004). "Nyelvbotlás"-korpusz. Spontán magyar beszéd hallás alapú megakadásainak gyűjteménye. *Beszédkutatás*, 1-100.

Gósy, M. (2005). *Pszicholingvisztika*. Budapest: Osiris.

Gósy, M. (2007). Alternative routes of lexical retrieval in spontaneous speech. *Studia Slavica Hung* **52**, 1-2, 129-136.

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

Hartsuiker, R. J., Kolk, H. H. J., Martensen, H. (2005). Division of labor between internal and external speech monitoring. In R. J. Hartsuiker, R. Bastiaanse, A. Postam, F. Wijnen (eds.), *Phonological Encoding and Monitoring in Normal and Pathological Speech*. Cornwall: Psychology Press.

Hieke, A. E. (1981). A content-processing view of hesitation phenomena. *Language and Speech* **24**, 147-161.

Horga, D. (1997). Samoispravljanje u govornoj proizvodnji. *Suvremena lingvistika* **19**, 91-104.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition* **14**, 41--104.

Levelt, W. J. M. (1989). *Speaking. From Intention to Articulation*. Cambridge, Massachusetts: MIT Press.

Marslen-Wilson, W. D. (1985). Speech shasowing and speech comprehension. *Speech Communication* **4**, 55-73.

Nooteboom, S. (2005). Listening to one-self: Monitoring speech production. In R. J. Hartsuiker, R. Bastiaanse, A. Postam, F. Wijnen (eds.), *Phonological Encoding and Monitoring in Normal and Pathological Speech*. Cornwall: Psychology Press, 167-186.

Oomen, C. C. E., Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Reseach* **30**, 163-184.

Nye, P. W., Fowler, C. A. (2003). Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *Journal of Phonetics* **31**, 63-81.

Postma, A., Kolk, H. H. J. (1992). Error monitoring in people who stutter. evidence against auditory feedback defect theories. *Journal of Speech and Hearing Research* **35**, 1024-1032.

Postma, A., Kolk, H. H. J. (1993). The covert repair hypothesis: prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research* **36**, 472-87.

Postma, A., Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech* **39**, 375-392.

Postma, A., Oomen, C. C. E. (2005). Critical issues in speech monitoring. In R. J. Hartsuiker, R. Bastiaanse, A. Postam, F. Wijnen (eds.), *Phonological Encoding and Monitoring in Normal and Pathological Speech.* Cornwall: Psychology Press, 157-167.

Mária Gósy
Institut za lingvistiku, Mađarska akademija znanosti i Sveučilište Eötvös Loránd, Budimpešta, Mađarska

DISFLUENTNOSTI I PRAĆENJE VLASTITOG GOVORA

SAŽETAK

Jedno od glavnih pitanja u znanosti o govoru jest na koji su način povezani jezična produkcija i razumijevanje. Oba procesa zahtijevaju brzo i točno prizivanje riječi i primjerenih gramatičkih struktura iz mentalnog leksikona. Teorije kažu da prizivanje riječi podrazumijeva izbor od najmanje dviju razina leksičkih informacija, semantičke reprezentacije i riječi-forme. Spontani govor karakteriziran je različitim fonetskim procesima kao što su artikulacija, varijabilnost fonetske forme riječi i različiti tipovi disfluentnosti. Kao govornici nailazimo na poteškoće tijekom govornog planiranja, dok se kao slušači moramo nositi s govornim poteškoćama drugih koje rezultiraju pogreškama. Govorne disfluentnosti obično su definirane kao pojave koje prekidaju tijek govora i ne dodaju propozicijski sadržaj nekom izrazu. Postoje različite forme disfluentnosti i mogu se razlikovati u različitim jezicima.

U ovom se radu prezentiraju rezultati niza eksperimenata usmjerenih prema govornim disfluentnostima provedenih s mađarskim govornicima/slušačima. Analiziran je korpus koji sadrži više od 5 000 uočenih pojavnica disfluentosti te snimljeni spontani govor u trajanju od 8 sati. U svrhu istraživanja problema provedeno je i nekoliko posebnih eksperimenata u koje su bili uključeni mladi, djeca i starije osobe.

Postavljena su sljedeća pitanja. (1) Na koji način različiti tipovi i pojave disfluentnosti ukazuju na poteškoće s kojima se govornici sreću tijekom govorenja? (2) Koji tipovi poveznica između fonetskih i fonoloških operacija rezultiraju pogreškama pri planiranju govora? (3) Odnose li se stanke na određene operacije u mentalnom leksikonu koje bi predvidjele fonetski rezultat? Vremenska analiza prizivanja riječi provedena je tehnikom izazivanja riječi koja je na vrhu jezika ("tip-of-the-tongue" elicitation), dok su stanke koje upućuju na poteškoće pri traženju riječi mjerene i u spontanom govoru. (4) Na koji se način odvija govornikova samokontrola? Koje strategije leže iza različitih rezultata procesa samokontrole? S jedne strane postoje disfluentnosti koje govornik ne uspijeva primijetiti, dok s druge postoje one koje govornik primjećuje, ali koje unatoč tome ostaju neispravljene. Postoje i one disfluentnosti koje su ispravljene tijekom govorenja, ali to ispravljanje govornik ne primjećuje. (5) Postoje li razlike u tipu i pojavi disfluentnosti ovisno o starosti govornika? (6) Kako govornikove disfluentnosti utječu na slušače? Govorna percepcija iznimno je brz

proces, pod uvjetom da, dok interpretira dolazeći valni oblik kao niz lingvističkih segmenata i suprasegmenata, mehanizam također ostaje stalno spreman za primanje i ispravljanje pogrešnih poruka. Kako su govornici sposobni pratiti vlastiti govor dok u isto vrijeme slušaju govor drugoga (upotrebljavajući tehniku simultanog ponavljanja tuđeg govora (shadowing)) ili dok su ometani pozadinskom bukom (Lombardijev efekt)? Koja vrsta strategije pomaže u razumijevanju govora s pogreškama? Naši rezultati omogućuju formiranje hipoteze koja objašnjava razlike u govornim disfluentnostima (kako u produkciji tako i u percepciji) ovisno o njihovu tipu, govornikovoj dobi, stvarnom kontekstu i razini planiranja na kojoj se pojavljuju.

Ključne riječi: *disfluencija, samoispravljanje (fonetika), govorna proizvodnja, fonetika, mađarski jezik*