Temporal Organization of Phonological and Phonetic Encoding: Evidence from Experiments and Spontaneous Speech

Mária Gósy
Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary
E-mail: gosy@nytud.hu

ABSTRACT

The interaction between pauses and the retrieval of the desired lexemes in the process of word production involves controversies that are worth investigating. The hypothesis of the present research was that certain pauses might 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 (i) marking the speaker's word finding trouble and (ii) preceding restarts and repetitions were measured in spontaneous speech. Results confirmed the existence of specific temporal organization underlying lexical access: a significantly different amount of time was measured depending on the subprocesses and on the mode of word retrieval between concept and articulation.

1 INTRODUCTION

Research in the past decades has yielded a lot of evidence for the hypothesis that lexical access in speech production occurs in two distinct steps: first the selection of a semantically and syntactically specified lexical representation, then the selection of its phonological content [1, 2, 3]. Models of word production may differ in various ways; however, they seem to agree on the two stages of the process concerning the activation of a syntactically specified representation prior to phonological specification. It is postulated that the lemma stage precedes the lexeme stage of processing. First a semantic then a syntactic representation is selected representing lemmas and then follows the selection of the phonological contents of the words representing lexemes. The tip-of-the-tongue phenomenon (TOT) is frequently cited to support the dual stage models of lexical access. There are experimental results, however, that do not support the successful retrieval of a word's syntactic features prior to successful retrieval of its phonological information. Evidence from TOT experiments showed that access to a word's partial phonological information can take place without selection of its syntactic features [4]. Temporal investigation of the retrieval of a word's features has provided further evidence for the actual existence of dual processing. Experimental data show that during noun-phrase production speakers retrieve the syntactic gender of a noun before its phonological properties. The phonological information was available for the subjects 40 milliseconds after syntactic properties had been retrieved [5]. This result provided direct evidence on the real time activation of syntax and phonology in noun production with the consequence of supporting the dual-stage word production models.

Accepting Levelt's discrete stage network model we assume that two lexical nodes – a lemma and a lexeme node – intervene between a word's semantic and phonological content. Whenever the speaker seeks for the appropriate word – either in a TOT condition or in spontaneous speech as a consequence of actual planning – the question arises how much time he needs for successful retrieval of the word depending on the various routes or strategies he uses during lexical access. The claim that formally related words compete during speech production is confirmed by the fact that phonologically related word forms facilitate word retrieval. Our hypothesis is that pauses can refer to specific operations in the mental lexicon, and can predict the phonetic output even in spontaneous speech. We suggest that the duration of pauses shows the activation time during lexical access. If this is so, temporal organization may highlight the interrelations of activated nodes of the processing. Two kinds of investigations were carried out in order to confirm the hypothesis. The first one concerned the acoustic-phonetic analysis of word retrieval by subjects being in a TOT condition in an experiment. Subjects in a TOT state have the 'feeling of knowing' a word despite their momentary inability to retrieve that word. When they are in this state, they are convinced about its semantic properties and can frequently retrieve partial phonological information about the word (like sounds or rhythm of the word, or the number of syllables). Since the Hungarian language has an agglutinative character its speakers often claim to know the exact morphological status of the word they are seeking for in a TOT condition. In spontaneous speech speakers come across not only the TOT phenomenon but also unsuccessful lexeme retrieval when despite of being aware of all necessary properties of the word they intend to produce they will utter another word completely mistaken as to its semantics but one that is appropriate as to its syntactic properties and very frequently its phonological features resemble to those of the intended word (e. g. intended Hungarian word: [øŋke:ɳɛʃ] 'arbitrary' uttered word: [øŋke:ntɛʃ] 'voluntary' or [konte:nɛr+bɛ] 'in container' instead of [kɔrɔnte:m+bɔ] 'in quarantine'). Pauses preceding assumed instances of unsuccessful lexical access were analyzed in continuous speech in terms of their duration. Those pauses were selected which occurred in positions

 ISBN 1-876346-48-5

where the speaker searched for the appropriate word, suffix or the second part of a compound. The temporal characteristics of the subjects' efforts to find the desired word in both conditions will be discussed.

2 METHOD AND MATERIAL

Expeirment I. 87 Hungarian-speaking adult subjects (46 females and 41 males, ages between 20 and 40) were given word definitions and asked to say aloud all responses that came to mind in the course of their attempts to retrieve the target words. All subjects were neurologically intact. The aim of the experiment was to evoke experimentally induced anomia [4]. They were asked to utter every item they retrieved after they had realized the semantics of the word that was about to come until they were able to find the lexeme itself. All participants were tested individually and their responses were tape-recorded in a sound proof chamber. The TOT condition could be reached in 1.9 cases per subject on average. 52 Hungarian words were selected that had the following characteristics: (i) they were infrequent words (49 nouns and 3 verbs) but all of them were supposed to be known for adult subjects (both native Hungarian words and loanwords), (ii) 49 non-compounds and 3 compound words, (iii) target words ranged in length from 1 to 5 syllables, and were distributed as follows: one syllable, 3.89%; two syllables, 26.97%; three syllables, 50.9%; four syllables, 14.38%; five syllables, 3.86%. E. g.: *kromoszóma* 'chromosome', *piramis* 'pyramid', *kaleidoszkóp* 'kaleidoscope', *urna* 'sepulchral urn', *donor* 'id.'. Subjects were in the TOT state altogether in 165 cases which is 3.64% of all possible cases. All different parts of the subjects' verbal communication were measured from silent intervals through laughing, groaning, or sighing as well as the duration of words and utterances up to the target word (for example "*wait, wait*" or "*oh, yes, yes it's on the tip of my tongue*" or "*I feel but I am not able to utter it*" or "*sorry, I cannot retrieve it*" or "*it will come*" or "*just I am saying*").

Experiment II. The second investigation was based on the analysis of a 7-hour continuous speech sample (consisting of a 2-hour monologue sample with the participation of 8 speakers and a 5-hour dialogue with the participation of altogether 7 speakers). Pauses were marked in the texts where the speaker was assumed to seek for the appropriate word (88 cases) or suffix (78 cases) or, repaired an articulation error (82 cases) and completed a cut-off word (84 cases). For instance "*ez nem volt egy könnyű* pause (867 ms) *feladat*" 'it was not a simple task'; *testrésznek* pause (685 ms) *testrésszel* ('for the part of body ~ with the part of body'); *találkozóhozs* silent pause (242 ms) *találkozóhoz* ('meetind ~ meeting'); *magya* pause (197 ms) *magyaráz* ('exp ~ explains'). Altogether 332 cases were selected for further acoustic-phonetic analysis. Pauses were analyzed occurring within words between two stems (compound words) and between stem and suffix or prefix (e. g. *feladat* pause /118 ms/ *nak* 'task + for'; *pipere* pause /126 ms/ *cikk* 'cosmetic + article'; *meg* pause /98 ms/ *adta* '/he/ gave /it/), altogether 148 cases. All types of disfluencies were categorized according to the speakers' repairing processes while the durations of all pauses were measured.

Silent and/or filled pauses in both investigations were measured by means of Computerized Speech Lab type 4300B with reference to sound pressure waveforms and wideband spectrograms. There was no problem with segmentation with any speaker or any type of text. The shortest duration between two words that was taken into consideration was set to 50 ms. Means and standard deviations were calculated for the durations in all cases and all categories. The statistical evaluation of the data was based on the ANOVA procedure that was carried out in SPSS for Windows 8.0 software package. In all cases confidence level was set at 99%.

3 RESULTS

3.1 The 'tip-of-the-tongue' elicitation experiment

Four categories could be defined for the 'route' subjects showed in order to reach the desired lexeme in the TOT condition: hesitation, phonetic, semantic and mixed processes. The route called 'hesitation' means that the speaker does not utter semantically identifiable things just silent or filled pauses. Fillings as they are typical of Hungarian are most frequently [ø] sounds with various durations and less frequently some sort of murmur. E. g. (! marks the moment of beginning the TOT state): ! [m] (2,770 ms) silent pause (4,250 ms) [øː] (370 ms) *orgazda* 'receiver of stolen goods'. The route called 'phonetic' means that the speaker tries to reach the lexeme uttering aloud speech sounds, sound combinations and syllables that come to his mind. E. g.: ! silent pause (920 ms) [k] silent pause (50 ms) [k] *kaloda* 'pillory' or ! [øː] (260 ms) [li] [lib] silent pause (870 ms) *libéria* 'livery' silent pause (456 ms) [lib] silent pause (50 ms) *librettó* 'libretto'. The route called 'semantic' means that the speaker utters synonyms, antonyms, phrases or semantically related short utterances that he activates hoping to help himself to get the word. E. g.: ! 4,180 *Egyiptom* 'Egypt' silent pause (1,060 ms) *fáraó* 'pharaoh' silent pause (1,365 ms) *Egyiptom, fáraó, piramis* 'Egypt, pharaoh, pyramid' (target) or ! silent pause (1640 ms) [øː] (650 ms) "it is not *fakír* 'fakir' but silent pause (2,250 ms) *mazochizmus*" 'masochism' (target). The route called 'mixed' means that the speaker activates both semantic and phonetic properties of the intended word in the TOT state. E. g.: ! silent pause (3,360 ms) *utópia* 'utopia' silent pause (5,170 ms) *uszály* 'tow-boat' silent pause (1,070 ms) *missile* (uttered in English) silent pause (14,700 ms) *uszály* 'tow-boat' silent pause (8563 ms) *torpedó* 'torpedo' (target). Responses that were phonologically related to the target word needed less time for activation than those being semantically related (cf. Table 1). These findings support the view that the interloper words or syllables are parts of the process that leads to TOT resolution.

The fastest processing was found using phonetic properties while the longest route resulted from the mixed type processing (cf. Table 1). It can be concluded that phonologically related word forms really facilitate

retrieval. The one-way ANOVA analysis indicated a significant effect of the type of processing on the time course of lexical retrieval [$F(3,125) = 7.973$, $p < 0.0001$]. Games–Howell post hoc comparisons revealed that the difference between the phonetic and semantic and the phonetic and mixed types is significant at the 0.01 level.

Route type	Durations between the lemma and lexeme levels	
	mean (ms)	std. dev. (ms)
phonetic	3467.55	3004.06
semantic	8631.51	6332.32
pause	5534.13	5517.99
mixed	9059.96	7006.94

Table 1: Amount of time speakers used in a word retrieval process in a TOT induced experiment

3.2 Data of spontaneous speech

Momentarily unsuccessful lexical access occurs in various forms in spontaneous speech depending on the actual place in the process between concept and articulation. The three types of difficulties in lexical access concern (i) semantic selection (between concept and lemma), (ii) morpho-phonological difficulty (between lemma and lexeme), and (iii) phonetic problem (articulation). The first one is characterized by various kinds of pauses before the uttered words while the others appear in terms of various forms of restarts with pauses between the uttered words. The latter are: completion of cut-off words, repetitions involving morphological changes and repetitions resulting in the corrected phonetic form of the word. The same kinds of trouble may apply also within a word considering the morphological richness of Hungarian, i.e. semantic selection problem within a compound word, suffix/prefix selection problem within the word or, momentary articulation trouble.

(i) Semantic selection. When the speaker speaking spontaneously is in trouble trying to find the appropriate word, this difficulty is expressed either by a pause before uttering the lexeme or by a restart of the incomplete word. The word finding trouble may arise for three reasons: either the speaker does not know exactly what to say next or, he is not able to decide which possible word to select (competing words have been activated) or, he is in a TOT-like condition. There are not always unambiguous clues for distinguishing the reasons when analyzing spontaneous speech. Restarts here are those repetitions where the intended word was cut off somewhere in the middle of the sound sequence and after a certain amount of time the speaker restarted and completed the word. This shows the uncertainty of the speaker concerning his lemma selection. The relatively frequent occurrence of restarts in Hungarian with the aim of completing a word can be explained by the statistical facts that the most frequent words in spontaneous Hungarian speech consist of four syllables where one or two syllables belong to the category of suffixes (or prefixes).

The analysis of the word seeking process revealed that nouns caused more trouble for the speakers (in 72.21% of all cases) than other categories (adjectives were found to be sought only in 23.61% of all cases while verbs in 4.16%, respectively). The pauses preceding the intended words are obviously a lot shorter than those obtained in the TOT elicitation experiment (the real TOT problem is solved by means of various ways in continuous speech). The mean value of the word seeking process was 677. 72 ms (std. dev.: 530.93 ms). What is particularly interesting here is the pattern of pauses preceding the intended word in continuous speech. The following subtypes of pauses were found (i) silent pause (17.14%), (ii) filled pause (12.85%), (iii) silent+filled (17.14%), (iv) filled+silent (12.85%) and (v) silent+filled+silent (40%). Statistical analysis confirmed significant differences between filled pauses vs. all the others and between the silent+filled+silent pause pattern vs. all the others ($p<0.01$). It can be assumed that the longer the pause the greater the speaker's word finding difficulty which is predicted by the silent+filled+silent pause pattern.

(ii) Morpho-phonological selection. The momentary failure of lexical access can also be caused by a false selection of suffix. Repetitions involving a morphological change show that the speaker selected the wrong suffix for the semantically correct word. The syntactic plan has already been completed, therefore the morphologically erroneous word needs correction. The speaker has to "go back" to an earlier phase of the speaking process in order to select another form of the suffix morpheme to complete the lexeme, and this processing takes more time.

(iii) Phonetic execution. Pauses often predict repairs at the phonetic level of lexical access. Slips of the tongue are corrected when the speaker realizes that the phonetic structure produced does not match the required one. These corrections are repetitions with the correct phonological structure of the lexeme. It is assumed that less time is necessary for correcting suffixes and slips of the tongue or for completing incomplete words than for selecting the appropriate semantics of the intended concept. If that assumption is correct the pauses in cases of repetitions should be shorter than those involving a word finding problem.

Empirical data and statistical analysis supported the hypothesis (cf. Fig. 1.). Mean duration of pauses in word seeking cases is 636.0 ms (std. dev.: 304.92), in restarts with completing the words is 172.48 ms (std. dev.: 366.32 ms), in slips 168.54 ms (std. dev.: 141.38 ms) while in repetitions with morphological changes is 500.56 (std. dev.: 366.32 ms). Statistical analysis yielded the result that the differences were significant [$F(3,332) = 83.439$, $p < 0.0001$]. According to Games–Howell post hoc comparisons differences between the pauses of restarts with morphological changes as opposed to the restarts with completing the word and with correcting slips were significant ($p < 0.0001$). The difference between the word seeking pauses and disfluencies caused by difficulties of morpho-phonological selection was also significant but less robust ($p = 0.015$). However, there was no such difference between the pause durations of restarts with completing the words and of correcting slips even at a 95% confidence level.

ISBN 1-876346-48-5 © 2003 UAB

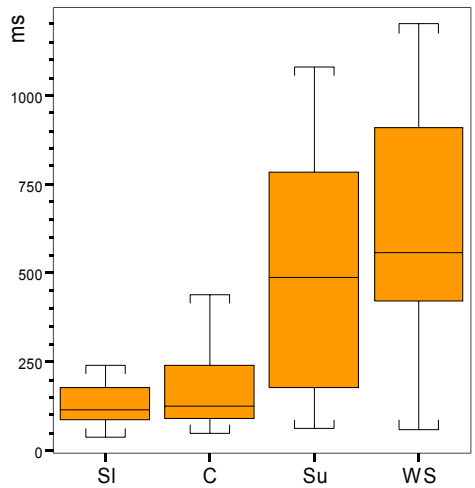

Fig. 1: Pause differences prior to word seeking (WS), completing words (C), correcting suffixes (Su) and slips (Sl)

Pauses within words. A special case occurs when a pause exists within a word (for example: *zené* silent pause /148 ms/ *ben* 'in music'). Because of the rich morphology of Hungarian, such pauses can occur at various points within the word. They can appear between the prefix and the word stem like *leg* 105 *népszerűbb* 'the most popular', *meg* 181 *alkotására* 'for its establishment' or, can precede the final suffix of the word like *zené* 72 *ről* 'about music', *kutatás* 178 *nak* 'for research', *fogalmaink* 67 *tól* 'from our concepts'. Pauses occur within the word stem without violating the syllable boundaries like *improvi* 148 *zációs* 'improvisational', *har* 82 *madik* 'third'. Finally, pauses often separate the two halves of compounds like *nép* 120 *zenében* 'in folk music', *interjú* 98 *módszer* 'method of interview' or *pipere* 160 *cikk* 'cosmetic article'. Syllabification is so strongly encoded that the speaker corrects the syllable structure if the interrupting pause appeared at the wrong speech sound, i.e. not at a syllable boundary. Let us see the following example: *szós* pause 579 *statisztikáját* 'his word statistics (acc.)'. The speaker uttered the first syllable *szó* together with the first consonant of the next syllable [ʃ] of the compound word. Realizing that he stopped within a syllable before continuing the articulation after the pause, the speaker repeats the [ʃ] consonant in order to recover the correct syllable. The recovery of the appropriate syllable structure appears in all cases where a pause interrupts the articulation of a word which can be explained by assuming that the phonemes of the words are retrieved from the lexicon with their syllabic position specified.

Statistically significant differences were found between the pauses occurring within compounds and non-compound words ($p = 0.006$) that refer to a TOT-like state in the case of compound words (mean value of pauses within non-compounds: 107.33 ms, std. dev.: 46.71 ms while within compounds: 290.83 ms, std. dev.: 165.94 ms). The duration of pauses differs also significantly depending on their place concerning the boundary between the word stem and the prefix or suffix (mean value of pause durations occurring between word stem and suffix: 170.66 ms, std. dev.: 183.50 ms; mean value of pause durations occurring between prefix and word stem: 343.16 ms, std. dev.: 233.64 ms; $p = 0.003$). We might assume that pauses after prefixes refer to word finding difficulties, and the data seem to confirm this assumption since there is no significant difference concerning the duration of pauses occurring after a prefix or within a compound word; however, significant difference was found between the pauses occurring after a prefix and within a non-compound word.

4 DISCUSSION

The temporal organization of lexical access is not explicitly modeled in speech-error based models. The present results show that the farther from articulation the problem arises in the course of lexical access the longer the pause before the repairs. The duration of pauses might be characteristic of the repairing process: the speaker needs the longest time when he has problems with semantic selection, he needs shorter time for correcting morphological troubles while the shortest pause is required to correct phonetic errors. The data of the tip-of-the-tongue experiment supported the above findings. The seeking time of the target word was much longer in cases the speaker activated semantically related words while it was significantly shorter in cases of activation of phonologically related words. The speaker has an internal loop of max. 100 ms to send an interrupt signal to the articulator in case of trouble [3]. This applies to the analyzed restarts that can easily be repaired such as slips and completing words. However, those troubles that need higher organized corrections like change of suffixes or selecting the appropriate words, cannot be done fast. It is assumed and argued for that in planning a sentence the retrieval of the morpho-phonological codes for each selected word takes place. Our findings have confirmed the hypothesis that pauses really predict the type of processes in lexical access.

Acknowledgments

This research has been supported by OTKA T 037287.

REFERENCES

[1] Fromkin, V. A., "The non-anomalous of anomalous utterances". *Language* **47**, pp. 27-52. 1971.

[2] Garrett, M. F., "Levels of processing in sentence production". In: Butterworth, B. (ed.). Language Production. Academic Press, London, pp. 35-71, 1980.

[3] Levelt, W. J.M., "Speaking: From Intention of Articulation". MIT Press, Cambridge, MA. 1989.

[4] Caramazza, A., Miozzo, M., "The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip-of-the-tongue' phenomenon". *Cognition* **64**, pp. 309-343, 1997.

[5] Turennout van, M., Hagoort, P., Brown, C.M., "Brain activitiy during speaking: from syntax to phonology in 40 milliseconds". *Science* **280**, pp. 572-574, 1980.