REPAIRING PHONOLOGICAL SPEECH ERRORS IN NOVEL PHRASES AND PHRASAL LEXICAL ITEMS

Sieb Nooteboom

Utrecht Institute of Linguistics OTS, Utrecht University, the Netherlands
s.g.nooteboom@uu.nl

ABSTRACT

Mental preparation for production is supposed to be more automatic for phrasal lexical items (PLIs) than for novel phrases (NPs). Automatic processes are thought to be less error prone and less closely monitored than novel processes. From this it is predicted that speech errors are less often detected and repaired by speakers when made in PLIs than when made in NPs. This prediction has been tested and confirmed in a corpus of phonological speech errors in spontaneous Dutch.

Keywords: production, phonology, speech errors, self-monitoring, phrasal lexical items

1. INTRODUCTION

This paper compares the repair rates of phonological speech errors made in phrasal lexical items (PLIs) and in novel phrases (NPs). PLIs are for example proverbs, sayings, idioms, collocations, clichés, all assumed to be included in what native speakers know about their language, whereas NPs are prepared by speakers in the act of speaking. The difference between the two phrase types is theoretically accounted for by Levelt and Meyer [4], focusing on idioms. They assume that each PLI is represented in the lexicon by a “superlemma”, being an internalized syntactic representation of the phrase concerned, with pointers to the lemmas of words constituting the phrase. This theory predicts on the one hand that the generation of a phrase, other things being equal, takes somewhat longer for a PLI than for a NP because of the extra step introduced by the superlemma. On the other hand, the theory predicts that context effects, activating words that are part of the phrase to be produced, have a much greater accelerating effect on the mental generation of the phrase in PLIs than in NPs. This is so because in PLIs, but not in NPs, such context effects generalize to all constituent words of the phrase by activating the superlemma. This was confirmed in experiments reported by Sprenger, Levelt and Kempen [8]. Their findings suggest that in the absence of priming one of the constituent words, preparation of idioms takes somewhat more, but in the presence of such priming significantly less time than preparation of NPs. One may assume that in preparing normal spontaneous speech, in virtual all cases PLIs somehow fit the preceding context, and that therefore, other things being equal, their mental preparation is faster and more automatic than the mental preparation of NPs.

The underlying idea of this paper, then, is that in normal spontaneous speech, mental preparation is more automatic for PLIs than for NPs. Automatic mental processes are thought to be less error prone and less closely monitored than novel mental processes [9]. From this it is predicted that (a) less speech errors are made in PLIs than in NPs, and (b) speech errors are less often detected and repaired in PLIs than in NPs. It is currently unknown whether or not less speech errors are made in PLIs than in NPs. Here we focus on the second prediction, comparing repair rates for phonological speech errors in spontaneous Dutch in PLIs and NPs.

2. COMPARING ERROR REPAIR RATES IN PLIS AND NPS

2.1. The corpus

The corpus of Dutch speech errors used here contains 2,455 errors in Dutch spontaneous speech, collected some twenty-five to thirty years ago in the Phonetics Department of Utrecht University [7]. For current purposes it is important to note that the collectors, all staff members of the Phonetics Department, were instructed to write down each error with its repair, if it was repaired. The collecting of speech errors is potentially error prone (cf. Cutler 1982). In particular, it seems very likely that repaired speech errors are more often noted by collectors than unrepaired speech errors, because repairs are highly conspicuous interruptions of continuous speech. However, there is little reason to assume that the possible overestimation of the proportion of repaired errors is different for PLIs and NPs. If it is not, the main question of this paper remains answerable.

2.2. Selecting a relevant subset of errors

In speech errors one can distinguish between the *source* of the error, i.e. the position where a misplaced element (phoneme, morpheme, word) stems from, and the *target* of the error, viz. the position where a misplaced element ends up. The corpus includes 1,370 paradigmatic and 1,085 syntagmatic errors. In the former the source of the error is not part of the verbal message, in the latter it is. Because for the current purpose we need to focus on errors spanning a sequence of more than a single word, all paradigmatic errors are excluded. To make the set of errors more homogeneous, all 163 lexical errors, where the misplaced element is a morpheme or a word, were excluded. Twenty-one phonological errors in a different language than Dutch, mostly English, were also excluded, leaving 901 phonological errors in spontaneous Dutch. Of these, 214 errors had source and target in the same word. These were excluded for the same reason paradigmatic errors were excluded, leaving 687 phonological errors. Of these, 28 errors concerned combinations of given and family name, or combinations like John and Mary. These were excluded because the status as PLI or NP of such name combinations seems hard to assess and may vary enormously from individual to individual. This left 659 phonological speech errors serving as the basis for answering the main question of this paper.

2.3. Assessing novelty and lexicality

An English example of one of the errors in the set of 659 errors could be *left after one inding with a tender elbow* for *left after one inning with a tender elbow* [2]. But the same error could equally well have been noted down as *inding with a tender* for *inning with a tender*, leaving out all material preceding and/or following the word sequence running from source to target or target to source. In order to make all speech errors comparable in this respect, in all errors all material preceding and/or following the sequence containing source and target of the error was removed. After that all speech errors were changed back into the error-free intended word sequences, in the example *inning with a tender*. Many of these were, of course, not themselves proper phrases. However, preliminary inspection of all word sequences showed that there were at least four classes that could be distinguished and should very likely be kept separate in further analysis. These were:

1 Typical (parts of) PLIs
2 Perhaps (parts of) PLIs
3 Typical (parts of) NPs
4 Combinations of function words (FWs)

Class 4 was set apart because it appeared subjectively always impossible to estimate whether these combinations of function words were or were not parts of PLIs.

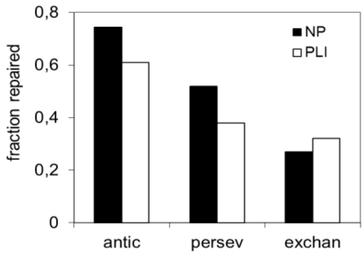
All 659 word sequences were presented on paper to three linguistically non-naïve judges, not including the present author. The judges were familiar with the four classes of word sequences. Each judge was asked to assign one of the four codes mentioned, to each word sequence. Results were further re-coded as follows: In 24 cases at least one of the judges had assigned case 4 (combination of function words). Where the judges differed this was because of uncertainty whether particular auxiliary or modal verbs are or are not function words. To be on the safe side, all 24 word sequences were given code 4 for further analysis. Of the remaining 635 cases there were 386 cases where all three judges agreed, 215 cases where two of the three judges agreed, and 34 cases where all three judges had a different judgment. These 34 cases were all assigned code 2, reflecting uncertainty as to lexicality. In all remaining 601 cases the majority of the judges was followed in the re-coding, giving 92 cases with code 1 (PLI), 75 cases with code 2 (PLI?), 468 with code 3 (NP), and, as said before, there were 24 cases with code 4 (combination of function words).

To check the validity of this classification, so-called Search Engine Count Estimates (SECEs) [1] were assessed for all word sequences, using Yahoo constrained to Dutch and to the precise symbol sequences by using double quotes. Yahoo SECEs were transformed by taking the $_{10}$log. In all those cases where the actual Yahoo count was 0, this 0 was set to 1 so that the $_{10}$log was 0. These data were analyzed with a simple Univariate one-way Analysis of Variance with $_{10}$log Yahoo SECE as dependent variable and category (1, 2, 3 ,4) as fixed factor, giving a significant effect of category on log frequency (df=3; F=97; $p<0.001$). A post hoc analysis using Tukey's showed that *PLI* and *PLI?* were not significantly different, whereas all other contrasts were. The average log SECE was significantly higher for *FWs* than for *PLIs*. Given that there is no reason to assign *FWs* the status of either *PLI* or *NP*, this demonstrates that frequency of usage in itself is not a good criterion for deciding on lexicality of a phrase. The 24 *FWs* were removed from the set of word sequences in all further analysis, leaving a set of 635 word sequences. This new data set was again submitted to a one-way Univariate Analysis of Variance with 10log Yahoo SECE as dependent variable and category (1, 2, 3) as fixed factor, again giving a significant effect of category on log frequency (df=2; F=110; p<0.001). A post hoc analysis using Tukey's showed no significant difference between *PLI* and *PLI?* and significant differences between both *PLI* and *PLI?* on the one hand and *NP* on the other. It was decided to collapse *PLI* and *PLI?* into *PLI* for further analysis. Removing PLI? would have left us with too few cases of PLI in further analysis. Of each of the 635 word sequences the following information, among other things, is now available: a) estimated lexicality (*PLI* vs *NP*), b) speech error class (anticipation vs perseveration vs exchange), c) repair status (repaired vs unrepaired). This was the input for further analysis.[1]

2.4. Lexicality and probability of repair

An initial breakdown of the repair data is presented in Figure 1. Here the actual fractions repaired as found in the corpus are presented separately for *PLI* and *NP* and for anticipations (*antic*), perseverations (*persev*) and exchanges *(exchan*).

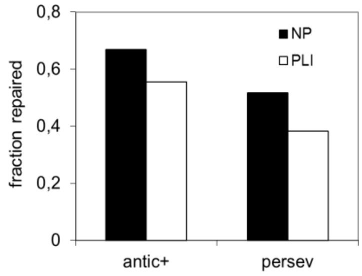
Figure 1: Fractions repaired as found in the corpus for *NPs* and *PLIs*, separately for phonological anticipations (*antic*; N=391), perseverations (*persev*; N=164) and exchanges (*exchan*; N= 80).



These data are at first sight somewhat strange. An exchange, like *Yew Nork* for *New York* [2] consists of an anticipation plus a perseveration, and thus has two chances to be detected instead of one. Therefore one expects the fraction repaired to be considerably higher, not lower, in exchanges than in the other speech errors. The very low repair rate for exchanges is to be explained before further analysis. The following explanation is suggested by [6]. First, when a speaker makes an exchange between two elements in his inner speech, resulting for example in an error like *Yew Nork*, he/she can detect and repair the error in inner speech before anything is spoken. If so, the outside world will not observe the error or the repair. Second, the speaker can detect the error after the first word and before the second word has been spoken. The first part of the error and the repair will then be overt, as in *Yew..uhh New York*. The crucial point here is that all such cases are classified as repaired anticipations, not as repaired exchanges, in a corpus of speech errors. Third, the speaker can detect the error only after the second word has been spoken. In that case the repair will again be overt, as in *Yew Nork uhh New York*. These latter cases will be classified in a corpus of speech errors as repaired exchanges. One can now explain the low repair rate of exchanges in Figure 1 by assuming that many exchanges in inner speech are detected after the first and before the second word has been spoken, and thus are classified as repaired anticipations. This will give a severe overestimation of the number of repaired anticipations and a severe underestimation of the number of repaired exchanges.

It follows that, instead of the counts made in the corpus, one would like to know the number of anticipations and exchanges detected in inner speech.

Figure 2: Fractions repaired of speech errors made in anticipations plus exchanges (*antic+*; N=333) and perseverations (*persev*; N=134), separately for *NPs* and *PLIs*. Both main effects are significant.



Unfortunately, there is no way to know whether a case like *Yew uhh New York* in a corpus derives from an anticipation or an exchange in inner speech. For this reason it is reasonable to collapse the numbers of anticipations and exchanges in the statistical analysis. This leads to the data set shown in Figure 2. Removing the exchanges, instead of collapsing them with the anticipations, would do little good, because it is almost certain that many, if not most, anticipations are half-way repaired exchanges. The data in Figure 2 were analyzed with a logistic regression using effect coding [3], with as dependent binomial variable *fraction repaired* and as fixed factors „*NP* vs *PLI*" and „*antic+* (anticipations plus exchanges) vs *persev* (perseverations)". The grand mean was used as intercept. The analysis results of the best fitting model are presented in Table 1. This best fitting model shows no interaction between the two fixed factors, and significant effects of both „*ant+* vs *pers*" (p<.003) and „*NP* vs *PLI*" (p<.02).

Table 1: Estimated parameters for the best fitting binomial logistic regression model of *fraction repaired* using effect coding. The grand mean was used as intercept. The table presents regression coefficients, standard errors, t values and p values.

effects	*coef.*	*s.e.*	*t*	*p*
intercept:	0.59	0.024	24.01	<.001
NP/PLI:	+/-0.13	0.055	+/-2.33	<.020
ant+/pers:	+/-0.15	0.049	+/-3.05	<.003
interact:	0.02	0.110	0.18	<.860

As expected, the fraction repaired is considerably and significantly higher for anticipations plus exchanges than for perseverations, confirming that most repaired anticipations stem from halfway repaired exchanges. That the repair rate is significantly lower for PLIs than for NPs confirms the main hypothesis tested in this paper.

3. DISCUSSION

Automatic processes are thought to be less error prone, and also less closely monitored for errors than novel processes [9]. Therefore one would expect that (1) less speech errors are made in PLIs than in NPs and (2) speech errors are less often detected and repaired by the speaker in PLIs than in NPs. The relative frequency of speech errors in PLIs and NPs has, as far as I know, not been investigated. This is yet to be done. The second prediction, that speech errors are less frequently detected and repaired in PLIs than in NPs, has been tested in this paper, using a corpus of phonological speech errors made in spontaneous Dutch. The results clearly show that, other things being equal, phonological speech errors are less often repaired in PLIs than in NPs. This confirms an earlier conclusion in [5]. The results also are in line with a theoretical account of the mental preparation of PLIs as compared to NPs in [4], where it is assumed that each PLI is lexically represented by a superlemma from which constituting lemmas are activated in one go. This makes, at least within proper context, mental preparation more automatic for PLIs than for NPs.

4. ACKNOWLEDGEMENTS

The author is grateful to Hugo Quené for his assistance in running the logistic regression analysis, and for his helpful comments on the investigation.

5. REFERENCES

[1] Dietmar, J. 2008. Objectivity, Reliability, and Validity of Search Engine Count Estimates. *International Journal of Internet Science* 3(1), 7-33. URL: http://www.ijis.net/ijis3_1/ijis3_1_janetzko.pdf
[2] Fromkin, V.A. 1973. Appendix. In Fromkin, V.A. (ed.), *Speech Errors as Linguistic Evidence*. The Hague: Mouton, 243-269.
[3] Johnson, K. 2008. *Quantitative Methods in Linguistics*. Malden MA: Blackwell.
[4] Levelt, W.J.M., Meyer, A. 2000. Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology* 12(4), 433-452.
[5] Nooteboom, S.G. 1999. Sloppiness in uttering stock phrases. *Proceedings of the XIVth International Congress of Phonetics Science* San Francisco, 683-686.
[6] Nooteboom, S.G. 2005. Listening to one-self: Monitoring speech production. In Hartsuiker, R.J., Bastiaanse, R., Postma, A., Wijnen, F. (eds.), *Phonological Encoding and Monitoring in Normal and Pathological Speech*. Hove and New York: Psychology Press, Taylor and Francis Group, 167-186.
[7] Schelvis, M. 1985. The collection, categorization, storage and retrieval of spontaneous speech error material at the Institute of Phonetics. *PRIPU* 10 Utrecht, 3-14.
[8] Sprenger, S.A., Levelt, W.J.M., Kempen, G. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54, 161-184.
[9] Wheatley, T., Wegner, D.H. 2001. Automaticity of action, psychology of. *International Encyclopedia of the Social and Behavioral Sciences,* 991-993.

[1] The list of word sequences with their relevant properties is available in EXCEL format at the following URL: http://www.let.uu.nl/~Sieb.Nooteboom/personal/PLIs&NovelPhrasesWordSequences.xls