

**Self-monitoring for speech errors in novel phrases and phrasal lexical items**

Sieb Nooteboom

Utrecht institute of Linguistics OTS

The Netherlands

In: Yearbook of Phraseology 2011, pp. 1-16.

## Self-monitoring in novel phrases and phrasal lexical items

### 1. Introduction

This paper concerns self-monitoring for phonological speech errors during the production of novel phrases and phrasal lexical items (PLIs). Novel phrases are assumed to be prepared by the speaker in the act of speaking, not part of what is stored in the speaker's mental lexicon. PLIs are assumed to be known as components of what a native speaker knows as part of his/her knowledge of the native language. The difference between novel phrases and PLIs is significant for language use. As is well known to foreign-language teachers, native-like fluency in a language is dependent on the number of PLIs known by the language user (e.g. Ketko 2000). It has been observed that English for a large part is *formulaic* (Bolinger 1975; Pawley and Syder 1983; Van Lancker-Sidtis 2004; Kuiper 2009), meaning that many utterances contain PLIs. Pawley and Syder estimate that there are hundreds of thousands of such PLIs in the lexicon of English. Jackendoff (1995) estimates 10,000 - 15,000 different English idiomatic phrases as having been used as stimulus items during 10 years in the TV-show "Wheel of Fortune" alone (idiomatic phrases or idioms are semantically non-compositional PLIs, Fraser 1970). There is no reason to think that the dependence on PLIs in linguistic performance will be different in other comparable languages.

Unfortunately for those who like clearly delimited classes, the distinction between novel phrases and PLIs is far from clear-cut. We may think of PLIs as *phrasal chunks* that are stored in the mental lexicon, whereas novel phrases are not so stored. But it is not always easy to know what precisely is stored in the mental lexicon and in what form. Proverbs provide more or less canonical examples of PLIs, being complete sentences that seem to be mentally stored as units, just as words are. But obviously proverbs retain their syntactic structure to the extent that speakers can make variations on them, as in *a new broom sweeps clean, new brooms*

*sweep clean*, *a new broom sweeps cleaner than you will like* etc. Most idioms have explicitly open places to be filled by the speaker in the act of speaking: *He / she / John / my mother-in-law / etc. is pushing up daisies*. Also in restricted collocations like *dry wine*, *white coffee*, *white noise* the meaning of the adjective is restricted by the noun, and in this sense lexically determined. But this non-compositionality of meaning is not always complete: in idioms the literal meanings of the words may contribute to the idiomatic meaning of the phrase (Nunberg, Sag and Wasow 1994). Also non-compositionality is not a defining property of PLIs. There are PLIs such as *take note of*, *offer hospitality* that seem to have compositional readings, but nonetheless definitely are PLIs in the sense that native speakers know that this is how one says such things.

How are PLIs represented in the mental lexicon and how are they activated during speech production? Swinney and Cutler (1979) proposed that idiomatic expressions, the most studied variety of PLIs, are stored in the mental lexicon as long ambiguous words. This implies that idioms perhaps might have word-like morphological properties but not sentence-like syntactic and semantic properties. In contrast, later theories assume that idiomatic expressions and presumably other PLIs are not only themselves lexical units, but also consist of constituent lexical units, and more importantly have sentence-like syntactic and semantic properties (Cacciari and Tabossi 1988; Cutting and Bock 1997; Levelt and Meyer 2000). According to those theories one would predict that idiomatic expressions and other PLIs can not only be themselves units involved in lexical speech errors, but also that their constituent words can be involved in speech errors as they can in novel phrases. Levelt and Meyer (2000) coined the term *superlemmas* for internalized syntactic representations of PLIs, and supposed that a superlemma, once activated from its associated lexical concept, activates the constituent word lemmas of the PLI. Kuiper et al. (2007) analyzed collections of speech errors involving *Multi-word Lexical Items* or MLIs, MLIs comprising both PLIs and compounds. Importantly they found that all the speech error types known to occur in novel

phrases also occur in MLIs. They found a number of error types characteristic of MLIs such as blends of MLIs and several types of errors where MLIs interact with semantically related single words. These error types and their distribution are predicted by Levelt and Meyer's superlemma theory, thus lending further support for this theory.

Assuming then that this theory is valid, one learns from it that the preparation in production of PLIs is, apart from the additional step required by the activation of the superlemma, quite similar to the preparation in production of novel phrases, with one additional important difference. Syntactic structure and word lemmas of a PLI are activated by the superlemma. This extra step in the generation process presumably takes time, and therefore one would predict that, other things being equal, production of a PLI takes more time than production of a novel expression. This is indeed suggested by research reported by Sprenger, Levelt and Kempen. (2006). However, their results also show that, although idiomatic expression may take somewhat longer time to be prepared for production than novel expressions when expressions are not primed by context, idiomatic expressions are prepared for production significantly faster than novel expressions when one of the words in the expression is primed by identity priming. This is accounted for by assuming that priming of idiomatic expressions impinges not only on the word concerned but on the expression as a whole, thus boosting all words in the expression. This causes that less mental computation is needed in preparing PLIs than in preparing novel phrases. In everyday speech most often PLIs will be related to preceding context or situation. We may assume then that in normal speech communication preparing a PLI is more routine, more automatic, than preparing a novel phrase. Automatic mental processes are more efficient, require less conscious guidance and monitoring than novel processes, and therefore use less attention capacity (Wheatley and Wegner 2001). An important reason that automatic processes require less monitoring than novel processes is that automatic mental processes are less error prone than novel ones. From these considerations one would

expect that the probability of a speech error occurring in a phrase would be less for PLIs than for novel phrases. Whether this is the case is as yet unknown. There does not seem to be an easy way to test this prediction on collections of speech errors in spontaneous speech, because in most collections of speech errors there is no easy way to quantify this probability. There is however, another related prediction one can derive. If PLIs are more automatic than novel phrases, and automatic processes require less monitoring because they are less error prone, then it is to be expected that speech errors made in PLIs will be less often detected and repaired by self-monitoring than speech errors in novel phrases. Below this prediction is tested against a corpus of speech errors in spontaneous spoken Dutch.

## **2. Are speech errors less often repaired in PLIs than in novel phrases?**

### **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 (Schelvis 1985). 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. Note that the collecting of speech errors is potentially error prone (cf. Cutler 1982). Some errors may more easily escape detection by the observers than others. More specifically, it seems likely that unrepaired speech errors are more often missed than repaired speech errors, because the repairs are conspicuous, easily observed, interruptions of normal fluent speech. Thus the observers' bias would probably cause an overestimation of the relative number of repaired speech errors, and an underestimation of the relative number of unrepaired speech errors. It is also possible that there is an observer bias in the sense that speech errors in novel phrases are more easily observed than speech errors in PLIs. Thus potentially both unrepaired speech errors and speech errors in PLIs would be underrepresented in a corpus of speech errors. However, the relevant question here is whether or not these

two biases are mutually dependent. I see no reason why the potential effect on error detectability of lexicalizedness on the one hand and of being unrepaired on the other would be mutually dependent. If they are independent, our specific hypothesis remains testable despite potential observers' biases underestimating the proportion of unrepaired errors and of errors in PLIs.

## 2.2. Paradigmatic and syntagmatic speech errors

Speech errors can be classified as paradigmatic and syntagmatic speech errors (Rossi and Defare 1998). An example of a paradigmatic error is when someone says *a verbal outfit* instead of *a verbal output*, where the substitution of one word by another cannot be traced to another element in the speaker's message. Examples of syntagmatic errors are exchanges like *teep a cape* instead of *keep a tape*, where two elements in the same message are interchanged, anticipations like *alsho share* instead of *also share*, where an element comes earlier than it should, often replacing another element, and perseverations like *John gave the boy* being spoken as *John gave the goy*, where an element is mistakenly repeated (all examples taken from Fromkin 1973). In syntagmatic speech errors one can distinguish between the *source* of the speech error, i.e. the position where a particular element should have been, and the *target*, i.e. the position where a misplaced element ends up. Here I will concentrate on syntagmatic errors, because paradigmatic speech errors generally involve only a single word. There is thus no way of knowing how much of the context should be taken into account when assessing whether this error occurred in a PLI. In syntagmatic errors at least one can examine the sequence of words including both source and target. Of the 2.455 errors in the corpus, there were 1.085 syntagmatic errors.

Of the 1.085 syntagmatic errors there were 163 lexical errors, all others were phonological speech errors. To keep the data set as homogeneous as possible, the lexical errors were removed. The remaining set of 922 phonological errors contained a number of errors in other languages than Dutch, mostly English. These were also removed, leaving 901 errors.

### 2.3. Length: the distance between source and target

Of these 901 remaining phonological errors 214 errors had the phonological source and target within the same word. Very often in the corpus only this single word containing source and target of the speech error, and where applicable the repair of the speech error, were noted down, making it impossible to find out whether or not this word was part of a PLI in the original context. For this reason all these cases were removed from the data set. This left 687 speech errors.

In many, if not in most cases, the complete expression, novel or lexicalized, within which a speech error occurs, is longer than the sequence of words including source and target of the error. It would be reasonable to assess the novelty or lexical nature of the complete expressions. Unfortunately, in many cases in the corpus of speech errors the complete expression is unknown, simply because the observer left out everything before and/or after the sequence of words including source and target. In order to follow the same procedure for all expressions, in all remaining cases everything before and after the sequence of words including source and target was removed. The error-containing sequences of words resulting from this procedure show considerable variation in length, from two to nine words.

### 2.4. Novel and lexicalized: first impression

In all remaining word sequences containing a speech error the speech error, and its repair if it was repaired, were removed by changing the phrase back to its intended form. Most of the resulting word sequences do not form complete sentences and very many do not even have a finite verb form. But a first inspection showed that there are at least four different classes of such word sequences that probably should not be collapsed in further analysis.

One class is formed by those word sequences that are in themselves almost certainly PLIs, mostly collocations, comparable to English *knife and fork*, *black and blue*, *head of lettuce*, *world wide*, *one day a week* or *Barack Obama*.

A second class is formed by word sequences that might well be PLIs but probably not for all users of the language. Examples in English might be *decimal*

*value* or *gross national product*. A particular class of uncertain cases is formed by combinations of given and family name, or combinations like *John and Mary*. There is no way of knowing whether such combinations were or were not so familiar to the speaker at the time in order for them to be considered PLIs. For this reason all 28 such name combinations were removed from the data set, leaving 659 word sequences for further analysis.

A third class is formed by those word sequences for which there is no reason to believe that they are PLIs or parts of PLIs. Examples in English would be *may lengthen a vowel*, *rapidly empty John*, *corner at Jacobs*, or *brake with my left foot*.

Finally, a somewhat special fourth class consists of combinations of function words, English examples being *in the*, *you at*, *they are on*, *from it*, *in on that*. It seems unlikely that these sequences constitute themselves PLIs. They may or may not have been part of longer PLIs, but they equally likely may have been part of novel phrases. In this respect they form a source of uncertainty in the data.

## 2.5 Assessing novelty and lexicality of word sequences

The data set of 659 word sequences, from which the speech errors and their repairs were removed by changing these word sequences back into the correct intended word sequences, was presented to three linguistically non-naive judges, not being the current author. These linguists were native speakers of Dutch and familiar with the notions of novel phrases and PLIs as used here. Each judge was asked to assign one of four possible codes to each word sequence, defined as follows:

- (1) This word sequence is itself a PLI, or very likely part of a PLI.
- (2) This word sequence might well be a PLI or part of a PLI, but I am not certain of it.
- (3) This word sequence is not a PLI or part of a PLI.
- (4) This word sequence is a combination of function words that very likely does not in itself form a PLI.

The resulting judgments were further reduced in the following way. In 24 cases at least one of the three judges had assigned code 4 (combination of function words) to



the word sequence, but the others had not. Where the judges differed this was mostly because of a difference of opinion whether certain auxiliary or modal verbs and certain adverbs are or are not function words. To be on the safe side, all these 24 cases were given the code 4 in order to keep them separate in further analysis. In all 635 remaining cases agreement among judges was as follows: 386 cases where all three judges agreed, 215 cases where 2 of the 3 judges agreed, and 34 cases in which all three judges had a different judgment. In these 34 cases assignments of necessity were 1, 2, and 3, reflecting the degree of (un)certainly about the lexicalizedness of the word sequence. Therefore these were assigned the code 2. In all remaining cases the majority of the three judges was followed, giving 92 cases with code 1 (PLI), 75 cases with code 2 (perhaps a PLI), and 468 cases with code 3 (novel phrase).

In order to find support for this intersubjective approach, and on the assumption that PLIs have a higher than chance frequency in text corpora, the Yahoo frequency of all 659 word sequences was assessed, with Yahoo counting limited to Dutch and to the exact word sequence, using quotation marks. It should be noted that frequency of usage of multi-word sequences cannot reliably be assessed from existing linguistic corpora such as the Corpus of Spoken Dutch (Oostdijk and Broeder 2004), because even a corpus of 10,000,000 words is simply not big enough for determining reliable frequencies of multiword sequences. Admittedly, frequency counts by web browsers provide, at best, a rough measure, first because they are based solely on documents accessible to these browsers which have not been selected with a view to creating a balanced corpus, second because in the counting many documents may be accessed multiple times, and third because the web browsers, in estimating the reported frequencies, extrapolate from the actual counts employing rules unknown to us (for pros and cons of using web browsers in assessing frequency estimates see Janetzko 2008). Nevertheless the Yahoo frequency often appears to provide intuitively plausible outcomes, and may thus be used as circumstantial evidence. A similar exercise with Google frequencies

basically gave the same results, and will not be reported here. The Yahoo frequencies found 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.

The data set for further analysis thus consists of 659 word sequences, each word sequence being the intended form of a word sequence in which a phonological speech error had occurred. Of each word sequence it is known whether it is (part of) a PLI, perhaps (part of) a PLI, or (part of) a novel phrase, or a combination of function words. Also the  $^{10}\log$  Yahoo frequency is known. Further it is known whether the original phonological speech error made in that sequence was an anticipation, a perseveration, or an exchange, and what the number of words in the word sequence is. These data are the basis of all further analysis, seeking an answer to the question if speech errors in PLIs are less often repaired than speech errors in novel phrases.

## 2.6. Analysis of the data

Before any further analysis the classification following from our three judges was checked against the  $^{10}\log$  frequency obtained from Yahoo. Figure 1 gives the basic breakdown of the data, where *lexic* stands for (part of) a PLI, *lexic?* for perhaps (part of) a PLI, *novel* for (part of) a novel phrase, and *fnctn wrds* for combinations of function words.

---

Figure 1 about here

---

The data were analyzed with a simple Univariate one-way Analysis of Variance, 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 *lexic* and *lexic?* were not significantly different, whereas all other contrasts were. The very high average log frequency of combinations of function words suggests that log Yahoo frequency in itself is not a good criterion for deciding whether or not a particular word sequence is (part of) a

PLI. Whereas one may expect that many PLIs (not being proverbs) have a relatively high frequency of usage, individual highly frequent word sequences are not necessarily PLIs. There is no way of knowing whether a particular combination of function words stems from a novel phrase or from a PLI. This category is therefore excluded from further analysis. Thus the number of word sequences is reduced from 659 to 635. To this set of data, excluding the anomalous combinations of function words, a new Univariate one-way Analysis of Variance was applied, followed by a post hoc analysis using Tukey's, with estimated lexical category as fixed effect and log Yahoo frequency as dependent measure. The effect of estimated lexical category is highly significant ( $df=2$ ;  $F=110$ ;  $p<0.001$ ). There is no significant difference between *lexic* and *lexic?* and both categories differ significantly from *novel*. The big and significant difference between *lexic* and *lexic?* on the one hand and *novel* on the other, supports the intersubjective classification following from our three judges. The fact that *lexic* and *lexic?* do not differ much and not significantly in their log frequency suggests that these two categories may be collapsed in our further analysis.

A initial analysis of the repair data is presented in Figure 2. Here the actual fractions repaired as found in the corpus are presented for *lexic+* (collapsing *lexic* and *lexic?* from Figure 1) and *novel* phrases, separately for anticipations (*antic*), perseverations (*persev*) and exchanges (*exchan*).

---

Figure 2 about here

---

The data in Figure 2 at first sight are somewhat mysterious. Apart from a tendency that, as predicted, the fraction repaired is higher in novel phrases than in PLIs, there is a much greater effect on fractions repaired of the speech error class, viz. anticipations versus perseverations versus exchanges. Notably fraction repaired is

very low in the exchanges, and here also higher in PLIs than in novel phrases. This is worrying because the difference in fraction repaired and the difference in effect of lexicalizedness on fraction repaired between error categories may cause trouble for our further statistical analysis. The difference in fraction repaired is also unexpected because where anticipations and perseverations contain a single speech error, an exchange contains two speech errors, viz. an anticipation and a perseveration. This gives not one but two chances that the error is detected and repaired. One thus would expect fraction repaired to be considerably higher for exchanges than for anticipations and perseverations. The following argument, provided by Nootboom 2005, makes transparent why in a corpus of speech errors repaired exchanges are more frequent and repaired anticipations less frequent than one would expect.

When a speaker in mentally preparing an utterance for speaking, makes a phonological exchange such as *Yew Nork* for *New York*, this exchange exists in his or her internal speech a short while before it is spoken. The speaker has now several chances to detect the error. First he or she may detect and repair the error in internal speech even before the first word *New* is spoken. If so, the chances are that the error is detected and repaired before it is realized, and the external world will never know that an error has been made and repaired in internal speech. Secondly, the speaker may detect an error after the first element has been spoken and before the second element has been spoken. Note that in this situation, the speaker may either detect the anticipation already made or the perseveration that is yet to come in overt speech but already present in the speaker's internal speech. Therefore the probability for error detection should be considerably higher for exchanges than for single anticipations and perseverations. However, we cannot see this in the corpus of speech errors because all such cases of repaired exchanges are classified as repaired anticipations. Thirdly, the speaker may detect that an error has been made after the second error (in our example *Nork*) has begun to be spoken. These latter cases are the ones that in the corpus are classified as repaired exchanges. The very low fraction repaired of exchanges can now be explained by assuming that most

exchanges in internal speech that are repaired at all, are detected and repaired after the first element and before the second element of the exchange is realized in speech. All these cases end up in the corpus as repaired anticipations (cf. Nootboom 2005).

For the purpose of this paper the relevance of the above reasoning is that it makes little sense to keep anticipations and exchanges separate as they have been in the corpus of speech errors. It would make sense to keep separate the anticipations and exchanges as they are made in internal speech, because the dependent measure in the further analysis, fraction repaired, apparently is not the same for these two classes of errors. Unfortunately there is no way to tell which repaired anticipations stem from misclassified exchanges and which do not. For this reason anticipations and exchanges were collapsed into a single category, *anticipations+*, to be kept separate from the category of perseverations. This leads to the breakdown of the data presented in Figure 3.

---

Figure 3 about here

---

The data in Figure 3 were analyzed with a logistic regression using effect coding, with as dependent binomial variable 'fraction repaired' and as fixed factors 'novel versus lexicalized+' and 'anticipation+ versus perseveration'. The grand mean was used as intercept. The best fitting model showed no interaction between the two fixed factors, and showed a significant effect of both '*anticip+* versus *persev*' ( $p < .003$ ) and '*novel* versus *lexic+*' ( $p < .02$ ). Table 1 gives the relevant analysis results. For a discussion of this type of logistic regression see Johnson 2008.

---

Table 1 about here

---

That the fraction repaired is considerably and significantly higher for anticipations plus exchanges than for perseverations can be explained by the many errors in this category that are half-way repaired exchanges for which the fraction repaired is expected to be much higher than for perseverations. That the fraction repaired is significantly higher for novel than for PLIs confirms the main hypothesis tested in this paper.

### **3. Discussion**

The basic assumption underlying the current study is that the preparation and production of PLIs such as proverbs, sayings, idiomatic expressions, collocations and clichés is more automatic than the preparation and production of novel phrases. This assumption is captured in the superlemma theory of Levelt and Meyer (2000). As it is known that automatic processes in the human mind are more efficient and less error prone than novel processes, one would expect that fewer speech errors are made in PLIs than in novel phrases. Whether this is indeed the case is yet to be investigated. But it is also known that more automatic processes are less guided by monitoring than novel processes (Wheatley and Wegner 2001). One therefore expects that, if speech errors are made in PLIs, they will be less often detected and repaired by self-monitoring than speech errors in novel phrases. The current study of repaired and unrepaired phonological speech errors in spontaneous Dutch shows that this is indeed the case.

In a first attempt to demonstrate the hypothesized lower fraction repaired for PLIs than for novel phrases, the effect was found for phonological anticipations and perseverations but not for exchanges. This potentially weakened the conclusion. Therefore the deviant pattern in exchanges was in need of explanation. It was observed that speech errors classified as repaired exchanges were less frequent and speech errors classified as repaired anticipations were more frequent than one would expect. This can be explained by assuming that impending exchanges in internal speech that are repaired after the first and before the second part of the speech error

is spoken, are classified as repaired anticipations from which they cannot be distinguished. If one corrects for this misclassification, on the basis of the distribution of repaired and unrepaired perseverations, it turns out that all three classes of speech errors, anticipations, perseverations and exchanges, have a more similar distribution (cf. Nootboom 2005). This implies that in internal speech there are more exchanges than one would derive from the classification of overt errors in corpora. In the further data analysis numbers of repaired anticipations and repaired exchanges had to be collapsed, because one has no way of knowing whether specific repaired anticipations in the corpus in fact were or were not half-way repaired exchanges in inner speech. However, the hypothesized lower fraction repaired for PLIs than for novel phrases clearly stands out in the final analysis, thus confirming the main hypothesis of the current study.

The fact that common phonological (and lexical) speech errors are made in PLIs such as proverbs, sayings, idiomatic expressions, collocations and clichés supports those theories of phrasal lexical expressions that assume that each such expression not only has its own lexical entry, called superlemma by Levelt and Meyer (2000), but also that each superlemma activates the lemmas for the words that are constituents of the expression. The finding that such speech errors are significantly less often repaired in lexicalized than in novel expressions can be explained by assuming that these errors are less often detected either in inner speech (cf. Nootboom and Quené 2008) or in overt speech (for the difference between self-monitoring inner and overt speech, see Nootboom 2010). This suggests that, as one might expect, the preparation and production of PLIs is more automatic and less closely monitored than the preparation and production of novel phrases, supporting the main assumption underlying this investigation.

**Acknowledgement**

The author is grateful to Hugo Quené for his assistance in running the logistic regression analysis, and for his helpful comments on the investigation and on an earlier version of this ms.

**Access to list of word sequences:**

<http://www.let.uu.nl/~Sieb.Nooteboom/personal/PLIs&NovelPhrasesWordSequence.s.xls>

[Data in sheet 1, explanation of column headers in sheet2]



**References:**

- Bolinger, Dwight. 1975. *Aspects of Language*. New York: Harcourt, Brace, Jovanovich.
- Cacciari, Cristina. & Patrizia Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language*, 27. 668-683.
- Cutler, Anne. 1982. The reliability of speech error data. *Linguistics* 19, 561-582.
- Cutting, J. Cooper & J. Kathryn Bock. 1997. That's the way the cookie bounces: Syntactic and semantic components of experimentally controlled idiom blends. *Memory and Cognition*, 25(1), 57-71.
- Fraser, Bruce. 1970. Idioms within a transformational grammar. *Foundations of Language* 6, 22-42.
- Fromkin, Victoria A. 1973. Appendix. In Victoria A. Fromkin (ed.) *Speech Errors as Linguistic Evidence*, 243-269. The Hague: Mouton.
- Jackendoff, Ray. 1995. The boundaries of the lexicon. In Martin Everaert, Evert-Jan Van der Linden, André Schenk, Rob Schreuder (eds.), *Idioms, Structural and Psychological Perspectives*, 133-169. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Janetzko, Dietmar. 2008. Objectivity, Reliability, and Validity of Search Engine Count Estimates. *International journal of internet science* 3(1), 7-33.  
[http://www.ijis.net/ijis3\\_1/ijis3\\_1\\_janetzko.pdf](http://www.ijis.net/ijis3_1/ijis3_1_janetzko.pdf). (Accessed 19 October 2010).
- Johnson, K. 2008. *Quantitative Methods in Linguistics*. Malden MA: Blackwell.
- Ketko, Hazel. 2000. Importance of "MultiWord Chunks" in Facilitating Communicative Competence and its Pedagogic Implications. In *The Language Teacher online*, December 2000. <http://www.jalt-publications.org/tlt/articles/2000/12/ketko>. (Accessed 22 February 2010).
- Kuiper, Koenraad. 2009. *Formulaic genres*. Basingstoke: Palgrave Macmillan.
- Kuiper, Koenraad, Marie-Elaine Van Egmond, Gerard Kempen & Simone A. Sprenger. 2007. Slipping on superlemmas: Multiword lexical items in speech production. *The Mental Lexicon* 2(3). 313-357.
- Levelt, Willem. J. M. & Antje Meyer. 2000. Word for word: Multiple lexical access in speech production, *European Journal of Cognitive Psychology* 12 (4), 433-452.
- Nooteboom, Sieb G. 2005. Listening to one-self: Monitoring speech production. In Robert J. Hartsuiker, Roelien Bastiaanse, Albert Postma and Frank Wijnen (eds.)

*Phonological encoding and monitoring in normal and pathological speech*, 167-186. Hove and New York: Psychology Press, Taylor and Francis Group.

- Nooteboom, Sieb G. 2010. Monitoring for speech errors has different functions in inner and overt speech. In Martin Everaert, Tom Lentz, Hannah de Mulder, Oystein Nilsen & Arjen Zondervan (eds.) *The Linguistic Enterprise. From Knowledge of Language to Knowledge in Linguistics*, 213-233. Amsterdam and Philadelphia: John Benjamins Publishing Company
- Nooteboom, Sieb & Hugo Quené. 2008. Self-monitoring and feedback. A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language* 58, 837-861.
- Nunberg, Geoffrey, Ivan Sag & Thomas Wasow. 1994. Idioms. *Language* 70. 491-537.
- Oostdijk, N. & D. Broeder. 2004. The Spoken Dutch Corpus and Its Exploitation Environment. In *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03). 14 April, 2003*. Budapest, Hungary: EACL-03.
- Pawley, Andrew K. & Frances H. Syder, 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Jack C. Richards and Richard . Schmidt (eds.), *Language and Communication*, 191-225. London: Longman.
- Rossi, Mario & Évelyne Peter-Defare. 1998. *Les lapsus ou comment notre fourche a langué* [Speech errors or how our slip has tongued]. Paris: Presses Universitaires de France.
- Schelvis, Marianne. 1985. The collection, categorisation, storage and retrieval of spontaneous speech error material at the Institute of Phonetics, Utrecht: *PRIPU* 10, 3-14.
- Sprenger, Simone A., Willem.J.M. Levelt. & Gerard Kempen. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54, 161-184.
- Swinney, David A. & Anne Cutler. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18. 523-534.
- Van Lancker-Sidtis, Diana. 2004. When novel sentences spoken or heard for the first time in the history of the universe are not enough: toward a dual-process model of language. *International Journal of Language and Communication Disorders* 39 (1), 1-44.

Wheatley, Thalia & Daniel M. Wegner. 2001. Automaticity of Action, Psychology of. *International Encyclopedia of the Social and Behavioral Sciences*, 991-993. Amsterdam: Elsevier.

Figures

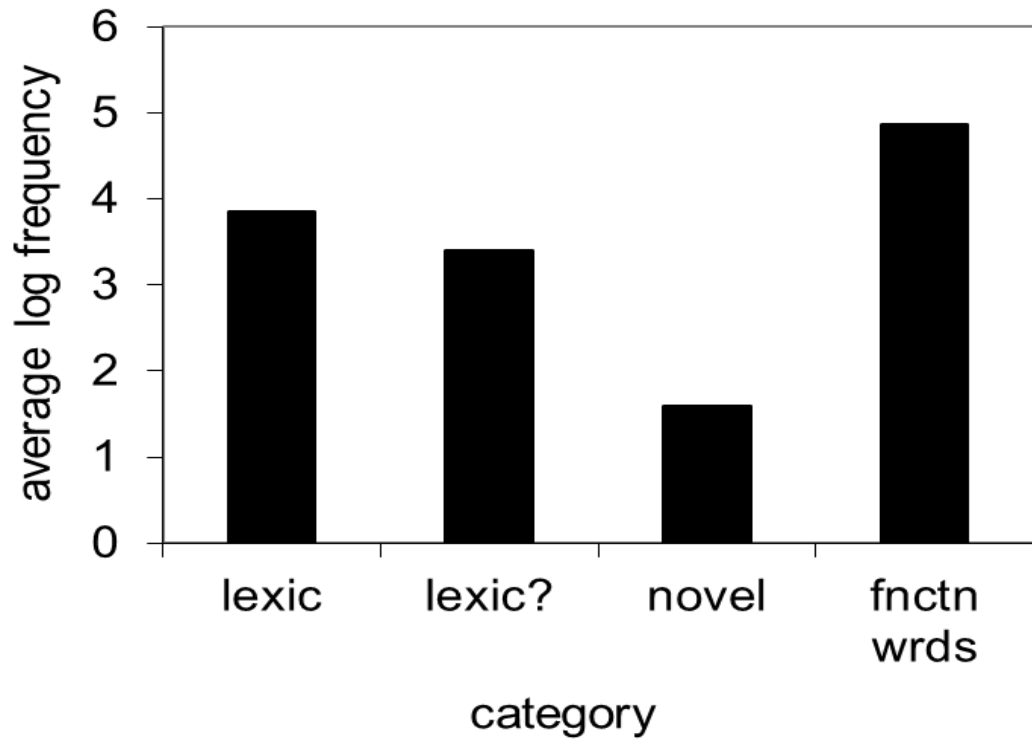


Figure 1. Average  $^{10}\log$  frequency obtained from Yahoo for four classes of word sequences, definitely lexicalized (lexic; N=92), perhaps lexicalized (lexic?; N=75), novel (novel; N=468), and combinations of function words (fcnctn wrds; N=24).

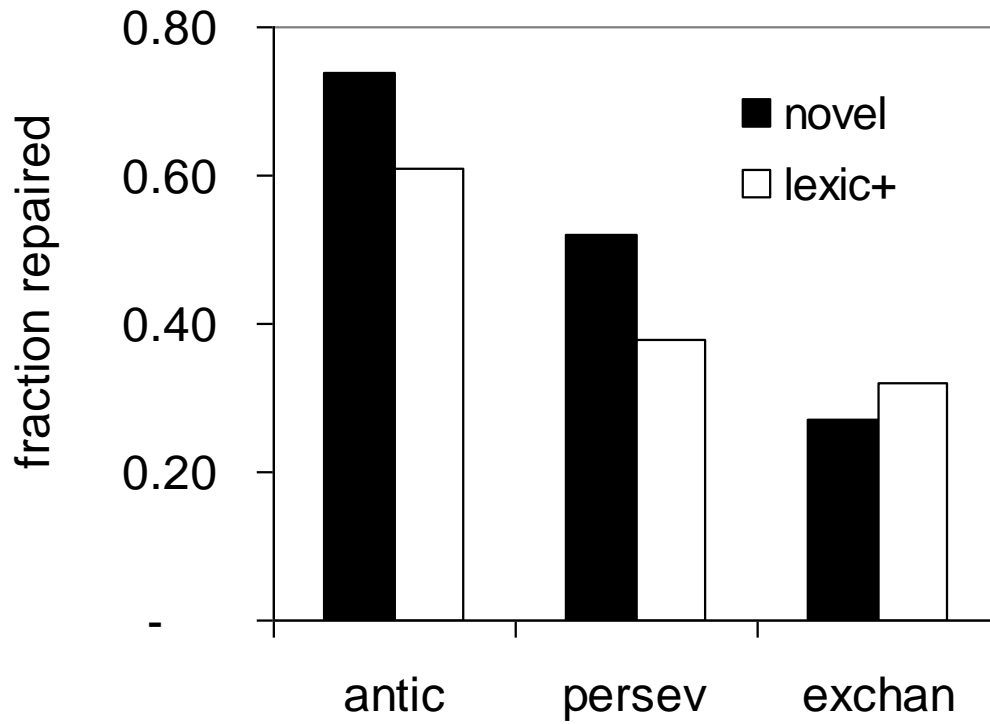


Figure 2. Fractions repaired as found in the corpus for novel and lexic+ phrases, separately for phonological anticipations (antic; N=391), perseverations (persev; N=164) and exchanges (exchan; N= 80).

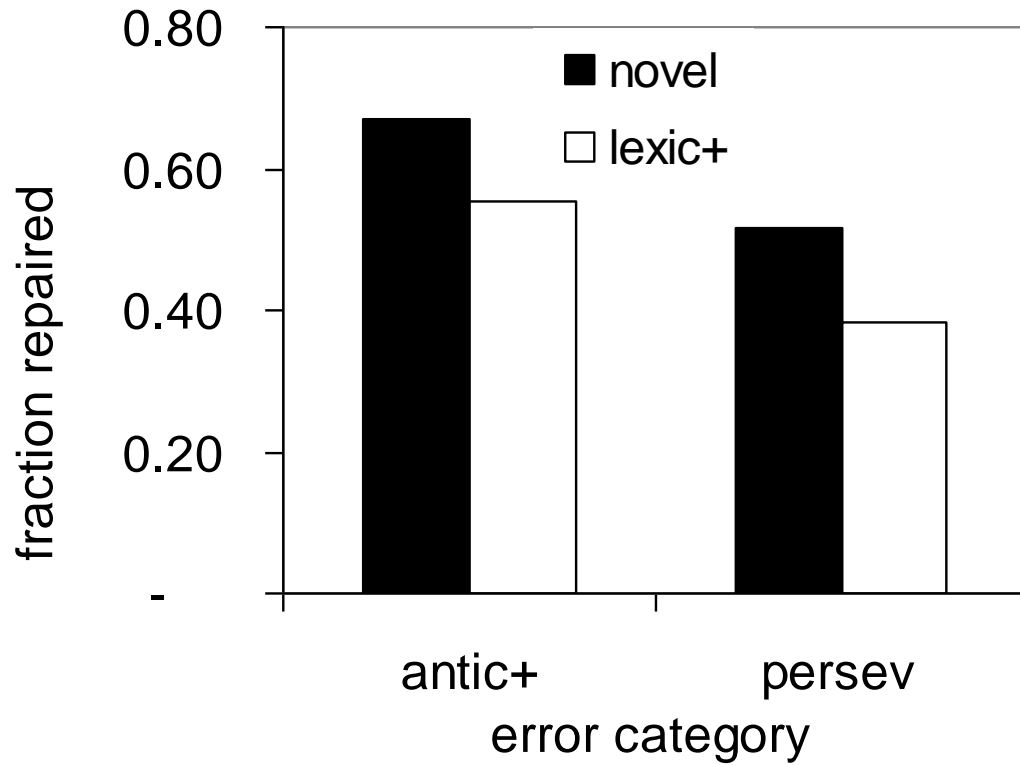


Figure 3. Fractions repaired of speech errors made in anticipations plus exchanges (antic+; N=333) and perseverations (persev; N=134), separately for novel and lexicalized (lexic+) expressions. Both main effects are significant.

## Tables

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'. For fixed effects, regression coefficients are given, with standard errors, t values and p values. Due to the structure of the data set with no repetitions for speakers or for word sequences, there are no random effects.

effects	coef.	s.e.	t	p
intercept=grand mean	0.594	0.024	24.01	<0.001
novel/lexic+:	+/-0.129	0.055	-2.33	< 0.020
antic+/-persev:	+/-0.151	0.049	-3.05	< 0.003
interaction:	0.020	0.110	0.18	< 0.860