Article:

Plug, L (2016) Informativeness, Timing and Tempo in Lexical Self-Repair. Language and Speech, 59 (4). pp. 516-543. ISSN 0023-8309

https://doi.org/10.1177/0023830915618427

INFORMATIVENESS, TIMING AND TEMPO IN LEXICAL SELF-REPAIR

Author: Leendert Plug

Affiliation: University of Leeds

Postal address: Linguistics and Phonetics, University of Leeds, Leeds LS2 9JT, United Kingdom

Telephone: +44 113 343 3189

E-mail: l.plug@leeds.ac.uk

Running head: Temporal organization in lexical repair

Acknowledgements: This work was supported by ESRC grant RES-061-25-0417 'Prosodic marking revisited: The phonetics of self-initiated self-repair in Dutch'. I would like to thank Paul Carter for his contribution to data selection, phonetic analysis and preliminary modelling; Mirjam Ernestus for her generous offers to provide frequency counts and recruit raters for the fill-in-the-gap task; and Christina Englert, Sanne Berends, Leen Plug, Gerard Berends, Anne van Hoek, Erik Verboom, Ilse Pit, Lisa Rommers, Nadia Klijn, Stef Piers, Thera Baayen, Marianne Plug and Ilona van der Linden for their participation in the rating exercises. I am also grateful to three Language and Speech reviewers for their constructive comments on three drafts of this article.

Abstract

This paper presents a study of the temporal organization of lexical repair in spontaneous Dutch speech. It assesses the extent to which offset-to-repair duration and repair tempo can be predicted on the basis of offset timing, reparandum tempo and measures of the informativeness of the crucial lexical items in the repair. Specifically, we address the expectations that repairs that are initiated relatively early are produced relatively fast throughout, and that relatively highly informative repairs are produced relatively slowly. For informativeness, we implement measures based on repair semantics, lexical frequency counts and cloze probabilities. Our results highlight differences between factual and linguistic error repairs, which have not been consistently distinguished in previous studies, and provide some evidence to support the notion that repairs that are initiated relatively early are produced relatively fast. They confirm that lexical frequency counts are rough measures of contextual predictability at best, and reveal very few significant effects of our informativeness measures on the temporal organization of lexical self-repair. Moreover, while we can confirm that most repairs have a repair portion that is fast relative to its reparandum, this cannot be attributed to the relative informativeness of the two portions. Our findings inform the current debate on the division of labour between inner and overt speech monitoring, and suggest that while the influence of informativeness on speech production is extensive, it is not ubiquitous.

Introduction

In this paper we report on a phonetic analysis of instances of lexical self-repair such as *I'm going by ca- by bus*, in which one lexical choice — here car — is rejected in favour of another — here bus. While a good deal is known about the various types of disfluency involved in the initiation of self-repair (see e.g. Nakatani & Hirschberg 1994, Shriberg 2001, Jasperson 2002, Benkenstein & Simpson 2003), relatively few studies have addressed the question of how the phonetics of the second — preferred — lexical item compare to those of the first — rejected — one. In this paper, we focus on how the two items compare in temporal terms. The temporal organization of self-repair is interesting for at least two reasons: first, it can provide us with valuable insights into the coordination of speech planning and production processes, and second, self-repair provides a useful context for exploring the relationship between informational redundancy and articulatory reduction.

Coordination of Speech Planning and Production Processes

For psycholinguists, self-repair 'may reveal principles of organization of the speech production process that would be hard to discover on the basis of laboratory data alone' (Levelt 1984: 105). Various studies have focused specifically on the temporal organization of instances of self-repair, in order to explore the functions and temporal coordination of 'self-monitoring' processes (see e.g. Blackmer & Mitton 1991, Oomen & Postma 2001, 2002, Seyfeddinipur et al. 2008).

For example, Blackmer & Mitton's (1991) finding that a substantial proportion of self-corrections involve no delay between the abandonment of the erroneous lexical item and the onset of the repair item cast doubt on the widely held assumption that repair planning must start at the abandonment of the erroneous item (see Levelt 1989). For many repairs it seems plausible that the speaker first detects the error after having articulated it, through a process variously called 'overt speech', 'auditory loop' or 'post-articulatory monitoring'

(Levelt et al. 1999, Oomen & Postma 2001, 2002, Nooteboom 2005b, Hartsuiker et al. 2005a, b), and subsequently plans a reformulation. However, when the reformulation follows the abandonment of the ongoing utterance without delay, repair planning must have started earlier. In these cases, the error is detected during the compilation of the speech plan, through a process variously called 'inner speech', 'inner loop' or 'pre-articulatory monitoring' (Levelt et al. 1999, Oomen & Postma 2001, 2002, Nooteboom 2005b, Hartsuiker et al. 2005a, b), simultaneous with ongoing articulation (Hartsuiker & Kolk 2001, Tydgat et al. 2011).

A point of debate in the literature on self-monitoring and repair is whether inner and overt speech monitoring are distinct processes (see Postma 2000 and Hartsuiker et al. 2005b for relevant reviews). According to Levelt (1983, 1989) and Levelt et al. (1999), both monitors employ the same comprehension system that is also used in the perception of speech produced by others: the only difference between the processes is in the nature of the input. Nooteboom (2005a, b, 2010), on the other hand, argues that inner and overt speech monitoring serve distinct purposes in the speech production process, and that these different purposes help to explain some of the variation in the temporal and prosodic make-up of instances of self-repair.

Nooteboom (2010: 215) suggests that the purpose of inner speech monitoring is to 'prevent errors … from becoming public'; therefore, major characteristics of the process are that it operates under time pressure, and that it aims to minimize disfluency in production. Once the erroneous form has been produced, on the other hand, it is clear to the speaker that fluency will need to be sacrificed and 'the speaker should take his or her time to make clear to the listener that an error has been made' (Nooteboom 2010: 216). This functional difference between inner and overt speech monitoring offers a straightforward explanation of the finding that repairs in which an erroneous word is interrupted after only one or two segments have significantly lower offset-to-repair durations — that is, shorter delays between the abandonment of the utterance and the first word of the repair — than repairs in which an erroneous word is completed before the utterance is interrupted (Nooteboom 2010: 223–224; see also Seyfeddinipur et al. 2008). Presumably, inner speech monitoring is responsible for the error detection in the former case, while overt speech monitoring accounts for the latter. Nooteboom (2010) also reports prosodic differences between 'early offset' and 'late offset' repairs, which he again attributes to the different functions of the two self-monitoring mechanisms that give rise to these repairs.

So far, few studies have examined whether the empirical patterns observed by Nooteboom (2010) in experimentally elicited speech errors generalize to other collections of self-repairs. Plug & Carter (2014) observe patterns consistent with Nooteboom's findings in spontaneous phonological repairs. In this paper, we investigate offset-to-repair duration and repair tempo in a dataset of lexical self-repairs sampled from spontaneous Dutch speech. If Nooteboom's account is on the right lines, and repairs that follow an early interruption of the reparandum are executed under a higher degree of time pressure than repairs that follow a later interruption of the ongoing utterance, we would expect to find significant effects of what we will call offset timing on offset-to-repair duration and repair tempo.

Informativeness and Articulatory Reduction

Repair tempo is also of interest because self-repair provides a useful context for testing predictions regarding the relationship between informativeness and articulatory reduction. It is generally understood that 'parts of the speech stream that carry little information are realized with less articulatory effort than more informative parts' (Pluymaekers et al. 2005a: 157; see also Kohler 2000, Barry & Andreeva 2000, Aylett & Turk 2006, Seyfarth 2014); in Lindblom's (1990, 1996) terms, informativeness is closely correlated with articulatory variation along a 'hypo–hyper continuum'. In lexical self-repair, a speaker interrupts the speech stream to replace one word by another, because the first is either factually or linguistically erroneous, or pragmatically infelicitous. On the face of it, this may suggest that the second, preferred lexical item — henceforth the repair item — is more predictable and better fitted to the context, and therefore less informative than the first — henceforth the reparandum item. Therefore, it makes sense to predict that in most cases, a repair item should be hypo-articulated relative to the reparandum item. In temporal terms, relative hypo-articulation goes together with temporal compression. In other words, based on considerations of informativeness, one might expect a local rise in articulation rate between the reparandum and repair items.

However, this prediction may be complicated by the frequency characteristics of lexical self-repairs. Kapatsinski (2010) shows that in American English lexical repairs, the lexical frequencies of reparandum and repair items are positively correlated, but reparandum items are on average more frequent than repair items. This makes sense in terms of the likelihood of lexical activation of more and less frequent items: the more frequent the intended lexical item, 'the more frequent a competitor needs to be to become activated before [it] and thus be erroneously uttered in production' (Kapatsinski 2010: 87). Various studies of the relationship between informativeness and articulatory reduction have used lexical frequency as a measure of informativeness (e.g. Jurafsky et al. 2001, Bell et al. 2003, Pluymaekers et al. 2005b, Baker & Bradlow 2009): the more frequent a lexical item is across large amounts of language use, the more predictable it is in individual instances, and therefore the more likely it is to be hypo-articulated (see Bybee 2002, 2010). In the case of lexical self-repair, this reasoning supports a prediction of a local fall in articulation rate between the reparandum and repair items: a decrease in lexical frequency means a increase in informativeness, and therefore an increase in the likelihood of relative hyper-articulation.

Moreover, there is clearly variation among repairs in terms of their informational salience. As indicated above, lexical self-repair can involve the correction of factual or linguistic error: the reparandum item can be grammatically appropriate but semantically incorrect, as in *I'm going by ca-* by bus, or grammatically inappropriate, as in *I'm goin*g on bu- by bus. Levelt (1983) calls such repairs 'error repairs'. Lexical self-repair can also involve the correction of a pragmatically infelicitous word choice, as in *Here's my gir-* my daughter, or an insufficiently specific one, as in *It's a bird* – uh, a parrot. Levelt (1983) calls such repairs 'appropriateness repairs'. As Levelt (1983, 1989) points out, error repairs can be considered more informative than appropriateness repairs: the former are crucial for the listener's understanding of the current utterance; the latter refine an utterance that is already propositionally accurate. In support of this idea, Levelt & Cutler (1983) report that error repairs are more likely than appropriateness repairs to be 'prosodically marked' — produced with noticeable pitch and intensity prominence on the repair item. In terms of temporal organization, we might expect error repairs to be more likely than appropriateness repairs to be produced with relative hyper-articulation of the repair item: that is, a local fall in articulation rate following the reparandum.

So far, the only examination of the relationship between informativeness and repair tempo is that of Plug (2011). Consistent with the first prediction above, Plug (2011) reports a predominance of temporal compression — that is, a relative speeding up after the repair initiation — in a collection of Dutch self-repairs. Plug reports no significant effect on repair tempo of the error–appropriateness distinction and the frequency differential between repair and reparandum items. However, a weakness of Plug's study is that it does not differentiate between several types of repair: as well as lexical repairs, its dataset contains phonological repairs — which involve the production of one lexical item only — and grammatical repairs — whose extent can be difficult to delimit. Furthermore, Plug (2011) considers the relevance of unigram frequency counts only. The study on which we report in this paper is firstly restricted to the most pertinent type of repair for testing predictions regarding the relationship between informational redundancy and articulatory reduction — lexical repair — and assesses the relevance of multiple semantic and probabilistic measures in analysing the repairs' temporal organization. Moreover, offset-to-repair duration is considered alongside repair tempo, as measures of informativeness are likely to have an impact on the speed of repair onset, too: monitoring for appropriateness issues is likely to be slower than that for errors (Levelt 1989, Postma 2000, Kormos 2000, Kapatsinski 2010) , and a high-frequency repair word is accessed more quickly than a low-frequency one (Harley & MacAndrew 2001, Kapatsinski 2010).

This Study

This paper reports on a study of the temporal organization of lexical repairs sampled from a corpus of spontaneous Dutch speech. The central question addressed in the study is how well we can predict the duration of the offset-to-repair interval and the tempo of the repair component — implemented here in the form of articulation rate in segments per second — based on the tempo of the reparandum and various other relevant factors. In particular, we address three general hypotheses.

- **HYPOTHESIS A —** Measures of offset timing (where 'early offset' is an interruption before the end of the reparandum item, and 'late offset' after its completion) are significant predictors of offset-to-repair duration and repair tempo.
- **HYPOTHESIS B —** A semantically-based classification of repairs (in which 'error' and 'appropriateness' repairs are distinguished) is a significant predictor of offset-to-repair duration and repair tempo.
- **HYPOTHESIS C —** Measures of lexical frequency and contextual predictability for reparandum and repair items are significant predictors of offset-to-repair duration and repair tempo.

HYPOTHESIS A is consistent with Nooteboom's (2010) observations on the temporal organization of phonological repair. Our findings in relation to this hypothesis may provide further support for a qualitative difference between repairs initiated through inner and overt speech monitoring. HYPOTHESIS B is consistent with Levelt's (1983, 1989) reasoning regarding the relative informativeness of semantically-based subtypes of lexical repair, and Levelt & Cutler's (1983) findings in relation to prosodic marking. HYPOTHESIS C is consistent with Kapatsinski's (2010) findings on the frequency characteristics of lexical repair, and findings on the relationship between informativeness — as quantified through measures of lexical frequency and contextual predictability — and articulatory reduction. Our findings in relation to these hypotheses may provide insight into the division of labour between multiple parameters of informativeness in constraining articulation.

Based on previous findings, we can formulate more concrete expectations as to the directions of hypothesized effects, as well as a number of expections regarding interactions between measures of offset timing, repair semantics, frequency and predictability. We do this below, following a description of the methods used in this study.

Method

Data Selection

The dataset for this paper comprises 209 instances of lexical repair extracted from four sub-corpora of the Spoken Dutch Corpus (Oostdijk 2002), containing spontaneous face-to-face conversations, interviews with teachers of Dutch, broadcast interviews, discussions and debates, and non-broadcast interviews, discussions and debates. We initially searched for words coded as interrupted and for a selection of lexical editing terms, as well as performing a number of unsystematic data trawls. We discarded a considerable number of potential instances because of poor audio quality or overlapping speech.

In order to make the dataset as homogeneous as possible, we applied the following inclusion criteria. First, we left aside instances in which the reparandum item was left incomplete and either no reasonable guess could be made as to its identity, or several candidate identities presented themselves. This selection was done by the author in the first instance, and was later verified by the linguist who assisted in the semantic classification of the repairs, as described below. Second, we left aside turn-initial and turn-final instances of repair, to minimize the influence of major prosodic boundaries on repair tempo (Jacewicz et al. 2000, Quené 2008). Third, we left aside instances in which the repair is accompanied by markers of disfluency that suggest hesitation or word searching between reparandum offset and repair, including marked 'stretching' of final reparandum sounds, repeated use of er, and multiple attempts at starting the repair item (see Fox Tree & Clark 1997, Shriberg 2001). Fourth, we left aside instances that could be attributed to segment substitutions — such as tar- car talk, where the speaker erroneously selects the lexical item tar, or mispronounces car (Shattuck-Hufnagel & Cutler 1999) — if a plausible substitution trigger — here talk — could be identified. The scope of our investigation is restricted, then, to semantically transparent, turn-medial, reasonably fluent, unambiguously lexical repairs.[1]

(1) contains representative examples from our dataset. The reparandum and repair items are in bold.

(1)
a. met de **au-** met de **bus** ('by ca- by bus')
b. als er met tekst **gebrui- gewerkt** wordt ('when one use- works with text')
c. de **koelka- koelcel** ('the refrigera- cold store')
d. die **drie** da- of die **twee** dagen ('those three day- or those two days')
e. een **leuke** k- een **mooie** keuken ('a nice k- a beautiful kitchen')
f. **een** telefoon- of **mijn** telefoonnummer opschrijven ('write down a phone- or my phone number')
g. in de **computerwe-** uh in de **bankwereld** ('in the world of compu- er of banking')

The examples in (1) illustrate that some cases the reparandum item is cut off prematurely, as in (a), (b), (c) and (g), and in others it is completed, as in (d) to (f). In some cases, lexical material preceding the reparandum item is repeated in the repair, as in (a), (d), (e) and (g);

[1] These restrictions should minimize the likelihood, raised by an anonymous reviewer, that a relatively high articulation rate in the repair stretch signals a return to 'normal' tempo, as opposed to constituting a marked 'speeding up'. We do not address this issue directly, restricting our attention to tempo relations within the narrow domain of the repair itself. We note that the former interpretation is inconsistent with impressionistic observations by Goffman (1981) and Cutler (1983), who describe repair tempo in terms of increases or decreases relative to surrounding talk.

and in some cases, the repair is initiated by an editing term such as of in (d) and (f) or uh in (g). We will return to some of these characteristics below.

Phonetic Analysis

We segmented all instances of repair in PRAAT (Boersma & Weenink 2010), as illustrated in Figure 1. We placed boundaries at the start and end of the reparandum and repair stretches, including any repeated lexical items, and at the start and end of the reparandum and repair items. Editing terms were segmented as part of the offset-to-repair interval, between the abandonment of speech and the start of the repair proper. We followed the segment-level segmentation criteria set out by Rietveld & Van Heuven (1997) throughout. We calculated the articulation rate for each segmented portion by dividing the number of surface segments articulated during the portion by its raw duration. Surface segments were transcribed by an experienced phonetician with no particular knowledge of Dutch, on the basis of auditory analysis and concurrent inspection of waveforms and spectrograms. All transcriptions were checked and approved by the author, who is a native speaker of Dutch. In what follows, we will call the articulation rate of the reparandum stretch Reparandum rate and that of the repair stretch — our primary dependent variable — Repair rate.[2] We will refer to the duration of the offset-to-repair interval as Offset-to-repair duration.

Figure 1 about here

Offset Timing

In order to assess whether repairs with a reparandum item that is interrupted early have different temporal characteristics from repairs with a completed reparandum item, following Nooteboom's (2010) findings on phonological error repairs, we classified each reparandum

[2] The repair stretch includes any lexical items repeated from the reparandum. Exploratory analysis not reported here (but see Plug & Carter 2011) suggested that modelling repair articulation rate including and excluding repeated items reveals similar data patterns. This is consistent with the findings of Plug (2011), who reports an explicit comparison of these alternative measures.

item as interrupted or completed prior to repair, as illustrated in (1). All morphologically complex words, including compounds, were treated as single words for this purpose: in other words, (1g) is considered interrupted even though the crucial reparandum morpheme, computer, is a free morpheme and is completed prior to the repair. Such complex reparandum items constitute less than 10% of the dataset, and exploratory analysis (not reported here) suggested that treating them differently would not alter the main findings reported below.

Following Plug & Carter (2014), we also explored the relevance of a proportional measure of reparandum item completeness, on the assumption that this might capture more fine-grained differences between 'early offset' and 'late offset' repairs. To implement this, we divided the number of segments produced between the start of the reparandum item to the abandonment of speech prior to repair by the number of segments in the (projected or completed) reparandum item. We ignored segment deletions for this purpose: the crucial question was which segment in a canonical realisation of the word in question was reached in the surface form. We referred to Heemskerk & Zonneveld (2000) for transcriptions. Note that the measure is not bounded by one: instances in which the speaker produces further lexical material following the reparandum item, but prior to repair, result in values above one. All other things being equal, the higher the value, the later the repair.

In what follows, we will call the binary completeness variable Completeness, and the proportional variable Proportional completeness.

Repair Semantics

In order to assess the predictive value of Levelt & Cutler's (1983) 'error' versus 'appropriateness' dichotomy, we classified all instances as error or appropriateness repair using the criteria set out by Levelt (1983) and, more recently, Kormos (1999). Instances in which the denotations of the two lexical items are mutually exclusive, as in (1a), (1d) and (1g) above, or in which the first lexical choice result in an ill-formed collocation, as in (1b), can be considered error repairs. Instances in which the denotations of the two lexical items are highly similar, as in (1c) and (1e), can be considered appropriateness repairs. In these cases, the first lexical choice is treated as ill-judged by the speaker, but is not factually or linguistically erroneous. Instances in which the second lexical item can be seen as more specific than the first, as in (1f), can also be considered appropriateness repairs.

The classification procedure we followed was the same as that described by Plug & Carter (2013). The classification was done by two raters: the author, who is a native speaker of Dutch, and a Dutch linguist with a research specialisation in discourse studies. The latter was not involved with any other aspects of this study. The dataset considered contained 222 instances. The second rater verified that the author's interpretations of incomplete reparandum items were correct in all cases. The two raters then classified all instances independently. They proposed the same classification for 201 instances (91%). They considered the 21 cases of disagreement in more detail, in some cases taking a wider context around the repair into consideration, and reached a consensus classification for 15. The remaining six instances, for which the raters agreed that either classification could be proposed, were excluded from further analysis. A further seven instances were excluded when we obtained predictability ratings (see above), leaving the 209 instances on which we report in this paper. Among these 209 instances, error repairs outnumber appropriateness repairs (N=128 and N=81, respectively). In what follows, we will refer to the error–appropriateness classification by its variable name, Repair type.

Following Plug & Carter (2013), we also assessed whether factual and linguistic errors give rise to different repair tempos, given the distinct levels of processing involved in error detection. For this purpose, the author further classified the 128 confirmed error repairs accordingly.[3] All instances in which the reparandum item would have resulted in a clearly ill-formed collocation, as in (1b) were classified as linguistic errors; all others, including (1a), (1d) and (1g), as factual error repairs. Repairs of factual errors outnumber those of linguistic errors (N=92 and N=36, respectively). In what follows, we will refer to this more fine-grained implementation of Repair type as Error type.

Frequency and Predictability

Lexical Frequency

In order to evaluate the influence of lexical frequency on the temporal organization of our repairs, we took two types of measurement. First, for comparison with Kapatsinski (2010), we took unigram frequency counts for the reparandum and repair items from CELEX (Baayen et al. 1995).[4] In addition to entering the (log-transformed) counts straight into our quantitative analysis, we subtracted the reparandum count from the repair count to yield a measure of the frequency differential between the two lexical items involved in the repair. Positive values correspond to a repair item that is more frequent than the item it replaces; negative values to a repair item that is less frequent.[5] In what follows, we will call these unigram frequency variables Reparandum word frequency, Repair word frequency and Word frequency delta.

Second, given the findings reported by Aylett & Turk (2004), Seyfarth (2014) and others, we took bigram counts for the reparandum and repair items with preceding and following words, from the Spoken Dutch Corpus (Oostdijk 2002). Again, we subtracted each reparandum count from its corresponding repair count to yield a measure of the bigram

[3] It was deemed unnecessary to involve the second rater in the further classification, as this could be based on notes recorded by both raters for the purpose of the error–appropriateness classification.

[4] Our analyses included both word and lemma counts. These revealed the same data patterns, so we report on results for one — the former — only.

[5] We also tried residualizing repair frequency values using reparandum ones, analogous to our modelling of Repair rate with Reparandum rate as a control variable. We found that for our unigram as well as bigram frequency variables, (standardized) delta values are almost equivalent to (unstandardized) residuals — therefore, both methods yield the same prediction of other variables.

frequency differential between the two lexical items involved in the repair.[6] In what follows, we will call these bigram frequency variables Reparandum prior bigram, Reparandum next bigram, Repair prior bigram, Repair next bigram, Prior bigram delta and Next bigram delta.

Contextual Predictability

In order to assess whether more context-sensitive measures of repair item predictability might have predictive value, we estimated cloze probabilities through a fill-in-the-gap task (Miellet et al. 2007, Schotter et al. 2014, Burdin & Clopper 2015). For this purpose, all instances were transcribed in their phrasal context with the reparandum item present and the position of the repair marked, but the identity of the repair item withheld.[7] Incomplete repairable items were completed for clarity, and repairable items were highlighted. Editing expressions and lexical items that were repeated as part of the repair were included. For example, the repairs in (1a) and (1b) above were partially rendered as (2a) and (2b) respectively.

(2) a. met de **au-** met de **bus**
➔ met de **auto** met de ___
b. als er met tekst **gebrui- gewerkt** wordt
➔ als er met tekst **gebruikt** ___ wordt

The question for raters was which lexical item they considered most likely to have occurred in the position marked by the underscore. Where relevant — for example, where the repair item had been mentioned before in a similar formulation, or where correct factual information could be gleaned from prior context — prior discourse was briefly summarized. In order to make the measure of predictability as fine-grained as practically possible, raters were asked to provide up to two candidate repair items, ranked as first and second choice.

---

[6] Our analyses also included measures of 'mutual information', derived from bigram and unigram counts using the formula provided by Pluymaekers et al. (2005a). These did not reveal any data patterns that the bigram measures do not capture, so we leave them aside here.

[7] The data set for this task comprised 216 instances. Subsequent analysis revealed that in two cases, a transcription error had been made, rendering the raters' judgements unreliable; and in five cases the repair involved a part-of-speech mismatch between reparandum and repair items, which means the repair could be analysed as grammatically rather than lexically motivated. These instances were excluded from further analysis, leaving the dataset of 209 instances on which we report in this paper.

It was deemed appropriate to show the repairable items in a full clausal context and with the repairable lexical item highlighted. Ideally, the raters' judgements should reflect speakers' estimations of listeners' ability to predict the repair item. While the speaker's estimation must be made before the repair — in other words, before the listener has been made aware of following clausal context — the speaker can be assumed to already have a fairly detailed plan of the remainder of the clause at that point (see Levelt 1989), which may well inform the estimation. Moreover, while a listener faced with a repair initiation does not receive explicit guidance as to what aspect of the preceding utterance might be problematic, the error or infelicity will be salient to the speaker. Of course, not providing raters with following context and not highlighting the reparandum items would also have made the task considerably more difficult and time-consuming.

Twelve native speakers of Dutch provided judgements. Two raters are retired secondary school teachers; all other raters are studying for, or have completed, a higher education degree. The inclusion of secondary school teachers is particularly appropriate since a subset of repairs in our dataset are from interviews with teachers conducted by a teacher, and some involve terminology that members of other professions may not be familiar with. Given their occupational backgrounds, then, these raters could be assumed to closely resemble the interviewers for whom the speakers were designing their talk in terms of relevant professional knowledge.

We quantified responses using a scale between zero and four. Four points were awarded if the rater correctly guessed the repair item, and provided it as first and only choice. Three points were awarded if the rater correctly guessed the repair item and provided it as first choice alongside an incorrect second choice. Two points were awarded if the rater correctly guessed the repair item, but provided it as second choice only. One point was awarded if the rater provided an answer, but did not correctly guess the repair item. No points were awarded if the rater did not provide any answer. Reliability analysis reveals that the raters' responses were highly consistent, yielding a Cronbach's Alpha of 0.917 and Intra-Class Correlation Coefficient of 0.915 (two-way random model, average measures, 95% confidence interval 0.899–0.930). We averaged scores across raters, and will call the resulting variable Repair item predictability.

Statistical Analysis

Our general method in modelling Offset-to-repair duration and Repair rate was to construct linear mixed effects regression models with and without individual candidate predictors from the set described above, and use likelihood ratio tests to assess whether the inclusion of the relevant predictor contributed significantly to the model fit (see Baayen 2008, Tagliamonte & Baayen 2012). We used the lme4 package (Bates et al. 2014) in R (R Development Core Team 2008) for this purpose.

For each final model, we also constructed a corresponding conditional inference regression tree using the party package in R (Hothorn et al. 2006). Given a dependent variable and a set of candidate predictor variables, the conditional inference regression tree algorithm establishes which predictor variables give rise to homogeneous sub-groupings of observations with respect to the levels of the dependent variable, and outputs a tree diagram in which each predictor variable that motivates a sub-grouping is represented as a node. The algorithm works recursively, in that given the identification of multiple significant predictors in a data set, the data is first split into two subsets according to the strongest predictor. As explained by Tagliamonte & Baayen (2012), this makes the algorithm robust in the face of collinearity among predictors. It is therefore a useful complement to linear regression modelling in analyses involving multiple correlated predictors, and can be used to assess the robustness of linear models' fixed effects. For recent applications in linguistic and phonetic studies, see Plug & Carter (2013, 2014), Tagliamonte (2014) and Strycharczuk et al. (2014).

In addition to the variables listed so far, which can be taken as our crucial candidate predictors, we included a number of other variables which might have some effect on the temporal organization of the repair. First, the articulation rate of the reparandum is expected to have a strong effect on that of the repair: therefore, Reparandum rate is a crucial control variable to include in a model of repair tempo. Second, we included the speaker's identity (Speaker) as a random effect.[8] Third, we included a measure of the difference in phonological length between the two crucial lexical items involved in the repair, on the assumption that all other things being equal, a longer target word might give rise to a higher articulation rate than a shorter one (Nooteboom 1972, Jacewicz et al. 2000). To implement this measure, we subtracted the number of segments in the (projected or completed) reparandum item from that in the repair item. We will refer to this variable as Lexical segments delta in what follows. Prior to statistical modelling, we (natural) log-transformed all values derived from duration measurements, segment counts and lexical frequency counts — including articulation rate — in order to make their distributions as close to normal as possible. We also centred values derived from lexical frequency counts to facilitate comparison of coefficients across variables. For ease of reference, Table 1 lists all of the variables described above.

Table 1 about here

Relationships among Predictor and Control Variables

Before modelling Offset-to-repair duration and Repair rate using the predictor and control variables listed in Table 1, we first inspected the latter variables' distributions in order to check their compatibility with similar variables analysed in previous studies, and assess

[8] We also assessed the relevance of speaker gender, language variety (Netherlands Dutch versus Flemish Dutch) and the sub-corpus from which each instance was sampled. None of these factors revealed significant data patterns, so we leave them aside in what follows.

whether any systematic relationships hold among them: these would need to be taken into careful consideration in subsequent modelling. Based primarily on the findings reported by Levelt (1989) and Kapatsinski (2010), we can formulate the following concrete expectations:

- **EXPECTATION 1 —** Unigram frequencies for reparandum and repair items are significantly correlated, and reparandum items more frequent than repair items.
- **EXPECTATION 2 —** High-frequency reparandum items are more commonly completed prior to repair than low-frequency items.
- **EXPECTATION 3 —** There is no significant relationship between repair semantics and reparandum item frequency.
- **EXPECTATION 4 —** Error repairs more commonly involve a premature abandonment of the reparandum item than appropriateness repairs.

We have discussed EXPECTATION 1 above: as pointed out by Kapatsinski (2010: 87), this pattern makes sense in terms of the likelihood of lexical activation of more and less frequent items. EXPECTATION 2 is based on Kapatsinski's (2010) crucial finding, which he takes as evidence for frequency of use leading to automaticity of production. In order to rule out a confounding effect of repair semantics (see EXPECTATION 4), Kapatsinski (2010: 91) explicitly tests the hypothesis that the significant relationship between frequency and reparandum completeness in his dataset can be attributed to a significant relationship between frequency and repair semantics — and reports no supporting evidence. EXPECTATION 4 is based on findings reported by Levelt (1989), Van Hest (1996) and Kormos (2000). Kormos (2000: 155) attributes a significant difference between appropriateness and error repairs in 'error-to-cut-off time' to a difference in detection speed, explained in turn in terms of the different levels of processing involved (see Postma 2000: 104, Kapatsinski 2010: 88). Levelt (1989: 481) questions this line of reasoning and suggests a 'pragmatic' account, centred on the assumption that '[b]y interrupting a word, a speaker signals to the addressee that that word is an error'.

In addition, we were interested in the relationship between our frequency counts on the one hand, and our elicited cloze probabilities on the other, given that frequency counts are often used as measures of contextual predictability — which our cloze probabilities should model more directly. Finally, an important question in the context of our study is whether any of our predictor variables are systematically related to Reparandum rate — our crucial control variable in modelling Repair rate.

Relationships among Frequency and Predictability Variables

First, as indicated above, Kapatsinski (2010) reports that in a collection of American English repairs, unigram frequencies for reparandum and repair items are significantly correlated, and reparandum items are generally more frequent than repair items. The former is clearly the case for our repairs, too (for the entire dataset, Pearson's r=0.794, p<0.001),[9] but the latter is not. Across the dataset, instances with a more frequent reparandum item — therefore a negative delta value — only marginally outnumber instances with a more frequent repair item (98 vs 97, with 14 instances yielding zero values for both items). Further inspection of the data suggests these generalizations hold for appropriateness, factual error and linguistic error repairs alike, and our bigram measures do not differ significantly from our unigram measures in this respect. EXPECTATION 1, then, is only partly met.

Figure 2 about here

It can be inferred from Figure 2 that Word frequency delta is positively correlated with Repair word frequency, with a regression line that starts just below the horizontal and ends just above it. We observe the same general patterns for our repair and reparandum item bigram measures, which are mostly significantly correlated with the unigram measures. Notably, Repair item predictability is weakly correlated with Repair prior bigram only (r=0.203, p=0.003). This confirms that lexical frequency counts are rough measures of

[9] All correlations reported as significant in this paper reach significance when computed using Pearson's r as well as Spearman's ρ. In some cases, the latter computation is appropriate given the distributions involved; however, for consistency we report Pearson's correlations throughout.

contextual predictability at best (Schotter et al. 2014), and validates the inclusion of cloze probabilities in this study.

Relationships across Predictor Variable Groups

Table 2 summarizes the results of our analysis of potential interactions between our measures of offset timing, repair semantics and frequency and predictability. For each of the three potential interactions across the variable groups, the table lists the most robust significant relationship between individual variables in the two groups in question. Relationships involving additional variables are mentioned below, as relevant.

Table 2 about here

With reference EXPECTATION 2 above, Table 2 shows that like Kapatsinski (2010) we find significant correlations between measures of the completeness of the reparandum item (Proportional completeness, as well as Completeness), and its frequency (Reparandum word frequency along with related bigram variables), such that high-frequency reparandum items are relatively likely to be completed prior to repair. Closer inspection, illustrated in Figure 3, suggests that this generalization holds for appropriateness repairs and, most clearly, linguistic error repairs, but not for factual error repairs. For linguistic error repairs, all instances with a Proportional completeness value above one have a frequency value close to the maximum observed across instances. For appropriateness repairs, the frequencies of completed reparandum items are more dispersed, although most are above the mean. Factual error repairs, on the other hand, show little differentiation of complete and incomplete reparanda in terms of lexical frequency.

Figure 3 about here

With reference to EXPECTATION 3 above, Kapatsinski's (2010) observation of an absence of association between lexical frequency and repair semantics is confirmed if factual and linguistic error repairs are grouped together, as in Kapatsinski's study: there is no significant relationship between Repair type and any of our probabilistic variables. However, separating factual and linguistic error repairs is informative. The significant association between Error type and Repair word frequency (and Reparandum word frequency, most bigram measures and Repair item predictability) reflects that linguistic error repairs involve more frequent and predictable lexical items than both appropriateness and factual error repairs (Tukey's HSD: $p=0.007$ and $p<0.001$, respectively). The latter two are not significantly different according to any probabilistic measure. The pattern can be gleaned from Figure 3 for Reparandum word frequency.

With reference to EXPECTATION 4, we find a significant relationship between Error type and Proportional completeness. At first sight, this seems consistent with Levelt's (1989) claim that error repairs are more likely than appropriateness repairs to involve a premature abandonment of the reparandum item. However, closer inspection suggests that again, linguistic and factual error repairs show different tendencies. Figure 4 shows, first of all, that appropriateness and factual error repairs have similar distributions for Proportional completeness. For both, about 70% of instances have incomplete reparanda, and for both, the means (marked by the peaks of the normal distribution curves) are below one. Appriateness repairs are associated with a higher mean than factual error repairs, as Levelt's claim predicts — but the difference does not reach significance in pairwise comparison (Tukey's HSD: $p=0.865$), and for both appropriateness and factual error repairs, premature abandonment of the reparandum item is the norm.

Linguistic errors, on the other hand, are least likely to be interrupted prior to repair: about 50% have completed reparandum items, and the mean across instances is above one —

significantly higher than the means of appropriateness repairs (p=0.027) and factual error repairs (p=0.007). Moreover, while for both appropriateness and factual error repairs, Proportional completeness values above two make up less than 10% of the distribution, over 20% of linguistic error repairs have a value in this range. Repairs with these values are initiated after the reparandum item has been completed and once, twice or more times the same number of segments has been articulated in subsequent lexical items — in other words, they are notably late.[10]

Figure 4 about here

Relationships Involving Control Variables

Finally before turning to Offset-to-repair duration, an important question is whether any of our predictor variables show a significant correlation with Reparandum rate — our crucial control variable in modelling Repair rate. Our analysis reveals only two relevant correlations, both weak: Reparandum rate is negatively correlated with Lexical segments delta (r=–0.220, p=0.001), such that lower reparandum rates are likely to have a positive delta value and higher reparandum rates are likely to have a negative delta value; and positively correlated with Proportional completeness (r=0.143, p=0.040). These effects are consistent with higher segment counts in the reparandum item giving rise to higher reparandum articulation rates. Given that articulation rate is calculated partly on the basis of segment counts, this is hardly surprising. Notably, there appear to be no systematic relationships between Reparandum rate on the one hand and Repair type, Error type, Completeness or any of our frequency and predictability measures on the other.

[10] The extreme values involve short reparandum items followed by several other lexical items prior to repair: see (3a) below for a clear example.

Modelling Results

We now turn to the results of our efforts to model Offset-to-repair duration and Repair rate using the predictor and control variables listed in Table 1. We will translate our three main hypotheses — repeated here for reference — into concrete expected data patterns for Offset-to-repair duration and Repair rate in turn.

- **HYPOTHESIS A —** Measures of offset timing (where 'early offset' is an interruption before the end of the reparandum item, and 'late offset' after its completion) are significant predictors of offset-to-repair duration and repair tempo.
- **HYPOTHESIS B —** A semantically-based classification of repairs (in which 'error' and 'appropriateness' repairs are distinguished) is a significant predictor of offset-to-repair duration and repair tempo.
- **HYPOTHESIS C —** Measures of lexical frequency and contextual predictability for reparandum and repair items are significant predictors of offset-to-repair duration and repair tempo.

Modelling Offset-to-Repair Duration

Elaborating on our main hypotheses above, and taking into consideration one additional finding reported in previous research, we can formulate the following concrete expectations for modelling Offset-to-repair duration:

- **EXPECTATION 5 —** Offset-to-repair duration is negatively correlated with the articulation rate of the reparandum.
- **EXPECTATION 6 —** 'Early offset' repairs have significantly lower offset-to-repair durations than 'late offset' ones.
- **EXPECTATION 7 —** Error repairs have significantly higher offset-to-repair durations than appropriateness repairs.
- **EXPECTATION 8 —** High-frequency repair words have significantly lower offset-to-repair durations than low-frequency ones.

EXPECTATION 5 is based on Oomen & Postma's (2001) finding of a reduction in average offset-to-repair duration when repairs are elicited under increased time pressure. This finding crucially informs our understanding of the temporal coordination between articulation and self-monitoring processes. EXPECTATION 6 elaborates on HYPOTHESIS A and is consistent with Nooteboom's (2010) finding for phonological error repairs. EXPECTATION 7 elaborates on HYPOTHESIS B. It is consistent with the findings reported by Kormos (2000: 157), which she accounts for by suggesting that the greater the change in 'informational content', the greater the 'processing effort' involved in the repair, and therefore the greater the likelihood of a delay in repair onset. On this reasoning, we might expect linguistic error repairs to have the lowest offset-to-repair durations, followed by appropriateness repairs and factual error repairs. EXPECTATION 8 elaborates on HYPOTHESIS C and is consistent with the notion that high-frequency words are accessed more quickly than low-frequency ones (Kapatsinski 2010, Harley & MacAndrew 2001).

We built linear mixed-effects regression models with Offset-to-repair duration as dependent variable. For this purpose, we excluded instances with an editing term between the reparandum and repair items, reducing the dataset to N=138. The distribution of the variable is not normal, as a sizeable subset of offset-to-repair intervals (N=20, or 14%) is zero. Excluding the raw zero durations (and log-transforming the remaining values) comes close to yielding a normal distribution (Shapiro-Wilks test: W=0.977, p=0.039). Therefore, we built two models: one to predict whether the interval duration is zero or not, and one to predict positive durations. In the latter case, the size of the dataset is further reduced to N=118.

In both cases, we started with a model containing only the random factor Speaker,[11] and assessed first whether any control variables significantly improved model fit. Neither Lexical segments delta nor Reparandum rate did: this means that EXPECTATION 5 above is not met. We then assessed for each of the candidate predictors listed in Table 1 above whether its addition to the model further improved its fit to the data. In both cases we expanded the model with the predictor causing the greatest significant improvement of the model's log likelihood, and then repeated the procedure, residualizing remaining predictors

[11] The models we report have random intercepts only; adding random slopes did not improve fit.

where relevant in light of the significant relationships pointed out above, and assessing whether incorporating interactions directly into the model improved fit. This yielded no further expansions. Table 3 summarizes the resulting models.

Table 3 about here

Table 3 shows that neither of our semantic variables features in the analysis: against EXPECTATION 7, we find no evidence for error and appropriateness repairs being characterized by different offset-to-repair durations. We do find some evidence to suggest that reparandum word completeness and probabilistic variables influence offset-to-repair duration. In the binary response model, the coefficient for Reparandum next bigram shows that an increase in bigram frequency is associated with an increase in the likelihood of a zero offset-to-repair duration. Subsequent conditional inference regression tree modelling suggests that the effect is due to a small subset of instances (14, or 10%) with very high bigram values having a relatively high incidence of zero offset duration: see Figure 5. Most of our bigram variables capture the same effect, but none of our unigram variables do, and neither does Repair item predictability. This is, then, a weak effect, and the fact that it is not reflected in the continuous model leads to our conclusion that EXPECTATION 8 is met, but not robustly so.

Figure 5 about here

In the continuous model, the coefficient for Proportional completeness shows that an increase in reparandum item completeness is associated with a decrease in offset-to-repair durations other than zero. This is inconsistent with Nooteboom's (2010) finding that 'early offset' repairs have significantly lower offset-to-repair durations than 'late offset' ones: EXPECTATION 6, then, is not met. Again, the observed effect is a weak one: it is not captured by Completeness, does not surface in a conditional inference regression tree analysis, and is not reflected in the binary response model. Moreover, further inspection suggests that the pattern is constrained by Error type: as Figure 6 shows, only linguistic errors show a negative correlation between Proportional completeness and Offset-to-repair duration ($r=-0.411$, $p=0.018$), and this is mostly due to the temporal characteristics of the small subset of instances with Proportional completeness values above two. Appropriateness and factual error repairs do not show significant correlations ($r=0.010$, $p=0.935$ and $r=0.155$, $p=0.157$, respectively).

Figure 6 about here

Modelling Repair Tempo

Elaborating on our main hypotheses above, and taking into consideration one additional finding reported in previous research, we can formulate the following concrete expectations for modelling Repair rate:

- **EXPECTATION 9 —** In a significant majority of instances, the articulation rate of the repair item is above that of the corresponding reparandum item.
- **EXPECTATION 10 —** 'Early offset' repairs have significantly higher repair item articulation rates than 'late offset' ones.
- **EXPECTATION 11 —** Error repairs have significantly lower repair item articulation rates than appropriateness repairs.
- **EXPECTATION 12 —** Highly predictable repair items have significantly higher articulation rates than less predictable ones.

EXPECTATION 9 is consistent with Plug's (2011) findings, and with the reasoning that on average, repair items are less informative than reparandum items. EXPECTATION 10 elaborates on HYPOTHESIS A and is based on the notion that 'early offset' repairs are produced under greater time pressure than 'late offset' ones (Nooteboom 2010). EXPECTATION 11 elaborates on HYPOTHESIS B and is consistent with Levelt's (1989) reasoning that error repairs are more informative than appropriateness repairs, and Levelt & Cutler's (1983) finding that the former are more frequently prosodically marked than the latter. EXPECTATION 12 elaborates on HYPOTHESIS C and is consistent with the frequent finding that highly predictable words are more prone to articulatory reduction than less predictable ones (Bybee 2002, Aylett & Turk 2004, Pluymaekers et al. 2005a, Seyfarth 2014).

In modelling Repair rate, we followed the same general procedure as that described above for Target-to-repair duration. In this case, we started with a model containing Speaker and Reparandum rate.[12] As might be expected, this reveals a significant correlation between reparandum and repair articulation rates; the model accounts for approximately 56% of the variance in Repair rate (r^2=0.557). The correlation is illustrated in Figure 7. Figure 7 further illustrates that consistent with Plug's (2011) results, we find that in a significant majority of instances (140, or 67%; χ^2=24.120, df=1, p<0.001 when compared with a 50%–50% split), the articulation rate of the repair stretch is above that of the corresponding reparandum. In other words, EXPECTATION 9 above is met.

Figure 7 about here

We checked whether Target-to-repair duration and Lexical segments delta are significant predictors of Repair rate, and found that the latter is, even after residualizing by Reparandum rate: the greater the value for Lexical segments delta — in effect, the longer the repair lexical item relative to the reparandum lexical item — the higher the articulation rate of the repair. Our 'base' model containing Speaker, Reparandum rate and Lexical segments delta accounts for approximately 62% of the variance in Repair rate (r^2=0.617). Prior to further modelling, we removed two outliers for Repair rate, reducing the size of the dataset to

[12] Exploratory analysis not reported here showed that whether Reparandum rate is measured including or excluding lexical items that are repeated in the repair has no significant effect on its predictive value in modelling Repair rate. This confirms that the presence or absence of lexical repetition is not a significant predictor of repairs' temporal organization (Plug 2011).

N=207.[13] We then applied the same stepwise algorithm as described for Offset-to-repair duration above. The resulting model is summarized in Table 4.

Table 4 about here

The model in Table 4 consists of our 'base' model plus Completeness only. While the model is significantly improved in terms of its fit to the data (log likelihood 15.343 to 18.018, χ^2=5.350, df=1, p=0.021), it still accounts for approximately 62% of the variance in Repair rate (r^2=0.619). The coefficient for Completeness shows that repairs with an incomplete reparandum item have a higher mean for Repair rate than repairs with a completed reparandum item. This is consistent with EXPECTATION 10. Still, the effect is too weak to emerge from a conditional inference regression tree analysis, or for a simple means comparison to reveal significance (F(1, 205)=0.048, p=0.827).[14] Figure 8 shows that it is really only observed among appropriateness repairs. Despite this apparent interaction between Completeness and Error type, adding the latter to the model — whether as an interaction term or as an additional main effect — does not improve fit. The same is the case for all probabilistic variables. As in the case of EXPECTATION 8 above, then, EXPECTATION 10 can be said to be met, but not robustly so. Clearly, neither EXPECTATION 11 nor EXPECTATION 12 finds support in our dataset.

Figure 8 about here

[13] These outliers gave rise to an apparent effect on Repair rate of speaker gender. They are highlighted in Figure 6, which shows their values are clearly separated from an otherwise continuous distribution of (log) articulation rate values between –3.5 and –5.0. The resulting distribution is not significantly different from normal (W= 0.990, p=0.143).

[14] The simple means comparison reported here was computed with the residuals of our 'base' model containing Speaker, Reparandum rate and Lexical segments delta as dependent variable. These were also used to construct Figure 8.

Discussion

In this paper we have reported on a phonetic analysis of instances of lexical self-repair, focusing on the repairs' temporal organization following the utterance interruption. The major aim of the analysis was to assess the impact on offset-to-repair duration and repair tempo of a number of factors that have been shown to constrain repair prosody specifically, or speech tempo more generally. Specifically, on the basis of previous research we addressed three general hypotheses, and developed these into a series of expected data patterns. We will return to our general hypotheses in our concluding section below; here, we take our concrete expectations as the starting point for discussion.

Relationships between Offset Timing, Repair Semantics, Frequency and Predictability

As regards the relationships among our independent variables, including possible interactions across our three predictor variable groups, our expectations were matched by our results as follows.

- **EXPECTATION 1 —** Unigram frequencies for reparandum and repair items are significantly correlated, and reparandum items more frequent than repair items. ➔ Partly met: significant correlation observed, but no difference in central tendency.
- **EXPECTATION 2 —** High-frequency reparandum items are more commonly completed prior to repair than low-frequency items. ➔ Met.
- **EXPECTATION 3 —** There is no significant relationship between repair semantics and reparandum item frequency. ➔ Not met: appropriateness and factual error repairs not significantly different, but linguistic error repairs have significantly more frequent reparandum items.
- **EXPECTATION 4 —** Error repairs more commonly involve a premature abandonment of the reparandum item than appropriateness repairs. ➔ Not met: appropriateness and factual error repairs not significantly different, and linguistic error repairs have significantly fewer incomplete reparandum items.

In relation to EXPECTATION 1, the fact that our dataset does not show the consistent differential between reparandum and repair item frequencies observed by Kapatsinski (2010) suggests that the lexical activation constraints he invokes to account for the differential are not strong enough to prevent routine activation of erroneous or inappropriate words that are less frequent than the words whose place they temporarily occupy. With reference to temporal organization, since neither a fall nor a rise in lexical frequency between reparandum and repair items is the norm, the frequency characteristics of our repairs do not support any gross generalization as to the relative articulation rates of the two stretches.

As regards EXPECTATIONS 2, 3 and 4, the results of our analysis of possible relationships between measures of offset timing, repair semantics, frequency and predictability highlight the importance of distinguishing factual and linguistic error repairs. For EXPECTATION 2, factual error repairs do not show the expected relationship between reparandum item frequency and completeness, while appropriateness and linguistic error repairs do. For EXPECTATIONS 3 and 4, we have found that linguistic error repairs are significantly different from both appropriateness and factual error repairs in terms of their frequency and offset timing characteristics. The observed differences between factual and linguistic error repairs are particularly notable, given that these have not been consistently distinguished in previous work on repair prosody, including Levelt & Cutler (1983), Kormos (2000) and Kapatsinski (2010).

On the whole, our results provide some support for Kapatsinski's (2010) account of the relationship between frequency of use and automaticity of production: we observe a gross positive correlation between reparandum item frequency and completeness, and as far as appropriateness and factual errors are concerned, this correlation cannot be attributed to a confounding effect of repair semantics. However, when linguistic error repairs are included, the correlation can be partly attributed to the fact that these repairs both involve the most highly frequent reparandum items, and are least likely to involve a premature abandonment. A closer look at the linguistic error repairs in our dataset suggests that the latter may be due to other factors than high lexical frequency alone. In particular, unlike factual error repairs, linguistic error repair recurrently involve the correction of a preposition, particle or pronoun: see the examples in (3).

(3) a. dat daar wat schot **op** ko- **in** komt ('that some progress is made on that', lit. 'that there some progress on co- in comes')

b. omdat het **met** zoveel afdelingen gi- **over** zoveel afdelingen ging ('because it concerned so many departments', lit. 'because it with so many departments wen- about so many departments went')

c. hij is daar gesneuveld en **haar** do- **zijn** dochter heeft … ('he died there and her dau- his daughter has …')

In addition to being high-frequency, these reparandum items also tend to be short, and in many cases, their crucial, initially erroneous collocation is not with a preceding lexical item, but with a following one. This following lexical item can be immediately adjacent, as in (3a), but it can also be separated from the reparandum item by multiple lexical items, as the separation of met 'with' and ging 'went' in (3b) illustrates. These characteristics may conspire to yield high proportional completeness values, independently of lexical frequency. Of course, our dataset contains relatively few linguistic error repairs (N=36, or 17%), so more research is needed to establish a comprehensive profile of this sub-type of error repair.

Modelling Offset-to-Repair Duration

In modelling the duration of the offset-to-repair interval, predicting whether this duration is zero or not as well as modelling positive durations, our expectations were matched by our results as follows.

- **EXPECTATION 5 —** Offset-to-repair duration is negatively correlated with the articulation rate of the repandum. ➔ Not met: no significant relationship observed.
- **EXPECTATION 6 —** 'Early offset' repairs have significantly lower offset-to-repair durations than 'late offset' ones. ➔ Not met: weak effect in the opposite direction.
- **EXPECTATION 7 —** Error repairs have significantly higher offset-to-repair durations than appropriateness repairs. ➔ Not met: no significant relationship observed.
- **EXPECTATION 8 —** High-frequency repair words have significantly lower offset-to-repair durations than low-frequency ones. ➔ Met, but weak effect only.

In relation to EXPECTATION 5, our analysis does not yield evidence to support Oomen & Postma's (2001) finding that offset-to-repair duration decreases with an increase in local speech tempo. It seems plausible that this is because the articulation rate of the reparandum item is not properly representative of the local speech tempo; articulation rate measurements taken over the entire pre-offset utterance may reveal a different pattern. In relation to EXPECTATIONS 6 to 8, given that EXPECTATION 4 above is not met in our data, and we find instead that linguistic error repairs are significantly more likely than factual error and appropriateness repairs to have a completed reparandum item, EXPECTATIONS 6 and 7 could plausible both be met. Given that linguistic error repairs also contribute the most frequent and predictable lexical items to our dataset, however, it would seem unlikely that EXPECTATIONS 6 to 8 could all be met. Consistent with this logic, the empirical evidence provides support for EXPECTATION 8 only. Because of the weakness of the observed effect we are reluctant to draw firm conclusions from this finding.

Our analysis also suggests some influence of offset timing on offset-to-repair duration, but not in the direction expected on the basis of Nooteboom's (2010) findings. Again, the observed effect is weak, and appears to be due to the characteristics of linguistic error repairs with notably late utterance interruptions, such as that in (3b) above. Again, further research is needed to establish a comprehensive profile of this subset of repairs. One possibility is that speakers do not necessarily treat repairs such as those in (3a) and (3b) as 'late' when the collocations they fix — as opposed to the individual words they replace — are not completed prior to the repair.

Modelling Repair Tempo

In modelling the articulation rate of the repair, our expectations were matched by our results as follows.

- **EXPECTATION 9 —** In a significant majority of instances, the articulation rate of the repair item is above that of the corresponding reparandum item. ➔ Met.
- **EXPECTATION 10 —** 'Early offset' repairs have significantly higher repair item articulation rates than 'late offset' ones. ➔ Met, but weak effect only.
- **EXPECTATION 11 —** Error repairs have significantly lower repair item articulation rates than appropriateness repairs. ➔ Not met: no significant relationship observed.
- **EXPECTATION 12 —** Highly predictable repair items have significantly higher articulation rates than less predictable ones. ➔ Not met: no significant relationship observed.

In relation to EXPECTATION 5, our analysis confirms Plug's (2011) finding of a predominance of temporal compression — that is, a local increase in speech tempo — following the reparandum offset. As indicated above, given that EXPECTATION 1 is not met, we cannot explain this predominance in general probabilistic terms: while a predominant fall in informativeness between reparandum and repair items is theoretically plausible, none of our probabilistic measures suggest that this is present in our dataset. No additional effects of lexical frequency, contextual predictability or a semantically-based interpretation of informativeness are observed either: neither EXPECTATION 11 nor EXPECTATION 12 finds support in our data. In relation to EXPECTATION 10, our analysis confirms that repair tempo is constrained to some degree by offset timing, and the direction of the effect is consistent with Nooteboom's (2010) account of the temporal organization of phonological error repair. However, the effect is again too weak for us to draw firm conclusions from them.

Limitations of This Study

Before we turn to conclusions, we should acknowledge that the data set for this study is rather small. Because of the strict selection procedure, it is a highly homogenous sample of repairs, unlike larger samples used in some previous studies (e.g. Nakatani & Hirschberg 1994). However, the substantial proportion of expected data patterns that are not observed in the data raises the question of whether they might be observed in a larger sample of self-repairs. Similarly, it is possible that cloze probabilities derived from a larger sample of fill-in-the-gap tasks might reveal data patterns in relation to informativeness that remain elusive in the current study. Clearly, further research is needed to address these questions; seen in this light, the current study can be taken as a methodological model for replication on a larger scale. Particular care should be taken in further studies to ensure a good balance in numbers between appropriateness, factual error and linguistic error repairs. In the initial design of the current study, distinctions among subtypes of error repair were not expected to be particularly relevant to the temporal organisation of repair. The results of this study suggest that they are, but the small size of the proportion of linguistic error repairs prevents us from drawing firm conclusions about their linguistic and phonetic characteristics.

Conclusions

In this paper we have reported on a study of lexical self-repair driven by the following general hypotheses:

- **HYPOTHESIS A —** Measures of offset timing (where 'early offset' is an interruption before the end of the reparandum item, and 'late offset' after its completion) are significant predictors of offset-to-repair duration and repair tempo.
- **HYPOTHESIS B —** A semantically-based classification of repairs (in which 'error' and 'appropriateness' repairs are distinguished) is a significant predictor of offset-to-repair duration and repair tempo.
- **HYPOTHESIS C —** Measures of lexical frequency and contextual predictability for reparandum and repair items are significant predictors of offset-to-repair duration and repair tempo.

Addressing these hypotheses has advanced our understanding of the general temporal organization of repair, as well as our understanding of factors influencing this organization. In relation to the former, our study confirms that in lexical repair, the repair component is routinely articulated at a higher tempo than the reparandum. This matches previous findings reported by Plug (2011) for a variety of repair types, and by Plug & Carter (2014) for phonological error repair. Interestingly, it also matches Goffman's (1981) informal description of the phonetics of self-repair. This means that there are now substantial empirical grounds to reject the intuition that self-repair should be accompanied by relative hyper-articulation, because of the speaker's presumed intention to mark the information conveyed in the repair as more important than that conveyed in the reparandum (Plug 2011).

In relation to factors influencing the temporal organization of repair, our findings complement those reported by Plug & Carter (2014) for spontaneous phonological error repairs in revealing some data patterns consistent with a tendency for repairs that are initiated early to be completed fast, and for repairs that are initiated late to be completed more slowly. As highlighted above, the data patterns are far from strong, and whether such a tendency can be attributed to a difference between 'early offset' and 'late offset' repairs in terms of the division of labour between inner and overt speech monitoring, as suggested by Nooteboom (2010), remains a matter of debate. The interaction between reparandum item completeness and lexical frequency, observed by Kapatsinski (2010) and confirmed by our analysis, complicates the issue: are high-frequency reparandum items typically completed because they are not detected as repairable in inner speech monitoring, or because once their articulation has started, it is difficult to stop — even if the 'stop' signal came early? If the latter is the case, this would mean that reparandum item completeness is an unreliable indicator of detection timing at best. This is also suggested by findings of repair initiation delay 'strategies' (Levelt 1989, Seyfeddinipur et al. 2008, Tydgat et al. 2011), which challenge the idea that speakers necessarily '[s]top the flow of speech immediately upon detecting trouble' (Levelt, 1989: 478). Still, on the face of it our data provide some further empirical support for Nooteboom's (2010) argument.

Similarly, our study reveals some evidence that informativeness constrains the temporal organization of lexical self-repair: it conditions offset timing and, to some extent, offset-to-repair duration. Further research is needed to establish whether the absence of more widespread influence is due to the small size of the current data set, or generalizable to larger repair samples. On the face of it, it is notable given our efforts to account for multiple dimensions of informativeness, given previous observations on prosodic differences between error and appropriateness repairs (Levelt & Cutler 1983), and given the clear predictions we could make for effects of frequency and predictability on repair tempo. Moreover, the absence of a consistent relationship between reparandum and repair items in terms of relative informativeness means that the observed predominance of temporal compression following the reparandum cannot be easily accounted for in probabilistic terms — at least not in our data. One conclusion to draw from our findings, then, is that while the influence of informativeness on speech production is extensive, it cannot be assumed to be ubiquitous. One possibility is that there are specific contexts in which the influence of informativeness is limited by other constraints. In the case of self-repair, Plug (2011: 296) has proposed that speakers may be orienting to a pragmatic constraint promoting fast repair completion, motivated by considerations of face-saving. Our results do not allow us to reject this proposal.

Our analysis highlights the complex relationship between the various measures of informativeness. For example, following Levelt's (1989) semantically-oriented reasoning, we should consider factual error repairs as more informative than appropriateness repairs because they implement a greater change to the ongoing utterance, while in probabilistic terms, appropriateness repairs can be considered more informative because they are likely to have less predictable repair items. Our data challenge the latter line of reasoning, in that no significant differences are observed between appropriateness and factual error repairs on any dimension of frequency or predictability. At the same time, they provide no empirical support for the former, as the two subtypes of repair show little difference in temporal organization.

Finally, our analysis yields two important recommendations for future research into repair and the influence of informativeness on articulation. First, as already mentioned in the context of this study's limitations, factual and linguistic error repairs should be carefully distinguished in research on repair, as they are demonstrably different on multiple linguistic and probabilistic counts. Plug & Carter (2013) have already shown that the distinction has explanatory value in investigating the pitch and intensity characteristics of lexical self-repair, and it would seem highly relevant in studies of repair patterns in the speech of second language learners (Van Hest 1996, Kormos 1999, 2000), too. Second, unigram frequency counts should be treated as rough measures of contextual predictability at best, and accompanied by more context-sensitive measures, such as cloze probabilities or at least bigram frequency counts, where possible.

References

Aylett, M. & A. Turk 2004. The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. Language and Speech 47: 31–56.

Baayen, R.H. 2008. Analyzing linguistic data: A practical introduction to statistics using R. Cambridge: Cambridge University Press.

Baayen, R.H., R. Piepenbrock & L. Gulikers 1995. The CELEX lexical database. Release 2 [CD-ROM]. Philadelphia: Linguistics Data Consortium.

Barry, W. & B. Andreeva 2001. Cross-language similarities and differences in spontaneous speech patterns. Journal of the International Phonetic Association 31: 51–66.

Bates, D., M. Maechler, B. Bolker & S. Walker 2014. Lme4: Linear mixed-effects models using Eigen and S4. Version 1.1-7. http://CRAN.R-project.org/package=lme4.

Beckman, M.E. & J. Edwards 1987. The phonological domains of final lengthening. Journal of the Acoustical Society of America 81: S67–S67.

Benkenstein, R. & A.P. Simpson 2003. Phonetic correlates of self-repair involving repetition in German spontaneous speech. In Proceedings of Disfluency in Spontaneous Speech 2003. Gothenburg. 81–84.

Blackmer, E.R. & J.L. Mitton 1991. Theories of monitoring and the timing of repairs in spontaneous speech. Cognition 39: 173–194.

Boersma, P. & D. Weenink (2010). Praat: Doing phonetics by computer. Version 5.1.34. http://www.praat.org/.

Burdin, R.S. & C.G. Clopper 2015. Phonetic reduction, vowel duration, and prosodic structure. In Proceedings of the 18th International Congress of Phonetic Sciences. http://www.icphs2015.info/Proceedings.aspx.

Bybee, J. 2002. Word frequency and context of use in the lexical diffusion of phonemically conditioned sound change. Language Variation and Change 14: 261–290.

Bybee, J. 2010. Language, usage and cognition. Cambridge: Cambridge University Press.

Cambier-Langeveld, T., M. Nespor & V.J. van Heuven 1997. The domain of final lengthening in production and perception in Dutch. In Proceedings of Eurospeech 1997. Rhodes. 931–934.

Cutler, A. 1983. Speakers' conceptions of the function of prosody. In Prosody: Models and Measurements, ed. A. Cutler & D.R. Ladd. Heidelberg: Springer. 79–91.

Fox Tree, J.E. & H.H. Clark 1997. Pronouncing "the" as "thee" to signal problems in speaking. Cognition 62: 151–167.

Goffman, E. 1981. Radio talk. In Forms of talk, ed. E. Goffman. Oxford: Blackwell. 197–330.

Harley, T.A. & S.B.G. MacAndrew 2001. Constraints upon word substitution speech errors. Journal of Psycholinguistic Research 30: 395–418.

Hartsuiker, R.J. & H.H.J. Kolk 2001. Error monitoring in speech production: A computational test of the Perceptual Loop Theory. Cognitive Psychology 42: 113–157.

Hartsuiker, R.J., R. Bastiaanse, A. Postma & F. Wijnen 2005a. Phonological encoding and monitoring in normal and pathological speech. In Phonological encoding and monitoring in normal and pathological speech, ed. R.J. Hartsuiker, R. Bastiaanse, A. Postma & F. Wijnen. Hove: Psychology Press. 1–14.

Hartsuiker, R.J., H.H.J. Kolk & H. Martensen 2005b. Division of labor between internal and external speech monitoring. In Phonological encoding and monitoring in normal and pathological speech, ed. R.J. Hartsuiker, R. Bastiaanse, A. Postma & F. Wijnen. Hove: Psychology Press. 187–205.

Hartsuiker, R.J., C.M. Catchpole, N.H. De Jong & M.J. Pickering 2008. Concurrent processing of words and their replacements during speech. Cognition 108: 601–607.

Hothorn, T., K. Hornik & A. Zeileis 2006. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical Statistics 15: 651–674.

Jacewicz, E., R.A. Fox & L. Wei 2000. Between-speaker and within-speaker variation in speech tempo of American English. Journal of the Acoustical Society of America 128: 839–850.

Jasperson, R. 2002. Some linguistic aspects of closure cut-off. In The language of turn and sequence, ed. C.E. Ford, B.A. Fox & S.A. Thompson. Oxford: Oxford University Press. 257–286.

Jurafsky, D., A. Bell, M. Gregory & W.D. Raymond 2001. Probabilistic relationships between words: Evidence from reduction in lexical production. In Frequency and the emergence of linguistic structure, ed. J. Bybee & P. Hopper. Amsterdam: John Benjamins. 229–254.

Kapatsinski, V. 2010. Frequency of use leads to automaticity of production: Evidence from repair in conversation. Language and Speech 53: 71–105.

Kohler, K.J. 1990. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In Speech production and speech modelling, ed. W.J. Hardcastle & A. Marchal. Dordrecht: Kluwer. 69–92.

Kormos, J. 1999. Monitoring and self-repair in L2. Language Learning 49: 303–342.

Kormos, J. 2000. The timing of self-repairs in second language speech production. Studies in Second Language Acquisition 22: 145–167.

Levelt, W.J.M. & A. Cutler 1983. Prosodic marking in speech repair. Journal of Semantics 2: 205–217.

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, Mass.: The MIT Press.

Levelt, W.J.M., A. Roelofs & A.S. Meyer 1999. A theory of lexical access in speech production. Behavioral and Brain Sciences 22: 1–75.

Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H Theory. In Speech production and speech modelling, ed. W.J. Hardcastle & A. Marchal. Dordrecht: Kluwer. 403–440.

Lindblom, B. 1996. Role of articulation in speech perception: Clues from production. Journal of the Acoustical Society of America 99: 1683–1692.

Marslen-Wilson, W. & L. Tyler 1981. Central processes in speech understanding. Philosophical Transactions of the Royal Society London, B259: 317–332.

Miellet, S., Sparrow, L., Sereno, S.C. 2007. Word frequency and predictability effects in reading French. Psychonomic Bulletin & Review 14: 762–769.

Nakatani, C.H. & J. Hirschberg 1994. A corpus-based study of repair cues in spontaneous speech. Journal of the Acoustical Society of America 95: 1603–1616.

Niebuhr, O. 2010. On the phonetics of intensifying emphasis in German. Phonetica 67: 170–198.

Nooteboom, S.G. 1972. Production and perception of vowel duration: A study of durational properties of vowels in Dutch. PhD thesis, Rijksuniversiteit Utrecht.

Nooteboom, S.G. 2005a. Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. Speech Communication 47: 43–58.

Nooteboom, S.G. 2005b. Listening to oneself: Monitoring speech production. In Phonological encoding and monitoring in normal and pathological speech, ed. R.J. Hartsuiker, R. Bastiaanse, A. Postma & F. Wijnen. Hove: Psychology Press. 167–186.

Nooteboom, S. 2010. Monitoring for speech errors has different functions in inner and overt speech. In The linguistic enterprise: From knowledge of language to knowledge in linguistics, ed. M.B.H. Everaert, T. Lentz, H. De Mulder, Ø. Nilsen & A. Zondervan. Amsterdam: John Benjamins. 213–233.

Oomen, C.C.E. & A. Postma 2001. Effects of time pressure on mechanisms of speech production and self-monitoring. Journal of Psycholinguistic Research 30: 163–184.

Oomen, C.C.E. & A. Postma 2002. Limitations in processing resources and speech monitoring. Language and Cognitive Processes 17: 163–184.

Oostdijk, N. 2002. The design of the Spoken Dutch Corpus. In New frontiers of corpus research, ed. P. Peters, P. Collins & A. Smith. Amsterdam: Rodopi. 105–113.

Plug, L. 2011. Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. Journal of Phonetics 39: 289–297.

Plug, L. & P. Carter 2013. Prosodic marking, pitch and intensity in spontaneous lexical self-repair in Dutch. Phonetica 70: 155–181.

Plug, L. & P. Carter 2014. Timing and tempo in spontaneous phonological error repair. Journal of Phonetics 45: 52–63.

Pluymaekers, M., M. Ernestus & R.H. Baayen 2005a. Articulatory planning is continuous and sensitive to informational redundancy. Phonetica 62: 146–159.

Pluymaekers, M., M. Ernestus & R.H. Baayen 2005b. Lexical frequency and acoustic reduction in spoken Dutch. Journal of the Acoustical Society of America 118: 2561–2569.

Postma, A. 2000. Detection of errors during speech production: A review of speech monitoring models. Cognition 77: 97–131.

Quené, H. 2008. Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. Journal of the Acoustical Society of America 123: 1104–1113.

R Development Core Team 2008. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Rietveld, A.C.M. & V.J. van Heuven 1997. Algemene fonetiek. Bussum: Coutinho.

Schotter, E.R., Bicknell, K., Howard, I. et al. 2014. Task effects reveal cognitive flexibility responding to frequency and predictability. Cognition 131: 1–27.

Seyfarth, S. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. Cognition 133: 140–155.

Seyfeddinipur, M., S. Kita & P. Indefrey 2008. How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. Cognition 108: 837–842.

Shattuck-Hufnagel, S. & A. Cutler 1999. The prosody of speech error corrections revisited. Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco.

Shriberg, E. 2001. To 'errr' is human: Ecology and acoustics of speech disfluencies. Journal of the International Phonetic Association 31: 153–169.

Slevc, L.R. & V.S. Ferreira 2006. Halting in single word production: A test of the perceptual loop theory of speech monitoring. Journal of Memory and Language 54: 515–540.

Smiljanić, R. & A.R. Bradlow 2009. Speaking and hearing clearly: Talker and listener factors in speaking style changes. Language and Linguistics Compass 3: 236–264.

Strycharczuk, P., M. van 't Veer, M. Bruil & K. Linke 2014. Phonetic evidence on phonology-morphosyntax interactions: Sibilant voicing in Quito Spanish. Journal of Linguistics 50: 403–452.

Tagliamonte, S. 2014. A comparative sociolinguistic analysis of the dative alternation. In Linguistic variation: Confronting fact and theory, ed. R. Torres Cacoullos, N. Dion & A. Lapierre, London: Routledge.

Tagliamonte, S. & R.H. Baayen (2012). Models, forests and trees of York English: Was/were variation as a case study for statistical practice. Language Variation and Change 24: 135–178.

Tydgat, I., M. Stevens, R.J. Hartsuiker & M.J. Pickering 2011. Deciding where to stop speaking. Journal of Memory and Language 64: 359–380.

Van Hest, E. 1996. Self-repair in L1 and L2 production. Tilburg: Tilburg University Press.

Tables

(a)	Offset-to-repair duration, Repair rate
(b)	OFFSET TIMING Completeness, Proportional completeness
	REPAIR SEMANTICS Repair type, Error type
	FREQUENCY AND PREDICTABILITY (a) Reparandum word frequency, Repair word frequency, Word frequency delta (b) Reparandum prior bigram, Repair prior bigram, Prior bigram delta (c) Reparandum next bigram, Repair next bigram, Next bigram delta (d) Repair item predictability
(c)	Speaker, Reparandum rate, Lexical segments delta

Table 1. Variables entered into the analysis: (a) dependent variables, (b) predictor variable groups, (c) control variables.

Relationship	Test	Coefficient	p
OFFSET TIMING ~ FREQUENCY AND PREDICTABILITY			
Proportional completeness ~ Reparandum word frequency	Pearson's correlation	r=0.410	<0.001
REPAIR SEMANTICS ~ OFFSET TIMING			
Error type ~ Proportional completeness	ANOVA	F(2, 206)=4.684	0.010
REPAIR SEMANTICS ~ FREQUENCY AND PREDICTABILITY			
Error type ~ Repair word frequency	ANOVA	F(2, 206)=13.791	<0.001

Table 2. Main significant relationships across the three predictor variable groups.

	Factor	Estimate	Df	Sum sq	F
(a)	Intercept	1.858			
	Reparandum next bigram	−0.082	1	0.952	8.020
(b)	Intercept	4.662			
	Proportional completeness	−0.171	1	2.714	5.356

Table 3. Summary of fixed effects in linear mixed-effects models predicting (a) whether the offset-to-repair interval is zero or not, and (b) the log-transformed offset-to-repair durations excluding raw zero values.

Factor	Estimate	Df	Sum sq	F
Intercept	−3.124			
Reparandum rate	0.257	1	0.756	20.088
Lexical segments delta	0.147	1	0.414	11.003
Completeness (Incomplete)	0.075	1	0.200	5.317

Table 4. Summary of fixed effects in a linear mixed-effects model predicting Repair rate. Completeness is given with the level to which the estimate refers.

Figures

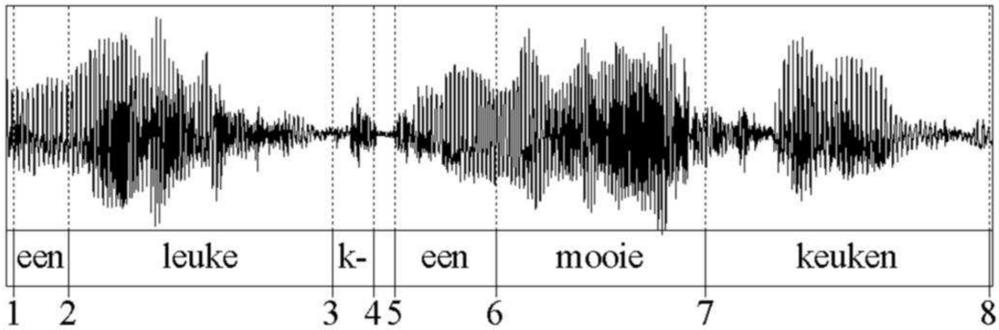

Figure 1. Segmented waveform of the repair in (1e). ‘1’ and ‘4’ delimit the reparandum: ‘4’ constitutes the reparandum offset. ‘4’ and ‘5’ delimit the offset-to-repair interval; ‘5’ to ‘8’ the repair. ‘2’ and ‘3’ delimit the reparandum item, leuke ‘nice’; 6 and 7 the repair item, mooie ‘beautiful’.

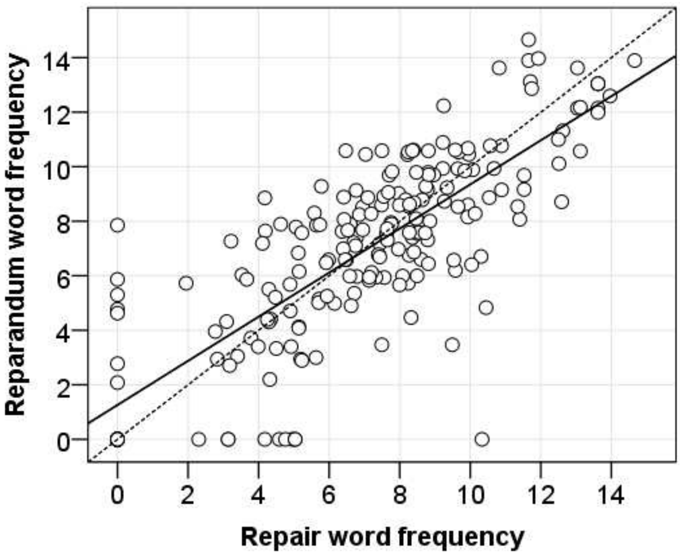

Figure 2. Reparandum word frequency plotted against Repair word frequency. The diagonal dotted line indicates values where the two rates are identical. The more shallow, solid line is the outcome of a simple linear regression.

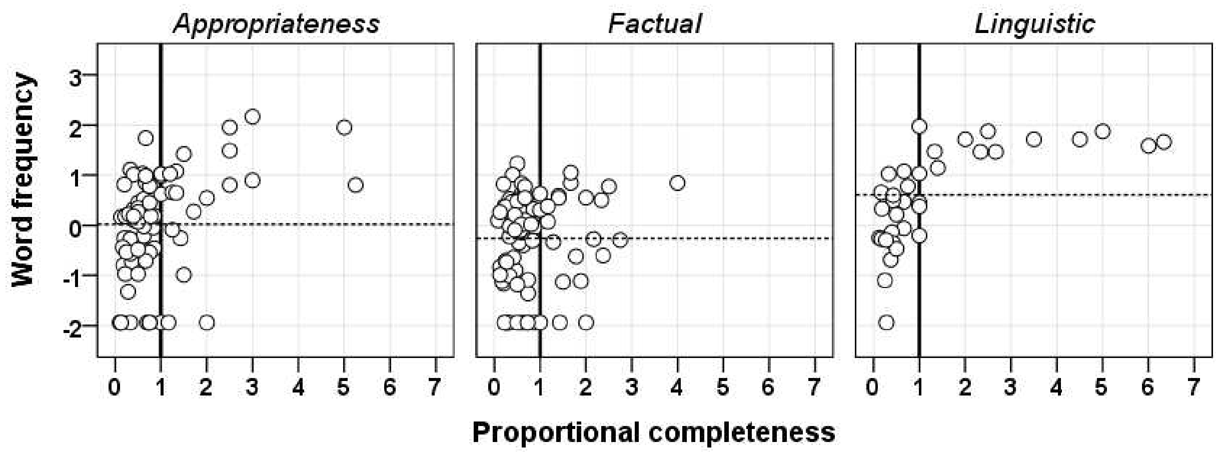

Figure 3. Reparandum word frequency plotted against Proportional completeness, split by Error type. The dotted horizontal lines mark the Reparandum word frequency means. The solid vertical lines represent the boundary between the two levels of Completeness.

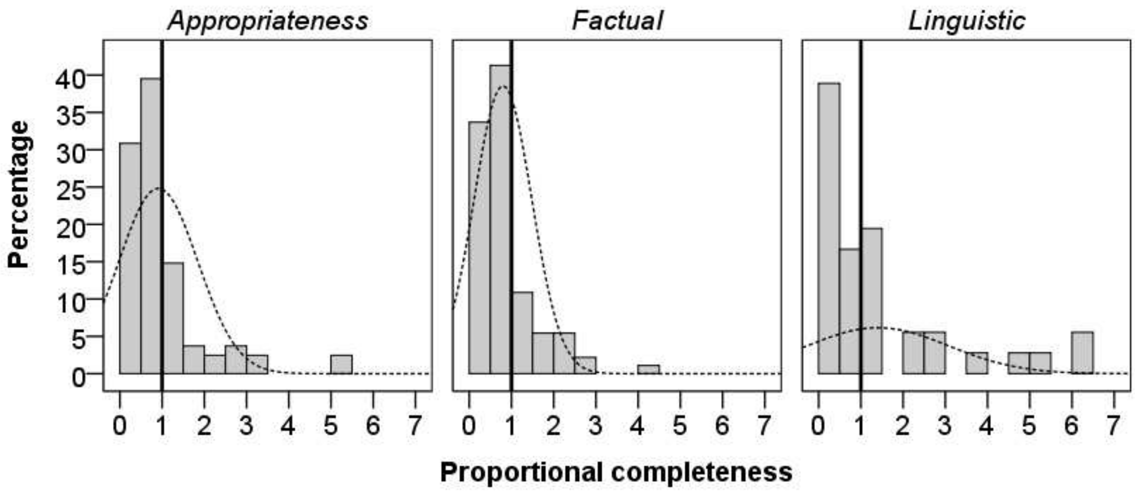

Figure 4. Percentage histograms for Proportional completeness, split by Error type. The dotted lines are fitted normal distribution curves. The solid vertical lines represent the boundary between the two levels of Completeness.

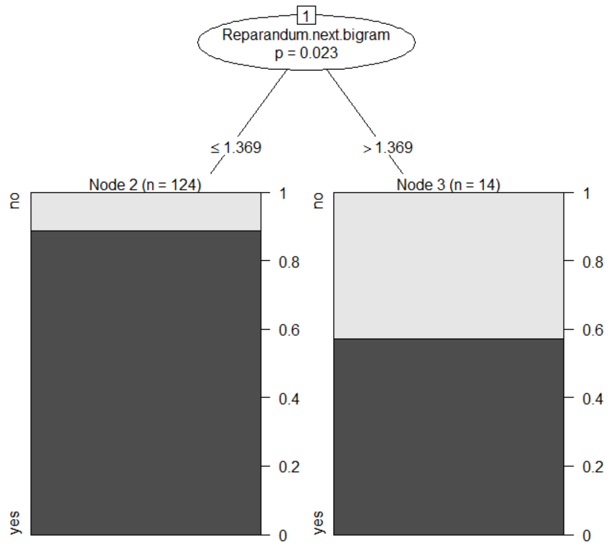

Figure 5. Conditional inference regression tree for Offset-to-repair duration, with split for Reparandum next bigram.

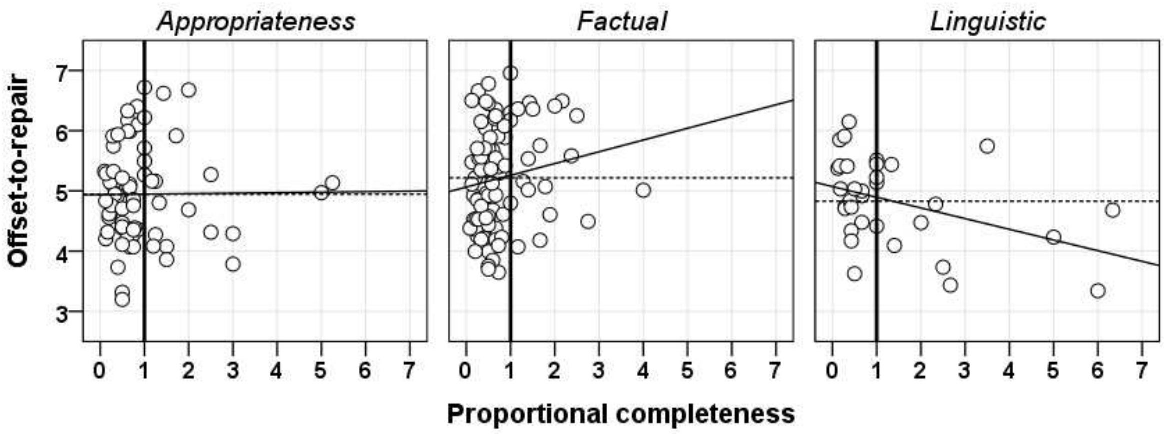

Figure 6. Offset-to-repair duration plotted against Proportional completeness, split by Error type. The dotted horizontal lines mark the Offset-to-repair duration means. The solid vertical lines represent the boundary between the two levels of Completeness. The solid slopes are the outcomes of simple linear regressions.

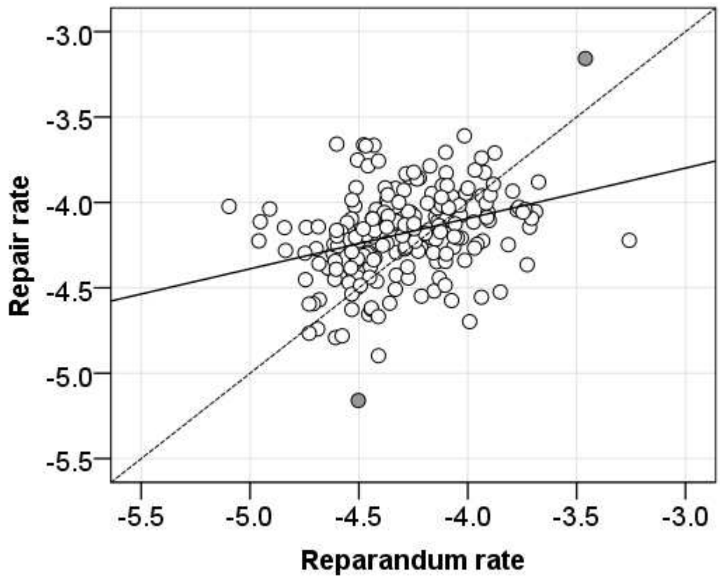

Figure 7. Repair rate plotted against Reparandum rate. The diagonal dotted line indicates values where the two rates are identical. The more shallow, solid line is the outcome of a simple linear regression ignoring all other factors. The two filled data points are outliers that were removed in subsequent analysis (see text).

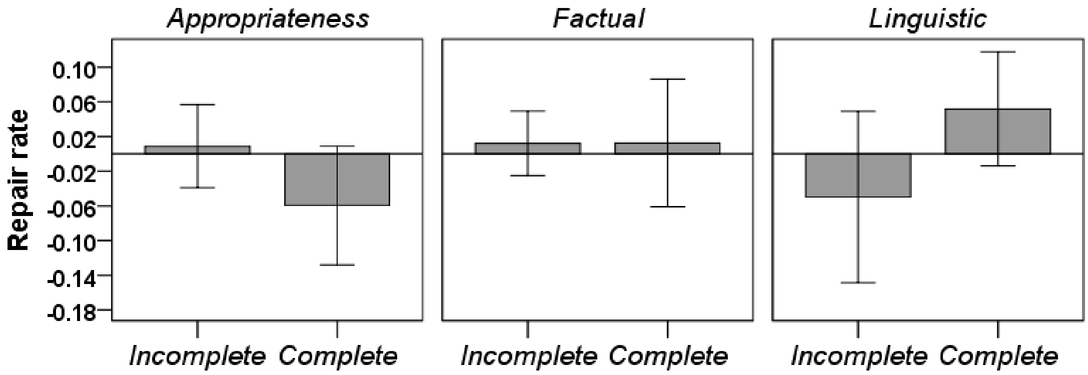

Figure 8. Repair rate means by Completeness, split by Error type. Plotted values for Repair rate are the residuals of a linear mixed-effects model predicting Repair rate containing Speaker, Reparandum rate and (residualized) Lexical segments delta. Error bars represent 95% confidence intervals.