TEMPORAL ORGANIZATION IN LEXICAL SELF-REPAIR IN COLOMBIAN SPANISH UNSCRIPTED SPEECH

Kelly Johanna Vera Diettes[1] | Leendert Plug[2]

[1]Universidad Nacional de Colombia | [2]University of Leeds
kjverad@unal.edu.co | l.plug@leeds.ac.uk

ABSTRACT

We report on an investigation of the temporal organization of lexical self-repair, focusing on the timing of the speech cut-off and repair onset. [1] showed that speech error repairs in which speech is interrupted early are completed quickly, while later repairs are generally slower. This supports the idea that early repairs are initiated through inner speech monitoring, and late repairs through overt speech monitoring. Research on lexical repair [2] suggested that different subtypes of repair are subject to different temporal constraints. Moreover, [3] proposed that speakers may strategically delay repair onset. This study investigates error-to-cut-off and cut-off-to-repair interval durations in lexical repairs sampled from a corpus of Colombian Spanish conversation. It assesses the correlation between interval durations and the predictive power of semantic variables. The results provide support for the findings of [1] but not entirely for [2]; also, they support the notion that speakers may delay interruption and repair.

Keywords: temporal organization, timing, self-repair, Spanish, unscripted speech.

1. INTRODUCTION

This paper reports on an investigation of the temporal organization of lexical self-repair in unscripted Colombian Spanish. In lexical self-repair, a speaker rejects one lexical choice for another, as in '*a bee w- (...) a butterfly was*'. We focus on the time it takes to interrupt speech after the start of the problematic 'target' word attempt (*target-to-cutoff*: '*bee w-*' in the example) and the time between the interruption and the start of the repair (*cutoff-to-repair*: '*(...)*').

[1] and [4] observed that target-to-cutoff durations in elicited Dutch speech error repairs are bimodally distributed, with two peaks of dispersion separated by around 500 ms. This is consistent with early repairs following error detection in 'inner speech' monitoring, while late repairs follow error detection in overt speech monitoring. This distinction has an impact on cutoff-to-repair timing too, such that repairs that follow error detection in overt speech monitoring require more planning time [5]. However, it has also been suggested that speakers may delay cut-off, possibly to allow themselves to plan the repair while prioritising fluency [3, 5]. This might generate repairs with long target-to-cutoff but short cut-off-to-repair intervals.

As [4] point out, there is no *a priori* reason to expect that findings on the timing of speech error repairs will generalize to lexical repairs, as different processing mechanisms are involved. [2] reports that in Dutch spontaneous lexical self-repairs, repairs in which the repairable word is completed before interruption have lower cutoff-to-repair durations than repairs with early interruptions. [2] further suggests that semantic sub-types of lexical repair may be subject to different temporal constraints. In particular, [2] shows that classifying repairs as error or appropriateness repairs, and further dividing error repairs into factual and linguistic error repairs, is informative. [2] excludes repairs with 'editing terms' [6], such as 'I mean' in '*a bee w- I mean a butterfly was*'. These complicate the analysis of the temporal organization of lexical self-repair: different subtypes of repair are likely to differ with respect to the occurrence of such terms in the cutoff-to-repair interval, and it seems reasonable to expect that cutoff-to-repair intervals containing an editing term are on average longer than silent ones.

To further our understanding of the temporal organization of lexical self-repair, we explore the distributions of target-to-cutoff and cutoff-to-repair intervals in a sample of self-repairs from a corpus of unscripted Colombian Spanish speech. We looked for indications of bimodality following [4, p. 20] and considered the relationship between these two intervals, taking account of the existence of multiple semantic sub-types of repair as well as the presence *vs* absence of editing terms.

2. DATA AND METHODS

2.1. Corpus

The data for this paper comprise 97 instances of self-initiated lexical self-repair extracted from sub-corpora of the Unscripted Colombian Spanish Interactive (UCSI) Corpus. The UCSI project aims to collect unscripted speech from the different dialects and subdialects of Colombian Spanish, based on the dialectal classification by [6]. Conversational speech is collected by means of two interactive tasks: A South American adaptation of the *Diapix task* [8, 9] and a consensual response task in which speakers need to verbally agree on answers to questions before reporting these in writing. Repairs were sampled from recordings of both tasks as completed by 7 speaker pairs, from 3 subdialects of Colombian Spanish. Initially 104 repairs were identified; five were later excluded (see below).

2.2. Segmentation

Each repair was orthographically transcribed by the first researcher, a native speaker of Colombian Spanish, then semi-automatically transcribed in Praat [10] using the Easyalign Spanish extension [11]. Resulting phoneme, word, syllable and phone tiers were hand-corrected by the first author. A point tier was added to locate crucial temporal landmarks, as illustrated in Figure 1. Here the ‘target’ of the repair is *cuatro* ‘four’, which is corrected to *dos* ‘two’. The speech is cut off in the speaker’s first production of *zapatos* ‘shoes’. The interval between 1 and 3 is the target-to-cutoff interval (‘T_C’); the interval between 3 and 4 is the cutoff-to-repair interval (‘C_R’). Interval durations were extracted and log-transformed prior to modelling, although the figures below will plot durations in ms.

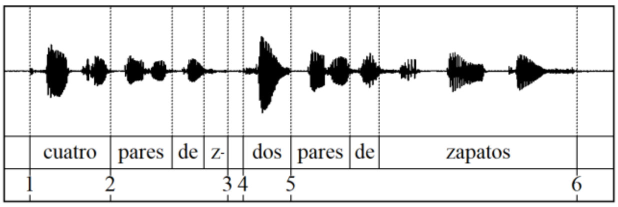

Figure 1: Segmented waveform of a repair. ‘1’ and ‘2’ delimit the target word; ‘3’ is the cutoff; ‘4’ and ‘5’ delimit the repair word; ‘6’ marks the end of the repair.

2.3. Semantic classification

Each instance was classified as an appropriateness or error repair, using the criteria described in [2]. Repairs with factual inaccuracies or linguistic ill-formedness are error repairs; in other cases, we assume that the initial lexical choice is rejected for pragmatic felicity or ‘appropriateness’ reasons. The classification was done by the first author and an independent second rater, a Colombian Spanish discourse analyst. Classifications matched for 92 repairs (90%). Consensus was reached for 9 out of the remaining 10; one repair was excluded as the raters agreed it could not be reliably classified.

2.4. Editing terms

We coded for the presence of editing terms following [6]. Examples include *o sea* ‘I mean’ and filled pauses such as *eee* [eː] and *mmm* [mː]. Some Colombian Spanish expressions could not be easily mapped to the classification scheme in [6]; as a result, 4 further repairs were excluded. This left 97 instances of repair, of which 26 contain editing terms.

2.5. Statistical analysis

We first inspected the distributions of the durations of T_C and C_R. We then performed a cluster analysis to investigate the relationship between the two intervals in more detail. Subsequently, we modelled the interval durations using linear mixed-effect regression models and cluster membership using a conditional inference regression tree. All the analyses were performed in R [12] using *lme4* [13], *k-means* and *party* [14]. Table 1 shows the variables entered into the analysis. Chi square tests confirmed that none of the independent variables are collinear, although we can note that none of the linguistic error repairs in our data set contain editing terms.

Dependent variables	**T_C, C_R**
Independent variables	**Completeness (Com)** 1. *Complete* 2. *Incomplete*
	Repair Type (ReT) 1. *Appropriateness* 2. *Error* 2.1 *Linguistic* 2.2 *Factual*
	Editing Terms (EdT) 1. *Yes* 2. *No*
Random effect	**Speaker**

Table 1: Variables entered into the analysis.

3. RESULTS

3.1. T_C and C_R duration distributions

The mean duration of T_C is 515 ms (range 99–2272); the mean duration of C_R is 307 ms (range 0.44–2982). Hartigans’ dip tests confirm that neither distribution is significantly multimodal (T_C: D=0.03, p=0.921; C_R, D=0.03, p=0.876), and the interval durations are not significantly correlated with each other (Pearson’s r=0.15, p=0.13).

We investigated the relationship between the two intervals in more detail through *k-means* clustering. This revealed three clusters, illustrated in Figure 2. First, there are repairs with a relatively short T_C and a relatively short C_R. We call these 'Early-Early' repairs: they have an early interruption and an early repair onset. Second, there are repairs with a relatively short T_C but a relatively long C_R. We call these 'Early-Late' repairs. Third, there are repairs with a relatively long T_C but a relatively short C_R. We call these 'Late-Early repairs'. Figure 2 shows that only one repair in our data set might potentially qualify for a 'Late-Late' classification.

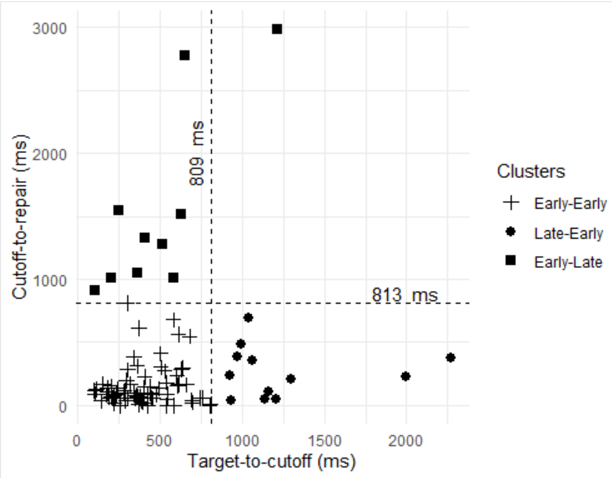

Figure 2: Scattergram showing T_C and C_R durations and cluster membership, with threshold lines separating the clusters.

3.2. Relevance of editing terms and repair semantics

As expected, there is a systematic relationship between the presence or absence of editing terms and C_R interval duration. Interestingly, EdT also seems systematically related to T_C duration. Figure 3 shows that repairs without any editing terms predominantly fall in the Early-Early cluster, while repair with an editing term are evenly distributed across the three clusters.

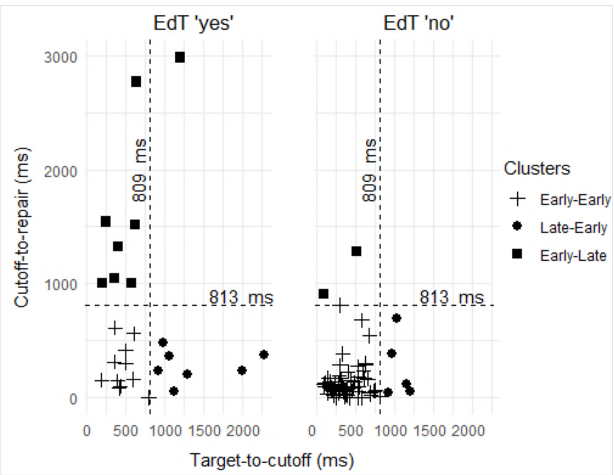

Figure 3: Scattergram as in Figure 2, split by EdT (left *Yes*, right *No*).

This was confirmed by our models. We first built a linear mixed-effects regression model for each of T_C and C_R with EdT (*Yes*, *No*) as well as ReT (*Appropriateness*, *Error*) and Com (*Complete*, *Incomplete*) as predictors. For each model we started with a random intercept for speaker identity and then added the predictors in turn, only keeping an additional predictor if model comparison indicated model fit was significantly improved. For each model, we tested for the relevance of an interaction between EdT and ReT, and we assessed whether implementing ReT with three levels (*Appropriateness*, *Factual*, *Linguistic*), added predictive value. The latter was not the case.

Optimal models are given in Table 2. The model for T_C confirms that EdT has a significant effect (higher duration with editing terms); so does Com (higher duration for repairs with completed target words). The model shows no significant overall effect of ReT, but reveals a significant interaction between ReT and EdT. The interaction is illustrated in Figure 4. It suggests that the effect of EdT on T_C duration is predominantly observed among appropriateness repairs.

The model for C_R shows a significant effect of EdT (higher duration with editing terms) only.

Target-to-Cutoff	**Estimate**	**Df**	**t**	**p**
Intercept	487.3	97	7.57	<0.01
EdT *Yes*	457.6	97	4.11	<0.01
Com *Incomplete*	-158.8	97	-2.31	0.02
ReT *Error*	27.5	97	0.35	0.72
EdT *Yes**ReT *Error*	301.2	97	-2.01	0.04
Cutoff-to-repair	**Estimate**	**Df**	**t**	**p**
Intercept	166.2	97	3.15	<0.01
EdT *Yes*	525.84	97	5.16	<0.01

Table 2. Optimal linear mixed effects models for T_C and C_R.

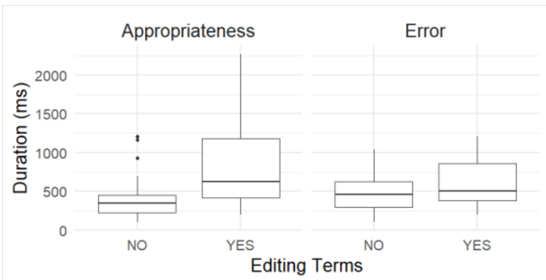

Figure 4. Box plot illustrating the significant interaction between EdT and ReT for T_C duration.

As a final step, we fitted a conditional inference regression tree with cluster membership (as in Figure 2) as response variable. This confirmed that the relationship between cluster membership and EdT seen in Figure 3 is at the level of significance ($p<0.001$). Neither Com nor ReT gave rise to significant splits in the tree.

4. DISCUSSION

This investigation assessed the temporal organization of lexical self-repair in Colombian Spanish unscripted speech, focusing on the target-to-cutoff and cutoff-to-repair intervals. In our data set, both have wide duration distributions. When considered separately, neither appears significantly bimodal; however, when considered together, the emergence of three clusters suggests that dividing both distributions into 'early' and 'late' ranges is informative. As expected, repairs with editing terms have longer cutoff-to-repair intervals than those without; interestingly, they also have longer target-to-cutoff intervals, at least among appropriateness repairs. Also as expected, repairs in which the target word is completed before the repair are associated with longer target-to-cutoff durations than repairs in which the target word is interrupted.

As indicated above, [1] and [4] observed that target-to-cutoff durations in elicited Dutch speech error repairs are bimodally distributed, with two peaks of dispersion separated by around 500 ms; moreover, speech error repairs in which speech is interrupted early are completed quickly, while later repairs are generally slower. In our spontaneous lexical repairs, we can identify 'early' and 'late' repairs, but unlike [2], we do not find that late repairs in terms of target-to-cutoff or target word completeness are also late in terms of cutoff-to-repair; and early repairs in terms of target-to-cutoff are not necessarily early in terms of cutoff-to-repair either.

The absence of a Late-Late cluster and the presence of a Late-Early cluster fit with the idea that speakers can delay interruption to maximize fluency [3, 5]. However, the additional existence of an Early-Late cluster suggests that speakers do not invariably use this delaying strategy. Further work is needed to expand our data set, so that we assess the robustness of the Late-Early and Early-Late temporal organizations, as well as the absence of a commonly observable Late-Late one.

Our analysis confirms that the presence or absence of editing terms is highly consequential for the temporal organization of repair. It might seem unsurprising that cutoff-to-repair intervals which contain editing terms are on average longer than silent cutoff-to-repair intervals, simply because producing an editing term takes time. [3] suggests that the use of editing terms might actually be an alternative 'repair delay' strategy to delaying the cutoff: after an early interruption, speakers can maintain fluency by using editing terms while planning the repair. In fact, Figure 3 shows that this accounts for most instances in our Early-Late cluster, as most contain one or more editing terms.

We also observed a positive effect of the presence of editing terms on target-to-cutoff duration, modulated by repair type. Concretely, it appears that our Late-Early cluster contains a majority of instances with an editing term (which must be relatively short), and these are mostly appropriateness repairs. If this pattern generalizes to a larger data set, more qualitative analysis is warranted to identify the specific subtypes of repair that have this temporal organization.

Unlike [2], we did not observe any more substantive effects of semantic repair type on interval durations, including cluster membership. [2] also suggests that the distinction between linguistic and factual error repairs is informative for understanding repair timing. Our results do not confirm this in quantitative terms, but our observation that none of our linguistic error repairs contain editing terms is consistent with linguistic and factual error repairs having distinct organizations. A larger data set should allow us to explore the nature of this distinction further.

5. ACKNOWLEDGEMENTS

This research was funded by Grant 16 of 2019 by the Ministry of Science, Innovation and Technology of Colombia—MINCIENCIAS. The first author is grateful to the Universidad Nacional de Colombia for granting her the study leave that allowed this research.

6. REFERENCES

[1] Nooteboom, S. 2010. Monitoring for speech errors has different functions in inner and overt speech. In Everaert, M., Lentz, T., De Mulder, H., Nilsen, Ø. (eds.), *The Linguistic Enterprise*. Amsterdam: John Benjamins, 213–233.

[2] Plug, L. 2016. Informativeness, Timing and Tempo in Lexical Self-Repair. *Language and speech*. 59/4, 516–543.

[3] Seyfeddinipur, M., Kita, S., & Indefrey, P. 2008. How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition*, 108/3, 837–842.

[4] Nooteboom, S.G. & Quené, H. 2017. Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language*, 95, 19–35.

[5] Nooteboom, S.G. & Quené, H. 2019. Temporal aspects of self-monitoring for speech errors. *Journal of Memory and Language*, 105, 43–59.

[6] Levelt, W. J. M. 1983. Monitoring and self-repair in speech. *Cognition*, 14, 41–104.

[7] Ruiz Vásquez, N.F. 2020. The Dialects of Colombian Spanish: A New Classification. *Lenguaje* (Cali, Colombia). 48/2, 160–195.

[8] Figueroa, M.A., García, D.A. and Salamanca, G.F. 2019. DiapixSp: adaptación al español y aplicación piloto de una herramienta de elicitación de habla espontánea y colaborativa. *Estudios de fonética experimental*. 257–288.

[9] Van Engen, K.J., Baese-Berk, M., Baker, R.E., Choi, A., Kim, M. and Bradlow, A.R. 2010. The Wildcat Corpus of Native-and Foreign-accented English: Communicative Efficiency across Conversational Dyads with Varying Language Alignment Profiles. *Language and speech*. 53/4, 510–540.

[10] Boersma, P., Weenink, D. 2017. Praat: Doing phonetics by computer. www.praat.org

[11] Goldman, J., & Schwab, S. 2014. EasyAlign Spanish: an (semi-)automatic segmentation tool under Praat. In Y. Congosto, M. L. Montero, & A. Salvador (Eds.), *Fonética experimental, Espacio Europeo de Educación Superior e Investigación* 1, 629-640. Arco/Libros.

[12] R Core Team. 2022. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[13] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *Lme4: Linear mixed-effects models using Eigen and S4*. Version 1.1

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

[15] Levelt, W.J.M. 1989. *Speaking. From intention to articulation*. Cambridge, Massachusetts: The MITPress.

[16] Hartsuiker, R. J., & Kolk, H. H. J. 2001. Error monitoring in speech production: A computationaltest of the perceptual loop theory. *Cognitive Psychology*, 42, 113–157.