How to repair a slip of the tongue?

Andy Lücking[1,2] **and Jonathan Ginzburg**[1]
[1]Université Paris Cité, CNRS,
Laboratoire de Linguistique Formelle (UMR 7110)
[2]Goethe University Frankfurt
luecking@em.uni-frankfurt.de, yonatan.ginzburg@u-paris.fr

Abstract

A slip of the tongue (SoT) is by no means a random occurrence and usually gets self-repaired immediately. The reparandum, however, remains available in context as potential anaphoric antecedent. So at least two puzzles for dialogue theory emerge: (i) how to deal with reparandum anaphora, and (ii) how is immediate repair possible? To provide answers, we make two extensions to the dialogue framework KoS (Ginzburg, 2012): Firstly, we spell out SoT repair as an "intra-utterance move" which utilizes a conversational rule drawing on *intended meaning*; Secondly, by reviewing current cognitive science work, we connect the linguistic types postulated by KoS to a pointer-based neurocognitive architecture and thereby sketch an explanatory dialogical model of SoT repair.

1 Introduction

Besides polished parlance, the domain of natural language use also knows slips of the tongue (SoT), or *lapsus linguae*. A well known example is senator Edward Kennedy's (1), transcribed here following Pincott (2012):[1]

(1) Our national interest ought to be to encourage the *breast* .. the best and the brightest

The SoT in (1) is a *substitution error* where the sound /r/ from *brightest* is anticipated and interferes the production of *best*, leading to an erroneously produced *breast*. However, the SoT is self-monitored and immediately repaired. The transcription used in (1) also exemplifies a methodological problem: why is the SoT transcribed as *breast* instead of, say, homophone *brest*? The reason very likely is just the joy of Freudian interpretations, a presumably dubious construal of speech errors which we do not follow further here—see also Cutler (1981) on "[t]he reliability of speech error data".

Unintentionally produced expressions such as SoTs are somewhat awkward for semantic theorizing: they are (arguably) not licensed by a grammar rule, nor are they part of the intended content of the to-be produced utterance.[2] They may also result in sounds which do not match the phonology of any word in the given language, although they virtually never violate the phonological constraints of that language (Wells, 1951). As a consequence they have been excluded from linguistic theory and competence as "grammatically irrelevant conditions" (Chomsky, 1965, p. 3). However, SoTs nonetheless influence turn-taking, other/self-repair and grounding, and are not on the whole arbitrary (Nooteboom, 1969; Harley, 2006). They can also figure as antecedents of anaphoric expressions: An addressee or overhearer of (1) can pick up the erroneously produced segment by means of a *Wh*-phrase or even a "salience anaphora" (cf. Asher and Wada, 1988).

(2) I heard *what you said first* / *it*.

The successful resolution of an anaphoric relation presupposes that the target *relatum* is available in context—in case of (2) and (1) this is the substitution error segment. Hence, for dealing with anaphora concerning SoTs, we need a notion of context that keeps track of lapses like of other speech items.

SoTs happen in every modality, be it spoken, written, or signed (Fromkin, 1980), but given the temporally detached communication mode in particular of writing, detected errors are usually erased right away—and even more easily so with electronic help—before any text is published.[3] Therefore it may be warranted to "idealize away" speech

[1]The corresponding video recording can be watched here: https://www.youtube.com/watch?v=SVJ0-cWr_PY, accessed 2nd May 2022.

[2]They *can* be used intentionally as part of, say, a joke, however.

[3]"In speaking, however, erasure is a physical impossibility,

errors from written, proof-read sentence-oriented grammars, but they are arguably impossible to ignore in (spoken) dialogue theory. Although it may not be part of a speaker's competence to *produce* speech errors, it *is* part of linguistic competence how to deal with them (see Ginzburg et al., 2014, p. 57 for a related argument concerning disfluencies). In fact, about one in three speech errors do get self-repaired (Levelt, 1983, p. 44). Furthermore, as also argued by Ginzburg et al. (2014), self-repaired speech errors pattern with other-repairs, a well-established type of clarification interaction (Schegloff et al., 1977). Hence, a unified account of self- and other-initiated repair needs to be provided by linguistic dialogue theory. We follow Ginzburg et al. (2014) in this respect, but here, following Postma et al. (1990), we distinguish SoTs from other cases of disfluencies/self-repair/self communication management:[4] while the former are proper speech errors in the sense that produced and planned speech diverge, the latter signal problems in the execution of a speech plan; differentiation between both may not be sharp, though.

Dependence on planned speech not only distinguishes SoTs from disfluencies in general, it also induces temporal constraints. Psycholinguistic research on speech lapses focuses on immediate repair (see Sec. 2). But repair detection can be delayed. An anonymous reviewer of SemDial came up with the following example:

(3) A: I think I'll wear my green dress. Can you bring it to me please?
B: OK [leaves to go get dress].
A: Wait, did I say green? Sorry, I meant my red dress.
B: OK, I'll get it. But your original choice was better.

We are somewhat dubious about whether this self-correction should be viewed as a SoT: we think this category should not include errors based on apparent *intention change*, as this one seems to be. We discuss one classification of speech errors below and hypothesize that SoTs do not felicitously allow for editing phrases like 'I didn't mean X', though drawing the line is clearly tricky.

In any case, it is clear that a repair can virtually be delayed for an arbitrary period of time: (speaking to Ann) "Did I really call you 'Barbara' last Christmas?". The temporal range of repair hence seem to be constrained by memory. In this regard, at least three temporal windows can be distinguished:

- immediate repair due to perceptual monitoring (Fig. 2) as in (1);
- repair within the reach of rehearsal of utterances within the phonetic loop within working memory (Baddeley, 2012), as in (3);
- referring to conversations which are stored as episodes within episodic memory (Ginzburg and Lücking, 2020) ("Barbara").

Of course, if a SoT remains unaltered (or undetected) without affecting the ongoing of the actual conversation, repair becomes superfluous; there is a decay of importance of repair, bound up with dialogical relevance. For this reason—memory issues aside—there is a strong prevalence of immediacy of repair. In fact, issuing non-immediate repair needs a special preparation to bring the reparandum into focus again—cf. the "Wait, did I say green?" phrase in (3). This is reminiscent of the pragmatic "one-moment-" or "just-a-minute-test" (Shanon, 1976, p. 248) for addressing presupposed contents. Hence, immediate SoT repair seems to be a uniform articulatory and time-bounded phenomenon which deserves a treatment on its own.

SoT repair usually is self-repair. This follows from its immediacy which is coupled to self-monitoring, but is also due to primary "editing rights" or even obligations of the speaker, as exemplified in (4), taken from a transcript of the TV show *Parks and Recreation*, where the addressee (Tom) claims a SoT concession from the speaker (Jerry):[5]

and its seeming social equivalent is only a polite convention that usually works only superficially. Nevertheless, all of us do try to cover up some of our lapses." (Hockett, 1967, p. 100)

[4]Terminology here involves important presuppositions. In generative linguistics and in NLP, it has been common to use the term 'disfluency' which carries the implication that the phenomena in question are somehow deviant from normal fluency. CA's term 'repair' makes the phenomenon more intentional, in line with works such as (Clark and FoxTree, 2002), which incorporate filled pauses into the lexicon. Allwood's term 'self communication management' goes the whole hog towards intentionalizing the phenomenon. The latter is, arguably, inappropriate for SoTs. We will mostly stick with 'repair', but occasionally use 'disfluencies' where the literature has already established this.

[5]`https://tvquot.es/parks-and-recreation/quote/u7lrn5nc/`, accessed 26th July 2022.

(4) JERRY: For my murinal, I was inspired by the death of my grandma.
TOM: [laughs] You said "murinal."
JERRY: No, I didn't.
ANN: Yes, you did. You said "murinal". I heard it.

We therefore formulate SoT repair as a speaker-independent linguistic resource below but acknowledge that it is mainly used (if at all) by the producer of a lapsus lingua.

In section 2 we briefly review some common types of SoTs. Section 3.1 introduces the basic ingredients of the formal dialogue theory KoS (Ginzburg, 2012), which is used in section 3.2 to adopt the analysis of backward-looking disfluencies from Ginzburg et al. (2014). Section 3.3 refines the previous analysis and bridges to a neural construal of SoT repair. The neural construal is taken up again in section 4, where simple networks are replaced with current semantic pointer-based architectures within spiking neuron populations. Inasmuch as slips of the tongue exemplify an interface phenomenon of dialogue theory and linguistic processing, a common analysis framework of this kind is needed.

2 Kinds of slips of the tongue

An enormous variety and detail of categories of speech errors has been observed (Crystal, 1997, p. 265). Pfau (2009) classifies the errors he found in his speech error corpus into four kinds (which in turn are partitioned into sub-kinds; examples are his):

- semantic anticipation or perseveration (e.g., substituting *potato* for *onion* or *vice versa*)
- errors involving feature mismatch (e.g., plural verb form following a singular subject noun phrase but which involves a plural genitive)
- stranding or shift of an abstract feature (e.g., perseveration of the plural feature onto a noun)
- errors involving accommodation (post error process where a follow-up error accommodates the error-induced context to grammatical constraints)

The first three classes roughly correspond to the most frequent error types Garnham et al. (1981) observed in the London-Lund corpus, namely substitution and anticipation at segment and word level. Some examples are collected in (5):

(5) a. "taddle tennis" instead of "paddle tennis" (segment, anticipation; Fromkin, 1973a, p. 112)

b. "I can't cook worth a cam" instead of "I can't cook worth a damn" (segment, perseveration; Fromkin, 1973a, p. 112)

c. "Seymour sliced the knife with a salami" instead of "Seymour sliced the salami with a knife" (word reversal; Fromkin, 1973b, Appendix)

d. "Take it out to the porch – eh – verandah." (word, substitution; Laver, 1969, p. 138)

We will therefore mainly focus on these kinds of SoT in the following.

Note that SoTs can also occur in sequence. Weir (2018) retells one of Nazbanou Nozari's—a cognitive scientist—stories about a research participant who was shown a picture of a sheep and called it "wolf". He corrected the incorrect classification to "steep" and then to "sleep". Remarkably, as pointed out by Weir (*op. cit.*), "'Wolf' is related to 'sheep' in meaning, 'steep' is related in sound, and 'sleep' in both meaning and sound." Hence, there are semantic and phonological crossover effects.

The given examples—exceptional cases aside—as well as received knowledge of SoTs show that they are a rather local phenomenon. Harley (2006, p. 740) provides a spot-on summary: "Sounds only exchange across small distances, whereas words can exchange across phrases; words that exchange tend to come from the same syntactic class, whereas sound exchange errors are not constrained in this way, but instead swap with words regardless of their syntactic class." This means that SoTs have to be accounted for *sentence-internally*. Accordingly, they get detected by monitoring mechanism during speech production (Hartsuiker and Kolk, 2001). The repair of a detected SoT follows a common pattern, which is, simplified from Levelt (1983, p. 45), shown in Fig. 1. Following an utterance which contains the *reparandum*, the repair is sometimes prepared by an editing phase[6] and provides the repairing expression, the *alteration*.

[6]In fact, in the corpus study of Switchboard by Hough (2015) fewer than 15% of self-repairs involved an editing phrase.

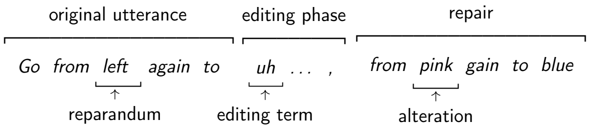

Figure 1: The structure of repair, simplified from Levelt (1983, p. 45).

Indicating correction with the editing phase can happen in several ways, including:

- the use of an editing term as in (5d) and Fig. 1; the editing term may also consist in the repetition of part of the original utterance before the reparandum;
- an aborted production during or after the reparandum and re-start, as in (1)—this is usually accompanied by stressed intonation starting with a guttural sound (Laver, 1969).

Repairing a SoT can happen as both, self-repair and other-repair, although correcting an unintentionally produced lapse of the dialogue partner may be regarded impolite, in particular when the intended utterance is recognized easily.

3 Formal model of SoT repair

3.1 Background: TTR and KoS

Given the intra-sentential domain of SoTs (see above), a highly incremental framework is needed. We use KoS (Ginzburg, 2012) in this respect. KoS is formulated in TTR (Cooper and Ginzburg, 2015; Cooper, 2022), a *Type Theory with Records*. Following the model of (perceptual) classification, a crucial notion of TTR is a *judgement*, this is in general that object a is of type T, notated $a : T$. More complex semantic issues such as an assertion that a situation is of a certain situation type draws on structured entities called *records* (token level) and *record types*. An assertion is then modelled in terms of an *Austinian proposition* as a judgement between *records* and *record types*:

(6) Austinian proposition :=
$$\begin{bmatrix} \text{sit} & : \textit{Rec} \\ \text{sit-type} & : \textit{RecType} \end{bmatrix}$$
and true iff *Rec* : *RecType*, i.e., the situation is of the type specified by the record type—in this case, *Rec* is a *witness* for *RecType*.

Linguistic parsing is construed along this way, too: sign types classify speech event tokens. Following work in *Head-Driven Phrase Structure Grammar* (Pollard and Sag, 1994), the basic sign architecture projects from lexical items to phrases and sentences and is as follows:

(7) *Sign* :=
$$\begin{bmatrix} \text{phon} & : \textit{List}(\textit{Phoneme}) \\ \text{cat} & : \textit{SynCat} \\ \text{constits} & : \textit{Set}(\textit{Sign}) \\ \text{dgb-params} & : \textit{RecType} \\ \text{cont} & : \textit{SemObj} \end{bmatrix}$$

The *constits* field collects all daughter elements of a (complex) sign, the *dialogue gameboard parameters* (dgb-params) provide an interface to context. Work on dialogue brought about significant refinements of the structure of context extending beyond a speaker addressing an addressee at a given time and place with a speech event, leading to *dialogue gameboards* (DGB; Ginzburg, 2012):

(8) *DGBType* :=
$$\begin{bmatrix} \text{spkr} & : \textit{Ind} \\ \text{addr} & : \textit{Ind} \\ \text{utt-time} & : \textit{Time} \\ \text{s-event} & : \textit{Rec} \\ \text{c-utt} & : \text{addressing(spkr,addr,s-event,utt-time)} \\ \text{FACTS} & : \textit{Set}(\textit{Prop}) \\ \text{Pending} & : \textit{List}(\textit{LocProp}) \\ \text{Moves} & : \textit{List}(\textit{LocProp}) \\ \text{QUD} & : \textit{PoSet}(\textit{InfoStruc}) \end{bmatrix}$$

FACTS represent shared assumptions in terms of a set of propositions. Dialogue moves that are in the process of being grounded or under clarification are the elements of the *Pending* list. Already grounded moves are moved to the *Moves* list. Within *Moves* the first element has a special status given its use to capture adjacency pair coherence and is referred to as *LatestMove*. The current question under discussion is tracked in the *QUD* field. It is structured as a partially ordered set whose topmost element is called *MaxQUD*. QUD not only tracks a question, but also an antecedent focal expression, the *focus establishing constituent* (FEC), hence its contents are objects of type *InfoStruc*:

(9) *InfoStruc* :=
$$\begin{bmatrix} \text{q} & : \textit{Question} \\ \text{FEC} & : \textit{LocProp} \end{bmatrix}$$

The sign-based classification of a phonetic speech event is a special kind of Austinian proposition called *locutionary proposition* (*LocProp*), a record–record type-pair consisting of a speech event u_0 as record and a sign as situation type in such a way that the *phon* value of the sign type correctly classifies the speech event and the entries within *dgb-params* are witnessed in a record w_0. Given these notational conventions, a record type of type *LocProp* has the following structure, where $s_0 = u_0 \wedge_{merge} w_0$ (i.e., s_0 is the *merge*, or unification, of u_0 and w_0):

(10) $$\begin{bmatrix} \text{sit} = s_0 : \textit{Rec} \\ \text{sit-type} : \textit{Sign} \end{bmatrix}$$

Updating an information state is licensed by conversational rules, pairs of DGBs of the form *preconditions* and *effects* (sometimes abbreviated as *pre* respectively *eff*):[7]

(11) $$\begin{bmatrix} \text{pre} & : \textit{DGBType} \\ \text{effects} & : \textit{DGBType} \end{bmatrix}$$

For example, if a question is posed, this question becomes—under smooth development, but not invariably (Łupkowski and Ginzburg, 2017)—the current *question under discussion*:

(12) Ask QUD-incrementation :=:

$$\begin{bmatrix} \text{pre} : & \begin{bmatrix} \text{q} & : \textit{Question} \\ \text{LatestMove}{=}\textit{Ask}(\text{spkr,addr,q}) & : \textit{IllocProp} \\ \text{u}_{\text{fec}} {\in} \text{MaxPending.sit.constits} & : \textit{LocProp} \end{bmatrix} \\ \text{eff} : & \begin{bmatrix} \text{QUD}{=}\begin{bmatrix} \text{q}{=}\text{pre.q} \\ \text{FEC}{=}\text{u}_{\text{fec}} : \textit{LocProp} \end{bmatrix} : \textit{InfoStruc} \end{bmatrix} \end{bmatrix}$$

A DGB is the agent-specific structure of context which constitutes the publicized part of information states:

(13) *TotalInformationState* :=

$$\begin{bmatrix} \text{private} : \textit{PRType} \\ \text{public} \ : \textit{DGBType} \end{bmatrix}$$

Given this formal background, an account of SoT repair can be given.

[7]The pair of preconditions and effects notated as a single record type abstracts over deductive and temporal aspects: they are means to *classify* interactions. Seen as processing resources (a point raised by an anonymous reviewer), they can be regarded as functional types (*if* preconds, *then* effects).

3.2 Previous work on backward looking disfluencies

A repair can potentially occur at any place of an ongoing utterance. Hence the preconditions of a dialogical repair are rather weak, presupposing only that Pending is non-empty. Should we add as an additional condition divergence from intended production? This conflicts with repair that involves repetition—a highly pervasive phenomenon (Hough, 2015). In a probabilistic setting, which we are not assuming here, this condition could be formulated as insufficient confidence in the reparandum. In the absence of that, we will not include a divergence condition in the rule for backwards looking repairs, but explicate it in terms of a trigger stated at the level of the private cognitive state. Based on work on meaning-oriented clarification requests, giving rise to a class of conversation rules called *Clarification Context Update Rule* (Ginzburg, 2012), the (potentially accommodated) MaxQUD of the eff(ect) of the repair resource amounts to the issue of *What did the speaker mean by u_{fec}?* This MaxQUD requires the next (if an editing phrase has been produced) or simultaneous (without editing phrase) move (the new LatestMove) to provide an answer—an utterance which is *co-propositional* with u_{fec}. This has been formalized as *Backward Looking Appropriateness Repair* by Ginzburg et al. (2014, p. 42):

(14) $$\begin{bmatrix} \text{pre} : & \begin{bmatrix} \text{spkr} : \textit{Ind} \\ \text{addr} : \textit{Ind} \\ \text{Pending}{=}\ \langle p0, \text{rest}\rangle : \textit{List}(\textit{LocProp}) \\ \text{u}_{\text{fec}} : \textit{LocProp} \\ \text{c1} : \text{member}(\text{u}_{\text{fec}}, p0.\text{sit.constits}) \end{bmatrix} \\ \text{eff} : & \begin{bmatrix} \text{MaxQUD} = \\ \begin{bmatrix} \text{q}{=}\ \lambda x ?\textit{Mean}(\text{pre.spkr},\text{pre.u}_{\text{fec}},x) \\ \text{FEC}{=}\text{u}_{\text{fec}} \end{bmatrix} \\ \quad : \textit{InfoStruc} \\ \text{LatestMove} : \textit{LocProp} \\ \text{c2} : \textit{CoProp}(\text{LatestMove}^{\text{cont}},\text{MaxQUD}) \end{bmatrix} \end{bmatrix}$$

(The superscript "cont" abbreviates the path LatestMove.sit-type.cont.) What does it mean that a question (MaxQUD) and a general semantic object (LatestMove$^{\text{cont}}$; including individuals, properties, propositions) are co-propositional? Ginzburg et al. (2014, p. 30) provide the following characterisation in terms of "answerhood" (where "utterances" denotes the range of expressions from fragments to full sentences):

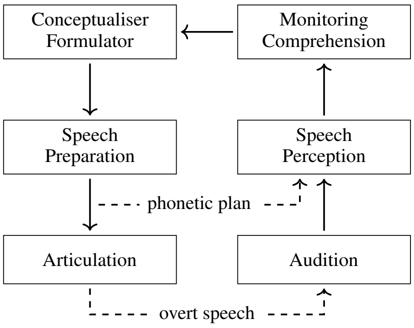

Figure 2: Dual perception loop (simplified from Levelt, 1989; Levelt et al., 1999).

(15) *Co-propositionality*

a. Two utterances u_1 and u_2 are *co-propositional* iff the questions q_1 and q_2 they contribute to QUD are co-propositional.

b. q_1 and q_2 are co-propositional if there exist a record r such that $q_1(r) = q_2(r)$.

That is, an utterance is co-propositional to an intended meaning MaxQud if the question projected from the utterance provides the same result (i.e., answer) when applied to a given record as MaxQud.

The rule in (14) already allows to analyze SoTs: an issue with the reparandum $\mathrm{u}_{\mathrm{fec}}$ has come up as MaxQUD. The content of the repair provides the alteration of the reparandum. In the aftermath of repair, Pending has to be modified in such a way that within the original utterance the alteration substitutes for the reparandum (*Pending replacement*).

3.3 Application

Since a speech error happens if the verbal utterance diverges from the planned one, speech error correction can occur if the divergence is detected. According to the *Dual Perception Loop* model (Levelt, 1989; Levelt et al., 1999) self-monitoring happens on two routes: the intended utterance is compared to both its phonetic plan ("inner speech") and to the perceived speech output, see Fig. 2.[8] This can be modelled in a pretty straightforward manner by incorporating a phonetic plan into the private share of interlocutors' total information states (*PRType*).

[8]Since in aphasic patients a dissociation between comprehension and error-detection ability has been observed, there is evidence that the fast inner loop does not rest on an internalized speech comprehension monitor but rather uses production signals (Nozari et al., 2011).

(16) $\mathit{PRType} \sqsupseteq \left[\text{PhonPlan} : \mathit{List}(\mathit{RecPhon})\right]$

where *RecPhon* is a reduced variant of a locutionary proposition which consists of a (mental) speech event and its phonological classification (according to the model sketched in Fig. 2, syntactic and semantic aspects pertain to the Conceptualiser and Formulator levels).

(17) $\mathit{RecPhon} := \left[\begin{array}{ll}\text{s-event} & : \mathit{Rec} \\ \text{phon-struc} & : \mathit{Sign}.\text{phon}\end{array}\right] \sqsubset \mathit{LocProp}$

A observes a speech error iff A.private.PhonPlan.i $\not\sqsubseteq$ A.public.Pending.i, for any list element with index i which is appended to the incrementally increasing list of *RecPhon*s respectively *LocProp*s.

Let us apply these tools in order to analyze example (1). The original utterance before the SoT occurs consists in the speech event e_0 = "Our national interest ought to be to encourage". The parse up to this point (speaking in terms of HPSG) has found an NP ("Our national interest") and an incomplete verb cluster headed by "ought", but the argument structure of "encourage" still requires an NP argument. If an NP argument follows, the verb cluster can be completed by means of a *head-argument-structure* and finally combined with the subject NP into a *head-filler-structure*. We abbreviate the chart loosely following Ginzburg et al. (2020) as T_{natint} as follows, including found (fnd) and still req(uired), anticipated information:

(18) $T_{\text{natint}} = \left[\begin{array}{ll}\text{e}_1 : [\text{Our national interest ought to be to encourage}] & \\ \text{e}_2 : [\text{Our national interest}] & \\ \text{e}_3 : [\text{encourage}] & \\ \text{e}_4 : \left[\begin{array}{l}\text{fnd1}=\text{e}_2 : \mathit{Sign}.\text{cat}=\text{NP} \\ \text{fnd2}=\text{e}_3 : \mathit{Sign}.\text{cat}=\text{V} \\ \text{req1}=\langle\text{NP,head-arg-struc}\rangle : \mathit{GramStruc} \\ \text{req2}=\langle\text{head-cluster-struc}\rangle : \mathit{GramStruc} \\ \text{req3}=\langle\text{head-filler-struc}\rangle : \mathit{GramStruc}\end{array}\right] & \end{array}\right]$

Since the chart type in (18) mentions still missing grammatical structures (type *GramStruc*; we only listed the ones needed for the example) it generates hypothesis about its continuation and can be used to construct *Pending* and *QUD* simultaneously and incrementally. K(ennedy)'s DGB therefore can be classified as (19).

(19) K.dgb1 =

$$\left[\begin{array}{l} \text{spkr}= \text{K} : \mathit{Ind}\\ \text{addr} : \mathit{Set}(\mathit{Ind})\\ \text{s-event}= e_0\\ \text{Pending}=\left\langle \left[\begin{array}{l}\text{sit}= e_0\\ \text{sit-type}= T_{\text{natint}}\end{array}\right]\right\rangle\\ \text{Moves}=\langle\ \rangle\\ \text{QUD}=\left\langle \left[\begin{array}{ll} \multicolumn{2}{l}{\text{q=?MaxPending}}\\ \text{FEC}= & \left[\begin{array}{ll} \multicolumn{2}{l}{\text{sit=[national interest]}}\\ \text{sit-type}= & \left[\begin{array}{l}\text{phon}=\langle\text{/national/,}\\ \text{/interest/}\rangle\\ \text{cat=NP} : \mathit{SynCat}\\ \text{cont} : \mathit{Ind}\end{array}\right]\end{array}\right]\end{array}\right]\right\rangle \end{array}\right]$$

Up to this point Kennedy's *PhonPlan* is satisfied, which continues as in (20):

(20) $$\left[\text{K.private.PhonPlan}=\left\langle\text{the,best,and,the,brightest}\right\rangle\right]$$

The definite article, the next utterance token, complies with both the PhonPlan and the NP requirement of the chart type. The utterance of the future reparandum could in principle be a noun and therefore complete the sentence:

(21) K.dgb2 =

$$\left[\begin{array}{l} \text{spkr}= \text{K} : \mathit{Ind}\\ \text{addr} : \mathit{Set}(\mathit{Ind})\\ \text{s-event}= e_0\\ \text{Pending}=\left\langle \left[\begin{array}{l}\text{sit}= e_0\\ \text{sit-type}= T_{\text{enc}}\end{array}\right]\right\rangle\\ \text{Moves}=\langle \mathit{Assert}(\text{K},\left[\begin{array}{l}\text{sit}= e_0\\ \text{sit-type}= T_{\text{enc}}\end{array}\right])\ \rangle\\ \text{QUD}=\left\langle \left[\begin{array}{ll} \multicolumn{2}{l}{\text{q=?MaxPending}}\\ \text{FEC}= & \left[\begin{array}{ll} \multicolumn{2}{l}{\text{sit=[national interest]}}\\ \text{sit-type}= & \left[\begin{array}{l}\text{phon}=\langle\text{/national/,}\\ \text{/interest/}\rangle\\ \text{cat=NP} : \mathit{SynCat}\\ \text{cont} : \mathit{Ind}\end{array}\right]\end{array}\right]\end{array}\right]\right\rangle\\ \text{FACTS}=\left\{\mathit{Classify}(T_{\text{enc}}, e_0)\right\} \end{array}\right]$$

where T_{enc} is a sentence parse:

(22) $$\left[\begin{array}{l}\text{e}_1 : \text{[Our national interest ought to be to encourage}\\ \quad\text{the bre(a)st]}\\ \text{e}_2 : \left[\text{fnd=e}_2 : \mathit{Sign}\text{.cat=S}\right]\end{array}\right]$$

However, the classification of the utterance e_0 by the type T_{enc} is unsatisfying:[9] it either involves a "novel" noun (*brest*), or it is semantically awkward (*breast*). Furthermore, there is a mismatch between K.private.PhonPlan and K.public.Pending following the definite article. Hence, an accommodation of *Backward Looking Appropriateness Repair* is triggered, leading to an update of K.dgb.QUD. The question *What did K mean by "breast"?* (or "*brest*") becomes MaxQUD and has to be addressed first: the following LatestMove—*best and brightest*—is constrained to provide a co-propositional value.

(23) K.dgb3 =

$$\left[\begin{array}{l} \text{Pending}=\left\langle \left[\begin{array}{l}\text{sit}= e_0\\ \text{sit-type}= T_{\text{enc}}\end{array}\right]\right\rangle\\ \text{u}_{\text{fec}}=\left[\begin{array}{l}\text{sit=[the breast]}\\ \text{sit-type} : \mathit{Sign}\end{array}\right]\\ \text{c1 : member(u}_{\text{fec}}\text{,MaxPending.sit.constits)}\\ \text{Moves}=\langle \mathit{Assert}(\text{K},\mathit{Mean},\text{u}_{\text{fec}},\\ \quad\left[\begin{array}{l}\text{sit=[the best and the brightest]}\\ \text{sit-type}= N_{\text{bab}} : \mathit{Sign}\end{array}\right])\ \rangle\\ \quad\quad\left[\begin{array}{l}\text{q}= \lambda x?\mathit{Mean}(\text{K},\text{u}_{\text{fec}},x)\\ \text{FEC=u}_{\text{fec}}\end{array}\right],\\ \text{QUD}=\left\langle \left[\begin{array}{ll} \multicolumn{2}{l}{\text{q=?MaxPending}}\\ \text{FEC}= & \left[\begin{array}{ll} \multicolumn{2}{l}{\text{sit=[national interest]}}\\ \text{sit-type}= & \left[\begin{array}{l}\text{phon}=\langle\text{/national/,}\\ \quad\text{/interest/}\rangle\\ \text{cat=NP} : \mathit{SynCat}\\ \text{cont} : \mathit{Ind}\end{array}\right]\end{array}\right]\end{array}\right]\right\rangle\\ \text{FACTS}=\left\{\begin{array}{l}\mathit{Classify}(T_{\text{enc}}, e_0)\\ \mathit{Classify}(T_{\text{bab}},\text{u}_{\text{fec}})\end{array}\right\} \end{array}\right]$$

(Where T_{bab} is the nominal sign type classifying the conjunct *best and brightest*.)

In the aftermath, and if the self-repair is accepted, *Pending Replacement* applies (Ginzburg et al., 2014), leading to a substitution of e_0 and T_{enc} according to the following re-parse:

(24) $$\left[\begin{array}{l}\text{e}_1 : \text{[Our national interest ought to be to encourage}\\ \quad\text{the best and the brightest]}\\ \text{e}_2 : \left[\text{fnd=e}_2 : \mathit{Sign}\text{.cat=S}\right]\end{array}\right]$$

Note that the SoT remains in FACTS, from which it can be retrieved as a constits element of T_{enc}, providing, for instance, an antecedent for reparandum anaphora.

The analysis of the SoT from (1) mainly in (23) recognizes two sources: the unlikely sign-based classification of the reparandum on the one

[9] One can express this by assigning the *Classify* relation a probability threshold (Cooper et al., 2015).

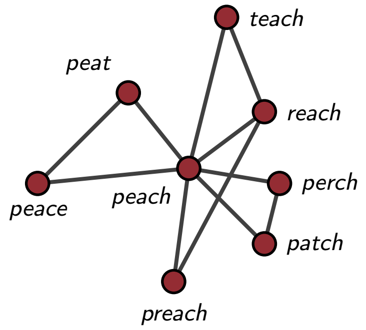

Figure 3: Extract of the phonological network around *peach* (adapted from Vitevitch et al. 2015, p. 32).

hand,[10] and the divergence of phonetic plan and self-monitored speech on the other hand. The *PhonPlan* accesses the mental state of an interlocutor, at least symbolically. We used it to exemplify a formal *post hoc* analysis of SoT repair. However, the non-arbitrary nature of *lapsus linguae* (see section 2) has been explained by activation-spreading models of sentence production (Dell, 1986), which also provide clues for their self-repair (Nooteboom and Hugo, 2020). The underlying rationale can be exemplified by means of a simple, phonetic example in Fig. 3: a word form like *peach* is phonetically similar to *teach* (the only phonetic difference is that the first is produced with an initial bilabial, the latter with an alveolar). Both sounds have not much in common with, say *apple* (although since apples and peaches are both fruits, they are associated in a semantic network). Hence, phonetic distances give rise to a phonological network. Since no exchange of content words with function words have been observed (Harley, 2006, p. 740), we assume that such networks are sorted according to part of speech. Now, if *peach* is to be articulated (cf. Fig. 2) it receives activation. This activation distributes to the neighboring nodes, however, which get co-activated. This co-activation may then lead to choosing the neighboring instead of the planned word and sending it to the articulator.

To summarize:

- How is it that the dialogue proceeds with the corrected utterances but the reparandum is still available as an antecedent? This is because SoT repair amounts to Pending replacement but the original utterance is still available in FACTS.
- Why is the phrase *the best and the brightest* interpreted as a reparandum? In fact, K. could also have uttered the conjunct NP *the breast, the best and the brightest*. Following work on SoT (e.g. Harley, 2006), we assume that the correction interpretation follows from a specific intonation.
- How is the SoT repaired so quickly and seamlessly? Besides the information encoded in the PhonPlan, Nooteboom and Hugo (2020) found evidence that co-activated items not only are the source of lapses but also a cue for their repair: the alteration will also be a node with high activation and therefore more easily accessible.[11]

While the activation spreading model provides crucial explanations for various kinds of *lapsus linguae*, it represents words—either as phonetic forms (segment errors) or as semantic markers ("Freudian" substitutions)—as single nodes. When construed neurally, this is a simplification: lexical items do not correspond to single neurons. Drawing on cognitive science insights, the contour of an integrative framework is emerging, which is sketched in the following section.

4 Activation spreading and semantic pointers

Following recent work in neurocognition, we assume that a *semantic pointer* is a notion that provides a needed level of abstraction within neuronal architectures (see Blouw et al., 2016 for a cognitive science summary). A semantic pointer is a compressed activation of neuronal spiking which is associated with a more elaborate region of neuronal activation, as illustrated in Fig. 4.

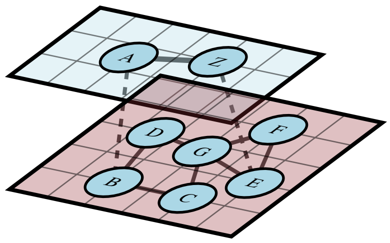

Figure 4: Semantic pointer (top layer): compressed activation of neuronal spiking (bottom layer).

[10] See Oliphint (2022) for a recent metaphysical account of words and the problem of distinguishing them from "non-words".

[11] Since the planned item is available in the PhonPlan and guides the monitoring process, retrieval loops are prevented, as remarked by an anonymous reviewer.

Visual processing, for instance, rests on a large population of spiking neurons which "encode" the visual input.[12] Via neuronal transformations (mathematically modelled as circular convolutions; Eliasmith, 2013) condensed levels of neuronal activation are produced from these large activation patterns which use a (much) less number of spiking neurons.[13] Further processing is sped up by using these abstract semantic pointers. However, the more detailed activation patterns can still be retrieved from the compressed encoding—hence the term semantic *pointer*. Semantic pointers can bind together various levels of activations, such as lexical, perceptual, and motor information. Linguistic forms—(mental representations of) labels such as phonetic strings—can be construed as semantic pointers as well.[14] Such "labeled semantic pointers" provide objects of fast linguistic processing, which may be unpacked in cognitive simulations (Connell, 2019; Goucha et al., 2017). Activation spreading from abstract labels to motor activation patterns implements embodiment (Mahon, 2015). A well-established example of a pointer-based model is the *hippocampal indexing theory* of Teyler and Rudy (2007). The hippocampus captures neocortical activity generated by an observed episode and projects back to these neocortical regions—hence, the hippocampus creates an index which can be unpacked to the full pattern of neocortical activity produced by the episode.

Instead of single nodes representing linguistic forms or meanings, we construe the nodes of such networks as semantic pointers, condensed levels of activation which can be unpacked by larger populations of neuronal spiking from which they are abstracted in the first place. SoT repair as outlined here and forward-looking disfluencies analysed by Ginzburg et al. (2014) appear as two sides of the same coin: a SoT is the result of too much, a disfluency of too little activation—probably everyone has experienced the latter as a tip-of-the-tongue feeling.

[12]But see Brette (2019) for a critical assessment of the encoding metaphor.

[13]Activation patterns are modeled in terms of mathematical vectors and synaptic weights. Now two vectors can be combined into a single one of the same dimensionality and later decomposed again. Hence there is some commonality of brain-based semantic pointer convolutions and popular data-based Deep Learning methods (Rasmussen, 2019).

[14]This view squares with the coordinative role of material symbols in cognition as argued by Clark (2006).

5 Discussion

We offer an account of SoT repair which rests on the notions of co-propositionality and intended meaning clarification from previous work on disfluencies in dialogue (Ginzburg et al., 2014). This account solves two linguistic puzzles which arise from correcting a lapse in a principled way: The utterance containing the reparandum is available within the assumptions shared by the interlocutors (FACTS—since the tongue slipped as a matter of fact) as an antecedent of reparandum anaphora; the repaired move including the alteration becomes MaxPending (the topmost move within the list of pending ones) and contributes to further dialogue progressing. The alteration is immediately retrievable since it is a pre-activated item. We noted that a repair is indicated by an explicit editing phrase, or by a specific intonation pattern, which signal that an utterance provides alterations (expressed by the meaning-pertaining question under discussion (QUD) $\lambda x?Mean(\mathrm{A},\mathrm{u}_{\mathrm{fec}},x)$) instead of, say, just continuing a dialogical exchange (incorporating phonetic repair-indicating details still has to be worked out, however). It has also been observed that SoTs follow phonetic or semantic constraints and that repair happens on very small time scales, virtually immediately. Psycholinguistic models provide explanations for these observations, mainly in terms of spreading activation architectures. In this respect it has been sketched how TTR types representing signs and *LocProp*s can be construed as labeled semantic pointers that compress larger populations of spiking neurons and are compatible with activation spreading. There is evidence that repair, and immediate repair, is part of dialogical competence (cf. sections 1 and 2). In order to provide an explanation not only of the semantic but also of the temporal aspects of this competence—cf. *Did I say X?* and the failure marking *I meant X* editing phrase for delayed repair—we think that formal models of meaning in dialogue eventually need to draw on processing models. This becomes much more pressing when considering multimodal interaction (which is the default form of dialogue): here temporal alignment of communicative means of various channels occurring both sequentially and (partially) simultaneously give rise to timing as an aspect of interaction *sui generis* (e.g. Lücking and Ginzburg, 2020; Rieser and Lawler, 2020). Timing in language, however, seems to be inextricably bound up with processing.

Using previous work on repair in general, we tried to develop formal tools for analysing SoTs which take the immediacy of repair into account. Besides their analytical properties, the formal tools make the case for a multi-level implementation. The symbolic dialogue theory—if construed cognitively—allows to represent linguistic, label-based compositional processing in a precise way, which is somehow carried out within speakers' brains (Frankland and Greene, 2020). However, linguistic labels are underpinned by statistical associations which help to speed-up processing (Connell, 2019)—but also may lead to production lapses, as reviewed in Sec. 3.3. Following work by Eliasmith (2013) and colleagues, these statistically associated labels can be construed as semantic pointers (Sec. 4); they can be unpacked to retrieve fully grounded semantic models, namely when condensed pointer-based representations in one brain area lead to replay (e.g., in recollection) or simulate (e.g., embodied sensori-motor processing) full representations usually from another brain area (Louwerse and Connell, 2011). The latter is involved in memory-based repair of stored episodes in contrast to immediate SoT repair, as mentioned in Sec. 1. In this sense, our formal, dialogical model of SoT repair suggests a specific interaction of statistical and symbolic semantic approaches, because both seem to target quite different aspects of meaning (Westera and Boleda 2019; see also Lücking et al. 2019): formal semantics provides the analytic backbone for defining semantic ontologies and providing scientifically precise, cognitively potent content representations, statistical regularities add inter-label associations which are important to capture temporal aspects of processing and understanding—and for producing *lapsus linguae* in the first place. To this we add semantic pointers to connect labels to the brain and to distinguish linguistic processing short-cuts from full mental simulations.

On a more general level this means that a (renewed) cooperation of semantics and cognitive science is required. Cognitive science develops processing models, but semantics and pragmatics contribute a precise structuring of the contents and contexts involved in processing. We think that formal dialogue theory, in particular KoS with its focus on spoken language, provides a useful semantic framework in this respect: KoS is already formulated in a way that is close to speech processing models (cf. notions such as LocProp and Pending) and fused with a WM model (Ginzburg and Lücking, 2020). Recent cognitive science work on the other hand seems to narrow down the gap between symbolic and neuronal levels of computation. Phenomena such as SoTs live on the interface of those levels and therefore are a lens into neuronally grounded dialogue semantics.

So how to repair a slip of the tongue? From all the co-activated items, retrieve the alteration which is co-propositional to the focal reparandum and complies with the inner PhonPlan, produce the alteration after an editing sound or phrase, move the original utterance to FACTS, and apply Pending Replacement to substitute alteration for reparandum.

Acknowledgments

We acknowledge the support by a public grant overseen by the French National Research Agency (ANR) as part of the program Investissements d'Avenir (reference: ANR-10-LABX-0083). It contributes to the IdEx Université Paris Cité ANR-18-IDEX-0001. We thank the SemDial reviewers for very helpful comments.

References

Nicholas Asher and Hajime Wada. 1988. A computational account of syntactic, semantic and discourse principles for anaphra resolution. *Journal of Semantics*, 6(1):309–344.

Alan Baddeley. 2012. Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63:1–29.

Peter Blouw, Eugene Solodkin, Paul Thagard, and Chris Eliasmith. 2016. Concepts as semantic pointers: A framework and computational model. *Cognitive Science*, 40(5):1128–1162.

Romain Brette. 2019. Is coding a relevant metaphor for the brain? *Behavioral and Brain Sciences*, 42. Published online by Cambridge University Press: 16 July 2018.

Noam Chomsky. 1965. *Aspects of the Theory of Syntax*, 50th anniversary edition edition. MIT Press, Cambridge, MA.

Andy Clark. 2006. Material symbols. *Philosophical Psychology*, 19(3):291–307.

Herb Clark and Jean FoxTree. 2002. Using uh and um in spontaneous speech. *Cognition*, 84:73–111.

Louise Connell. 2019. What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience*, 34(10):1308–1318.

Robin Cooper. 2022. *From perception to communication: An analysis of meaning and action using a theory of types with records (TTR)*. Oxford University Press.

Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. 2015. Probabilistic type theory and natural language semantics. *Linguistic Issues in Language Technology – LiLT*, 10:1–43.

Robin Cooper and Jonathan Ginzburg. 2015. Type theory with records for natural language semantics. In Shalom Lappin and Chris Fox, editors, *The Handbook of Contemporary Semantic Theory*, 2 edition, chapter 12, pages 375–407. Wiley-Blackwell, Oxford, UK.

David Crystal. 1997. *The Cambridge Encyclopedia of Language*, 2 edition. Cambridge University Press, Cambridge, UK.

Anne Cutler. 1981. The reliability of speech error data. *Linguistics*, 19(7-8):561–582. Guest editorial.

Gary S. Dell. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3):283–321.

Chris Eliasmith. 2013. *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford University Press, Oxford.

Steven M. Frankland and Joshua D. Greene. 2020. Concepts and compositionality: In search of the brain's language of thought. *Annual Review of Psychology*, 71(1):273–303. PMID: 31550985.

Victoria Fromkin, editor. 1980. *Errors in Linguistic Performance*. Academic Press, New York.

Victoria A. Fromkin. 1973a. Slips of the tongue. *Scientific American*, 229(6):110–117.

Victoria A. Fromkin, editor. 1973b. *Speech Errors as Linguistic Evidence*. Mouton, The Hague and Paris.

Alan Garnham, Richard C. Shillcock, Gordon D. A. Brown, Andrew I. D. Mill, and Anne Cutler. 1981. Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics*, 19:805–817.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford, UK.

Jonathan Ginzburg, Robin Cooper, Julian Hough, and David Schlangen. 2020. Incrementality and HPSG: Why not? In Anne Abeillé and Olivier Bonami, editors, *Constraint-Based Syntax and Semantics: Papers in Honor of Danièle Godard*. CSLI Publications, Stanford, CA.

Jonathan Ginzburg, Raquel Fernández, and David Schlangen. 2014. Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics*, 7(9):1–64.

Jonathan Ginzburg and Andy Lücking. 2020. On laughter and forgetting and reconversing: A neurologically-inspired model of conversational context. In *Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue*, SemDial/WatchDial.

Tomás Goucha, Emiliano Zaccarella, and Angela D. Friederici. 2017. A revival of *Homo loquens* as a builder of labeled structures: Neurocognitive considerations. *Neuroscience & Biobehavioral Reviews*, 81:213–224. The Biology of Language.

T. A. Harley. 2006. Speech errors: Psycholinguistic approach. In Keith Brown, editor, *Encyclopedia of Language & Linguistics*, 2 edition, pages 739–745. Elsevier, Oxford.

Robert J. Hartsuiker and Herman H. J. Kolk. 2001. Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, 42(2):113–157.

Charles F. Hockett. 1967. Where the tongue slips, there slip I. In (Fromkin, 1973b), pages 93–119.

Julian Hough. 2015. *Modelling Incremental Self-Repair Processing in Dialogue*. Ph.D. thesis, Queen Mary, University of London.

John D. M. Laver. 1969. The detection and correction of slips of the tongue. In (Fromkin, 1973b), pages 132–143.

Willem J. M. Levelt. 1983. Monitoring and self-repair in speech. *Cognition*, 14(1):41–104.

Willem J. M. Levelt. 1989. *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.

Willem J. M. Levelt, Ardi Roelofs, and Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1):1–38.

Max Louwerse and Louise Connell. 2011. A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, 35(2):381–398.

Andy Lücking, Robin Cooper, Staffan Larsson, and Jonathan Ginzburg. 2019. Distribution is not enough – going Firther. In *Proceedings of Natural Language and Computer Science*, NLCS 6.

Andy Lücking and Jonathan Ginzburg. 2020. Towards the score of communication. In *Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue*, SemDial/WatchDial.

Paweł Łupkowski and Jonathan Ginzburg. 2017. Query responses. *Journal of Language Modelling*, 4(2):245–292.

Bradford Z. Mahon. 2015. What is embodied about cognition? *Language, Cognition and Neuroscience*, 30(4):420–429. PMID: 25914889.

Sieb G. Nooteboom. 1969. The tongue slips into patterns. In (Fromkin, 1973b), pages 144–156.

Sieb G. Nooteboom and Quené Hugo. 2020. Repairing speech errors: Competition as a source of repairs. *Journal of Memory and Language*, 111.

Nazbanou Nozari, Gary S. Dell, and Myrna F. Schwartz. 2011. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, 63(1):1–33.

Jared S. Oliphint. 2022. Using a two-dimensional model from social ontology to explain the puzzling metaphysical features of words. *Synthese*, 200.

Roland Pfau. 2009. *Grammar as Processor*. Number 137 in Linguistik Aktuell / Linguistics Today. John Benjamins, Amsterdam and Philadelphia.

Jena E. Pincott. 2012. Slips of the tongue. *Psychology Today*.

Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. CSLI Publications, Stanford, CA.

Albert Postma, Herman Kolk, and Dirk-Jan Povel. 1990. On the relation among speech errors, disfluencies, and self-repairs. *Language and Speech*, 33(1):19–29. PMID: 2283918.

Daniel Rasmussen. 2019. NengoDL: Combining deep learning and neuromorphic modelling methods. *Neuroinformatics*, 17:611–628.

Hannes Rieser and Insa Lawler. 2020. Multi-modal meaning – an empirically-founded process algebra approach. *Semantics and Pragmatics*, 13(8):n/a.

Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382.

Benny Shanon. 1976. On the two kinds of presuppositions in natural language. *Foundations of Language*, 14(2):247–249.

Timothy J. Teyler and Jerry W. Rudy. 2007. The hippocampal indexing theory and episodic memory: updating the index. *Hippocampus*, 17(12):1158–1169.

Michael S. Vitevitch, Rutherford Goldstein, and Elizabeth Johnson. 2015. Path-length and the misperception of speech: Insights from network science and psycholinguistics. In Alexander Mehler, Andy Lücking, Sven Banisch, Philippe Blanchard, and Barbara Job, editors, *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, pages 29–45. Springer, Berlin and New York.

Kirsten Weir. 2018. Lab work: The secrets behind slips of the tongue. *Monitor on Psychology*, 49(3).

Rulon Wells. 1951. Predicting slips of the tongue. In (Fromkin, 1973b), pages 82–87.

Matthijs Westera and Gemma Boleda. 2019. Don't blame distributional semantics if it can't do entailment. In *Proceedings of the 13th International Conference on Computational Semantics – Long Papers*, pages 120–133, Gothenburg, Sweden. Association for Computational Linguistics.