

# Dysfluencies as Intra-Utterance Dialogue Moves

Jonathan Ginzburg Raquel Fernández David Schlangen  
yonatan.ginzburg@univ-paris-diderot.fr

June 4, 2013

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Dealing with Dysfluencies</b>	<b>3</b>
2.1	Background . . . . .	3
2.2	Desiderata for a Theory of Dysfluencies . . . . .	4
2.2.1	Dysfluencies are recognised incrementally, for which information about their meaning is required . . . . .	4
2.2.2	Dysfluencies have immediate discourse effects . . . . .	5
2.2.3	Dysfluencies are related to other dialogue moves . . . . .	5
2.2.4	Dysfluencies are in the grammar . . . . .	7
2.3	Previous Work on Dysfluencies . . . . .	9
<b>3</b>	<b>Dysfluencies as Intra-active Meaning</b>	<b>11</b>
3.1	Dialogue GameBoards . . . . .	11
3.2	Grounding and Clarification . . . . .	16
<b>4</b>	<b>From CRs to Dysfluency: Informal Sketch</b>	<b>25</b>
<b>5</b>	<b>Dysfluency Rules</b>	<b>26</b>
5.1	An Incremental Perspective . . . . .	26
5.2	Backward Looking Dysfluencies . . . . .	28
5.3	Some More BLD Examples . . . . .	31
5.4	Forward Looking Dysfluencies . . . . .	33
<b>6</b>	<b>Conclusions</b>	<b>37</b>
	<b>References</b>	<b>39</b>

## Abstract

Although dysfluent speech is pervasive in spoken conversation, dysfluencies have received little attention within formal theories of grammar. The majority of work on dysfluent language has come from psycholinguistic models of speech production and comprehension and from structural approaches designed to improve performance in speech applications. In this paper, we argue for the inclusion of this phenomenon in the scope of formal grammar, and present a detailed formal account which: (a) unifies dysfluencies (self-repair) with Clarification Requests, without conflating them, (b) offers a precise explication of the roles of all key components of a dysfluency, including editing phrases and filled pauses, (c) accounts for the possibility of self-addressed questions in a dysfluency.

## 1 Introduction

Although dysfluencies are pervasive in spoken conversation, they have typically been viewed by theoretical linguists as the *untouchables* of language—elements not fit to populate the grammatical domain. Their very existence is a significant motivation for the competence/performance distinction (Chomsky, 1965) and for the assumption that spoken language is not the input for language acquisition (Chomsky, 1972). Indeed even quite recently researchers highly skeptical of the competence/performance distinction could suggest that ‘[t]he competence approach uncontroversially excludes performance mishaps such as false starts, hesitations, and errors from the characterization of linguistic knowledge’ (Seidenberg, 1997).

In contrast to this malign attitude to dysfluencies, in a seminal paper Schegloff, Jefferson, and Sacks (1977) initiated the study of such utterances among conversation analysts, showing that self-corrections share many properties with clarificational and correctional utterances made by the other interlocutor. Over the last twenty years there has been increasing interest in the study of self-corrections, hesitations, and other dysfluencies among psycholinguists (e.g. Levelt (1983), Clark and FoxTree (2002), and Bailey and Ferreira (2007)), phoneticians (e.g. Candea, Vasilescu, Adda-Decker, et al. (2005) and Horne (2012)) and computational linguists and researchers on speech processing (e.g. Shriberg (1994), Heeman and Allen (1999), and Johnson and Charniak (2004)).

In this paper, we present a detailed formal grammatical account which:

1. unifies dysfluencies (self-repair) with Clarification Requests (CRs), without conflating them,
2. offers a precise explication of the roles of all key components of a dysfluency, including editing phrases and filled pauses,
3. accounts for the possibility and range of self-addressed questions in a dysfluency.

Beyond the need for assuming an *incremental* perspective towards language processing, an assumption that has in any case become increasingly influential in recent years (Kempson, Meyer-Viol, & Gabbay, 2000; Rieser & Schlagen, 2011a), our account will involve positing no additional mechanisms beyond those already needed for the interpretation of dialogue. We will see that dysfluencies manifest precisely the characteristics one expects of a grammatical phenomenon: they exhibit both significant cross-linguistic variation at all linguistic levels and also potential universals and, far from constituting meaningless *noise*, participate in semantic and pragmatic processes such as anaphora, conversational implicature, and discourse particles, as illustrated in (1). In all three utterances in this example, the semantic process is dependent on the *reparandum* (the phrase to be repaired) as the antecedent:

- (1) a. Peter was, well, he was fired. (Example from Heeman and Allen (1999); anaphor refers to material in reparandum)

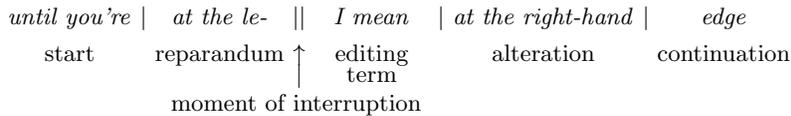


Figure 1: General pattern of self-repair

- b. A: Because I, any, anyone, any friend, anyone, I give my number to is welcome to call me (Example from the Switchboard corpus (Godfrey, Holliman, & McDaniel, 1992)) (implicature based on contrast between repair and reparable: ‘It’s not just her friends that are welcome to call her when A gives them her number’)
- c. The other one did, no, other ones did it. (Example from BNC (file KB8, line 1705); material negated by “no” originates in the reparable )

The structure of the paper is the following: in Section 2 we review the ‘syntax’ of dysfluencies, give a classification of types of dysfluencies, make some observations about what desiderata for a discourse theory of dysfluencies are, in particular arguing that it needs to be grounded within a grammar, and we critically review previous work on dysfluencies. Section 3 provides background about the formal dialogue theory we utilize, KoS<sup>1</sup> (Ginzburg, 2012), and in particular explains how it can be used to analyze clarification interaction. In Section 4 we offer an informal sketch of our analysis of dysfluencies. Section 5 spells out this analysis for the two classes of dysfluencies that we argued earlier need to be distinguished. Section 6 offers some brief conclusions.

## 2 Dealing with Dysfluencies

### 2.1 Background

As has often been noted (see e.g. Levelt (1983), and references therein for earlier work), speech dysfluencies follow a fairly regular pattern. The elements of this pattern are shown in Figure 1, annotated with the labels introduced by Shriberg (1994) (who was building on Levelt (1983)).

Of these elements, all but the moment of interruption and the continuation are optional. The presence of elements and their relations can be used as the basis for classifying dysfluencies into different types (McKelvie, 1998; Heeman & Allen, 1999):

- If the alteration differs strongly from the reparable and does not form a coherent unit together with the start, or if alteration and continuation are not present at all, the dysfluency can be classified as an *aborted utterance* / *fresh start*.
- If the alteration ‘replaces’ the reparable, the dysfluency is a *repair*.
- If the alteration elaborates on the reparable, it is a *reformulation*.

The following gives examples for these three classes, in the order they were mentioned:<sup>2</sup>

- (2) a. { *I mean* } [ [ *I, + I,*  ] + [ *there are a lot, + there are so many* ] ] *different songs,*
- b. [ *We were + I was* ] *lucky too that I only have one brother.*

<sup>1</sup>KoS is not an acronym, despite emphasizing a Conversationally Oriented Semantics.

<sup>2</sup>These examples, and most others in this section, are taken from the Switchboard corpus (Godfrey et al., 1992), with dysfluencies annotated according to Godfrey et al. (1992): ‘+’ marks the moment of interruption and separates reparable from alteration, ‘{ }’ brackets editing terms and filled pauses, and ‘[]’ brackets the dysfluency as a whole.

- c. *at that point, [ it, + the warehouse ] was over across the road*

Within the class of repairs, a further distinction can be made (Levelt, 1983):

- *appropriateness-repairs* replace material that is deemed inappropriate by the speaker given the message she wants to express (or has become so, after a change in the speaker’s intentions or in the state of the world that is being described), while
- *error-repairs* repair material that is deemed erroneous by the speaker.

Finally, these types of dysfluencies can be, with a nod to the similarly named distinction in the DAMSL annotation scheme (Allen & Core, 1997), labelled *backward-looking* dysfluencies, as here the moment of interruption is followed by an alteration that refers back to an already uttered reparandum. We can distinguish from these types those dysfluencies where the moment of interruption is followed not by an alteration, but just by a completion of the utterance which is delayed by a filled or unfilled pause (hesitation) or a repetition of a previously uttered part of the utterance (repetitions). We will call this kind of dysfluency *forward-looking*;<sup>3</sup> the following gives some examples of such dysfluencies.

- (3)
- a. From Shriberg (1994):  
Show flight arriving in uh Boston.
  - b. From Besser and Alexandersson (2007):  
And also the- the dog was old.
  - c. From Levelt (1989):  
A vertical line to a- to a black disk
  - d. From Switchboard Corpus (file sw2020):  
Yeah. / {D Well, } [ I, + I ] don’t really have anything against rap music. / I, -/ the one thing I do object to about rap music [ is, + is ] when it becomes militant,
  - e. From Switchboard Corpus (file sw2028):  
{C So, } it’s been inordinately warm, {F uh, } here, [ for, + {F uh, } for ] this time of year.

## 2.2 Desiderata for a Theory of Dysfluencies

We now make some observations about dysfluencies that a theory of their semantics and pragmatics must address.

### 2.2.1 Dysfluencies are recognised incrementally, for which information about their meaning is required

As with many kinds of linguistic structure, the structure of a dysfluency (as indicated in Figure 1) is not given *en bloc*, but rather must be recognised incrementally. The listener faces what Levelt (1983) called the *continuation problem*, which is roughly the problem of how to integrate the material from the alteration into the previous material; the solution of this problem requires computation of what the reparandum is.

Levelt (1983, p. 492) proposes rules based on lexical identity (“word identity convention”) and categorial identity (“category identity convention”). We will be proposing to add to these rules content-based conventions for identifying the reparandum. The semantics of the reparandum can also be more directly relevant to the semantics of the alternation, namely in cases where anaphora in the alternation involves reference to an entity introduced in the reparandum which is not meant to be repaired or corrected (i.e., the antecedent is part of the “anticipatory retracing”), as in the following examples:

---

<sup>3</sup>Levelt (1983) refers to such dysfluencies as “covert repair”.

- (4) From Shriberg (1994):  
Our dog likes- he loves the beach.
- (5) From Heeman and Allen (1999) (repeated from above (1)):  
[ Peter was + { well } he was ] fired.
- (6) From Milward and Cooper (1994):
- a. The three main sources of data come., uh ..., they can be found in the refer-  
ences [reconstructed from actual utterance]
  - b. Every boy should uh... he should have taken a water bottle with him.[constructed]
- (7) From TRAINS:<sup>4</sup>  
9.1-5 M: so we should move the engine at Avon engine E to  
10.1 S: engine E1  
11.1 M: E1  
12.1 S: okay  
13.1-3 M: engine E1 to Bath to  
13.4-5 : or we could actually move it to Dansville to pick up the boxcar there  
14.1 S: okay

### 2.2.2 Dysfluencies have immediate discourse effects

Recent psycholinguistic studies have shown that both the simple fact that a dysfluency is occurring and its content can have immediate discourse effects, which show in different behaviour of listeners. E.g., Bailey and Ferreira (2007) found that “filled pauses may inform the resolution of whatever ambiguity is most salient in a given situation”, and Brennan and Schober (2001) found that in a situation with two possible referents, the fact that a description was self-corrected enabled listeners to draw the conclusion that the respective other referent was the correct one, before the correction was fully executed.

Similarly, Arnold, Kam, and Tanenhaus (2007) showed that during reference what we call forward-looking dysfluencies allow listeners to infer that the speaker is having difficulty with lexical retrieval, which in a reference identification task leads listeners to look at those objects that are more difficult to name; a finding that has been replicated in a corpus study on more naturalistic dialogues reported in Schlangen, Baumann, and Atterer (2009). (Interestingly, as Arnold et al. (2007) report, the effect of the dysfluencies to make reference to difficult-to-describe objects more likely goes away if listeners are told their partners suffer from aphasia and have problems finding words.)

### 2.2.3 Dysfluencies are related to other dialogue moves

Figure 2 illustrates the continuity between more typically described types of (discourse) correction and clarification on the one hand and dysfluencies on the other. It shows (constructed) examples of ‘normal’ discourse correction (a), two uses of clarification requests (b & c), correction within a turn (d), other-correction mid-utterance (e), and two examples of self-correction as discussed above (f & g). The first four examples clearly are instances of phenomena within the scope of discourse theories. What about the final two?

There are clear similarities between all these cases: (i) material is presented publicly and hence is open for inspection; (ii) a problem with some of the material is detected and signalled (i.e. there is a ‘moment of interruption’); (iii) the problem is addressed

---

<sup>4</sup>Poesio (1995) about (7): “[in the fresh start in utterances 13.4-5] *S* replaces the proposal introduced in 9.1-13.2 with a new one, but in doing so he assumes that the engine at Avon, engine E1 is part of the common ground. If the repair process were to take place before discourse referents are established and reference resolution is performed, the referent would be removed, and we would end up with a pronoun without antecedent.”

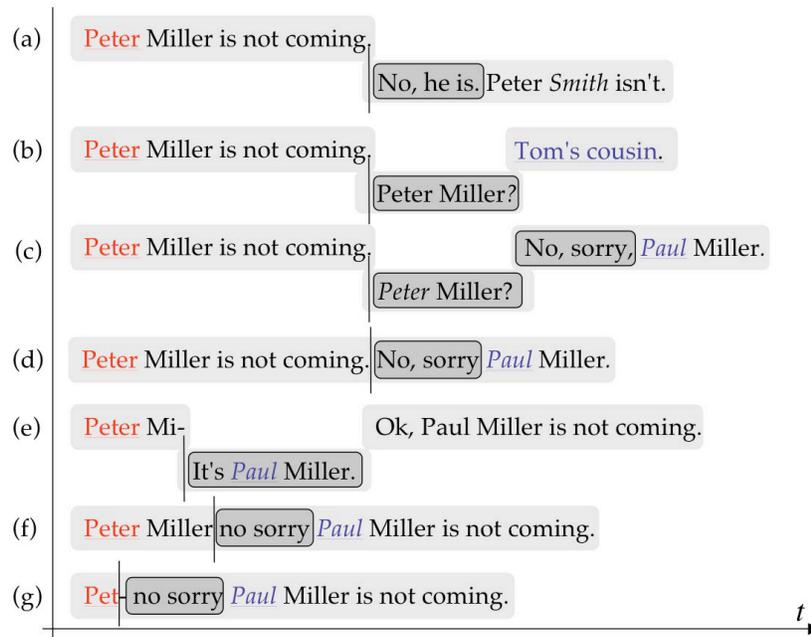


Figure 2: Continuity between (discourse) corrections and clarifications and dysfluencies

and repaired, leaving (iv) the incriminated material with a special status, but within the discourse context. That (i)-(iii) describe the situation in all examples in Figure 2 should be clear; that (iv) is the case also for self-corrections can be illustrated by the next example (repeated from above), which shows that self-corrected material is also available for later reference:

(8) [*Peter was + {well} he was*] fired

Moreover, even though this is not the most frequent form such “within-utterance repairs” take, it is quite possible for the other dialogue participant to take over the turn during both backward looking and forward looking dysfluencies, which further argues for not artificially separating them from other dialogue moves. The following (constructed and attested) examples illustrate this:

(9) (constructed)  
 A: And then Peter performed a hystorect- ehm \*hytorese\*  
 B: \*hysterectomy\*  
 A: er yeah right hystorectomy on the patient.

(10) (constructed)  
 A: Now take thee... um right.  
 B: auger?

(11) a. From BNC (file: KPJ 550-551):  
 A: Chilli, has, has, has never really been [pause] er  
 B: A big seller.  
 b. From Pentomino corpus (file: 20061123.pento\_nonoise):  
 P: so that goes - remember where we were having so much fun where they were adja- those

- E: kissing ?  
P: the kissing pieces ?  
E: yeah
- c. From BNC (file: KPU 471-474):  
A: Well Tuesday is my busiest day. I'm getting  
B: What?  
C: some more in.
- d. From BNC (file: KS1 789-791):  
A: I'm pretty sure that the  
B: Programmed visits?  
A: Programmed visits, yes, I think they'll have been debt inspections...

We take this as evidence that it would be desirable to have a model that brings out these similarities between these phenomena, while respecting their differences.

### 2.2.4 Dysfluencies are in the grammar

In the introduction, we already mentioned that grammarians have usually assumed that an analysis of dysfluencies is outwith the scope of the grammar; indeed their existence is an important motivation for the competence/performance distinction. The question of whether to include a set of linguistic utterance types X within the grammar has frequently preoccupied grammarians, but has rarely been addressed systematically.<sup>5</sup> We offer here various arguments for why the view of a dysfluency-free grammar is untenable, though, as will become clear, the discussion raises some deep issues that we cannot resolve here.

For a start, it is instructive to think about dysfluencies by analogy with friction. Non-dysfluent speech is analogous to frictionless motion. Some of the time it is useful to ignore the effects of friction, but the theory of motion is required to explicate the existence and quantitative effects of friction. Whereas it seems plausible that not all dysfluencies are consciously produced *by the speaker*, for the addressee they always form part of the string of phonemes perceived which needs to be parsed and interpreted.

More concretely, dysfluencies display an important characteristic of grammatical processes, namely cross-linguistic variation. This has been documented in some detail in comparative work between morphosyntactic aspects of repair on a wide range of languages by Fox and collaborators (e.g. (Fox et al., 1996; Wouk et al., 2009; Fox et al., 2010))<sup>6</sup> and in phonetic analysis of hesitation markers (Candea et al., 2005).<sup>7</sup> Here we

---

<sup>5</sup>See Jackendoff (2005), who provides various arguments *contra* the core v. periphery distinction.

<sup>6</sup>In a study of seven languages with significantly different typological characteristics Wouk et al. (2009) find important correlations between the diversity of length in a language's lexicon and the site of repair initiation: for instance, Chinese displays a strong preference for initiating repair in monosyllabic words, in contrast to Japanese where the preference is for initiation in multisyllabic words. Fox et al. (2010) demonstrate significant differences across English, Hebrew, and German in the distribution of words where recycling (reutterance of a word, typically as a hesitation device) and replacement (repairs where the alteration is distinct from the reparandum, used in self-correction) occur: for instance, English's majoritarian category for recycling is the subject pronoun, whereas for both German and Hebrew it is the preposition; German replacement favours verbs and determiners, in marked contrast to English and Hebrew, which favour nouns. Patterns such as these seem strongly related to word order and complexity of inflectional morphology.

<sup>7</sup>For phonetic analysis of cross-linguistic variation see Candea et al. (2005), who compared fillers in Arabic, Mandarin Chinese, French, German, Italian, European Portuguese, American English and Latin American Spanish:

Language-specific features can be observed in the segmental structure of the fillers. French, for example, prefers a vocalic segment as filler realization, whereas English prefers vowels followed occasionally by a nasal coda consonant [m]... In Portuguese as well, more complex diphthongized segments can be found. To conclude, for some languages the vocalic support of the fillers might be a segment exterior to the vocalic system of the language (i.e. Italian in our corpus). However,

briefly note some evidence concerning hesitation markers and editing phrases. Concerning the former, we note that there is some variation in how hesitation is typically expressed in various languages, as exemplified in (12). Indeed, some languages, e.g. Mandarin and Japanese, use demonstratives for this role:

- (12)
- a. ‘uh’ ‘um’ (English) (Clark & FoxTree, 2002)
  - b. ‘euh’ ... (French): tu sais c’était un peu euh : : l’ambiance santa-Barbar-euh (De Fornel and Marandin (1996), example (1a))
  - c. ‘em’, ‘eh’ (Modern Hebrew): spkr1: im male male eh em ta’alot mayim kaele ktanot shama besin hem eh ohavim eh (662-667, TripToFarEast:44, <http://hebrewcorpus.nmelrc.org/>)
  - d. Mandarin: ‘en’, ‘nage’ (literally ‘that’), ‘zhege’ (literally ‘this’) (Zhao & Jurafsky, 2005)
  - e. Japanese: ‘ano’, ‘(so)no’, ‘kou’ (Yoshida & Lickley, 2010)

With respect to the latter, a child acquiring English needs to discover that ‘no’ can be used in a self-correction, but, for instance the closely related word ‘nope’ cannot. Similarly, a trilingual acquiring English, German, and French will need to learn that ‘enfin’ can be used in a self-correction, whereas ‘finally’ and ‘schließlich’, which are often interchangeable with ‘enfin’, cannot be so used:

- (13) Quand ma belle mère **enfin** quand ma femme apelle (De Fornel & Marandin, 1996, example (2a))

Conversely, we suggest that dysfluencies are also involved in grammatical universals. We postulate the following:

- (14)
- a. if NEG is a language’s word that can be used as a negation and cross-turn correction, then NEG can be used as an editing phrase in backward-looking dysfluencies.
  - b. No (English): The other one did, no, other ones did it. (BNC, KB8, line 1705)
  - c. Non (French): Il a trente-cinq francs par semaine **non** vingt-cinq pardon (*He had 35 francs per week, no 25 sorry.*)(De Fornel and Marandin (1996), example (2b))
  - d. Nein (German): Dann muß Du nach links **nein** rechts gehen. (*Then you have to go left, no right.*)
  - e. lo (Hebrew): ani, **lo** at batmuna. (*I, no you are in the picture.*)
  - f. No (Catalan): Centenars - **no**, milers de persones es manifesten a Barcelona per forçar la negociació dels convenis. (*hundreds - no, thousands of people take part in a demonstration in Barcelona to force the negotiation of the agreements.*)

These considerations argue for the fact that the elements participating in dysfluencies are subject to phonological, syntactic, and semantic constraints internal to individual languages, as well as exhibiting universal properties common to many languages. They strongly suggest, then, that dysfluencies are part and parcel of grammatical systems of natural languages.

Of course part of the reluctance to accord dysfluency-containing utterances the status of utterances internal to the grammatical system derives from the assumption that the task of grammar is to characterize the “well formed” utterances of a given language. The

---

all the eight languages seem to accept as fillers’ vocalic support at least one of the vowels of their vocalic system (Candea et al., 2005, p. 48).

force of this view has weakened with the increasing recognition that ‘grammaticality’ is a gradable rather than a classifying notion (Keller, 2000). Thus, Lappin (2013) propose a gradient notion of grammaticality that arises via a set of scoring procedures for mapping the logprob value of a sentence on the basis of the properties of the sentence and the corpus containing the sentence. Such a view can be generalized into a view of grammar as a mechanism that enables us to characterize the coherently interpretable conversational events.

### 2.3 Previous Work on Dysfluencies

Dysfluencies have received a fair amount of attention both in psycholinguistics and in computational linguistics. In this section we give a brief overview of the most prominent approaches in these fields. To the best of our knowledge, none of the existing approaches has studied dysfluencies from a semantic point of view, incorporating them into the grammar, and has put forward a general framework that offers a treatment of dysfluencies alongside other dialogue moves—as we shall propose here.

It is not surprising that computational linguists have been concerned with dysfluencies because automatic natural language understanding systems that deal with spoken input cannot succeed unless dysfluencies can be handled. The main concern of computational linguists has been to detect and process dysfluencies automatically. To this end, many corpus studies have been performed, which have provided very valuable information on the structural properties, the distributional characteristics, and the frequency of different types of dysfluencies (Godfrey et al., 1992; Shriberg, 1994, 1996; Besser & Alexandersson, 2007). This information has been exploited to recognise dysfluencies automatically either by means of rules (McKelvie, 1998; Core, 1999) or by leveraging statistical information (Stolcke & Shriberg, 1996; Heeman & Allen, 1999).

Detecting the presence of dysfluencies is of course only the first step in being able to handle them appropriately. In computational linguistics, the predominant approach to processing dysfluencies after they have been detected has been to filter them out before or along parsing, prior to any process of semantic interpretation (Stolcke & Shriberg, 1996; Heeman & Allen, 1999; Charniak & Johnson, 2001). While this kind of filtering approach may have practical advantages (as the interpretation module does not have to deal with dysfluencies), theoretically such a model is implausible, given that rather long segments can be self-corrected (as in the next example), so that this model would entail the claim that interpretation can lag behind for arbitrarily long intervals, running against much evidence in psycholinguistics for the immediacy of interpretation (as we mentioned in Section 2.2.1). The filtering approach has therefore received strong criticism from authors in psycholinguistics (Lickley, 1994; Ferreira & Bailey, 2004).<sup>8</sup>

- (15) A.1: {D Well,} the first thing for me is [ I wonder, + I see ] a couple of different ways of talking about what privacy is, {F um,} if [ privacy is something that disturbs your private state, + {E I mean} an invasion of privacy is something that disturbs your private state, ] / that’s one thing, / {C and} if privacy is something that comes

Recently, in computational linguistics a proposal was put forward (Hough & Purver,

---

<sup>8</sup>Although by and large computational linguists have adopted a filtering approach for practical reasons, we should point out that they have also been critical of it on theoretical grounds. For instance, Core and Schubert (1999) point to examples such as “*Take the oranges to Elmira um I mean take them to Corning*”, where filtering out the reparandum would leave an anaphoric pronoun without a referent; see also footnote 4 regarding example (7). In fact, recent approaches in computational linguistics have started to exploit rather than eliminate dysfluencies for language understanding (Schlangen et al., 2009) and language generation (Callaway, 2003; Skantze & Hjalmarsson, 2010).

2012) that sketches a treatment of dysfluencies in an incremental grammar formalism, dynamic syntax (Kempson et al., 2000), and hence fulfills our desideratum of placing these constructions in the grammar. However, this approach, although promising, fails to bring out the similarities between self-corrections and corrections and clarification requests by the other dialogue participant, as it lacks connection to a discourse model.

Within psycholinguistics, researchers have looked into a wide variety of aspects related to dysfluencies. From the point of view of language production, the main concern has been how speakers monitor and correct their speech (Levelt, 1983, 1989; Van Wijk & Kempen, 1987). Regarding language comprehension, some authors have investigated the pragmatic effects triggered by dysfluencies (we have already mentioned several studies in Section 2.2.2 showing that dysfluencies can lead listeners to draw inferences on the information state of the speaker). Clark and FoxTree (2002) claimed that filled pauses (in our terminology, *forward-looking* dysfluencies) are lexical items with the conventionalised meaning “a short / slightly longer break in fluency is coming up”, but no semantic formalisation of this claim is given; the claim is contested, e.g. by Finlayson and Corley (2012). Others have been concerned with how dysfluencies are recognised and processed by the human parser (e.g., Levelt (1983), Ferreira, Lau, and Bailey (2004), Bailey and Ferreira (2007)).

As we mentioned in Section 2.2.1, Levelt (1983) suggested syntactic conventions that would allow listeners to solve the *continuation problem* they face when a repetition or a repair (what we are calling *backward-looking* dysfluencies) is processed: what is the reparandum and where does the repair start? He proposes two syntactic constraints, word identity and category identity, that would guide listeners in identifying the onset of the reparandum. Word identity applies when the first word of the repair is identical to a word in the original utterance, which would then be taken as the point where the reparandum starts. Category identity is meant to apply in cases where there isn’t an identical word but only a match in the syntactic category of a word in the original utterance and the first word of the repair. Levelt sees the interruption moment as a sort of coordinating connective: “The original utterance and the repair are, essentially, delivered as two conjuncts. The syntax of repairing is governed by a rule of well-formedness, which acknowledges this coordinating character of repairs” (Levelt (1989), p. 499). Ferreira et al. (2004), building in part on Levelt’s ideas, propose a more concrete model cast in the formalism of Tree Adjoining Grammar. Their “disfluency reanalysis” approach centres around a parsing operation of “Overlay”. According to this approach, the incremental parser, upon encountering new material that cannot be attached to an existing node in the syntactic tree being constructed, attempts to overlay the tree corresponding to the alteration material on top of the reparandum tree. For this, the parser relies on recognising root node identities between the syntactic trees of the reparandum and the alteration. The new tree prevails but, crucially, “[t]he reparandum tree has some effect on processing because it was not deleted but rather covered up with the replacement/repair tree. The unique bits of that tree are therefore still somewhat visible to the processor, and so they can affect its operations” (Ferreira et al. (2004), p. 742). This arguably accounts for some processing effects such as a “lingering” effect of the argument structure of a repaired verb.<sup>9</sup>

Since these proposals are strictly concerned with syntactic constraints, it is difficult to judge whether they could allow for some degree of transparency to reach the interpretation processing module. Nevertheless they are interesting because they leave open the possibility that the meaning of the dysfluency and the reparandum could indeed influence the process of dysfluency recognition (hence fulfilling one of our desiderata discussed above). However, both Levelt’s and Ferreira and colleagues’ models also seem to miss the similarities between self-correcting dysfluencies and other types of corrections we have

---

<sup>9</sup>Ferreira et al. (2004) only deal with one-word repairs concerning verb replacements such as *you should put-drop the frog*.

discussed above; they also cannot explain why it seems possible to take over the turn both in backward-looking dysfluencies and forward looking ones, as was shown above.

As will become clear below, our approach incorporates the insights of these models regarding structural parallelism and makes a clear step forward by adding an account of the *semantics* of dysfluencies which, in addition, connects them to other dialogue moves. We start by providing in the next section background on the dialogue framework we use here, namely KoS (see e.g. Ginzburg and Fernández (2010) and Ginzburg (2012)), describing in particular how this framework deals with “between-utterance” clarification moves (of the types (a)-(c) from Figure 2). In Section 4 we then sketch the (very few) extensions that are needed to capture dysfluencies as well, which we will develop formally in Section 5. We defer to future work the important tasks of specifying a grammar that can incorporate incremental parsing and interpretation of dysfluency-containing utterances and the identification of reparanda.

### 3 Dysfluencies as Intra-active Meaning

#### 3.1 Dialogue GameBoards

KoS is formulated within the framework of Type Theory with Records (Cooper, 2005, 2012; Cooper & Ginzburg, (forthcoming)), a model-theoretic descendant of Martin-Löf Type Theory (Ranta, 1994) and situation semantics (Barwise & Perry, 1983; Cooper & Poesio, 1994; Ginzburg & Sag, 2000). TTR enables one to develop a semantic ontology, including entities such as events, propositions, and questions. With the same means TTR enables the construction of a *grammatical* ontology consisting of utterance types and tokens and of an interactional domain in which agents utilize utterances to talk about the semantic universe. What makes TTR advantageous for our dialogical aims is that it provides access to both types and tokens at the object level. This plays a key role in developing metacommunicative interaction, as we shall see below, in that it enables simultaneous reference to both utterances and utterance types.

For current purposes, the key notions of TTR are the notion of a *judgement* and the notion of a *record*.

- **The typing judgement:**  $a : T$  classifying an object  $a$  as being of type  $T$ .
- **Records:** A record is a set of fields assigning entities to labels of the form ((16a)), partially ordered by a notion of *dependence* between the fields—dependent fields must follow fields on which their values depend. A concrete instance is exemplified in ((16)). Records are used here to model events and states, including utterances, and dialogue gameboards.<sup>10</sup>

$$(16) \quad \text{a.} \quad \left[ \begin{array}{l} l_1 = val_1 \\ l_2 = val_2 \\ \dots \\ l_n = val_n \end{array} \right]$$

- $\left[ \begin{array}{ll} x & = -28 \\ \text{e-time} & = 2\text{AM, Feb 17, 2011} \\ \text{e-loc} & = \text{Nome} \\ \text{c-temp-at-in} & = \text{o1} \end{array} \right]$

<sup>10</sup>Cooper and Ginzburg ((forthcoming)) suggest that for events with even a modicum of internal structure, one can enrich the type theory using the *string theory* developed by Tim Fernando (e.g. Fernando (2007)).

- **Record Types:** a record type is simply a record where each field represents a judgement rather than an assignment, as in (17a).

$$(17) \quad \text{a.} \quad \left[ \begin{array}{l} l_1 : T_1 \\ l_2 : T_2 \\ \dots \\ l_n : T_n \end{array} \right]$$

The basic relationship between records and record types is that a record  $r$  is of type  $RT$  if each value in  $r$  assigned to a given label  $l_i$  satisfies the typing constraints imposed by  $RT$  on  $l_i$ . More precisely,

$$(18) \quad \text{The record} \quad \left[ \begin{array}{l} l_1 = a_1 \\ l_2 = a_2 \\ \dots \\ l_n = a_n \end{array} \right] \text{ is of type:} \quad \left[ \begin{array}{l} l_1 : T_1 \\ l_2 : T_2 \\ \dots \\ l_n : T_n \end{array} \right]$$

iff  $a_1 : T_1, a_2 : T_2, \dots, a_n : T_n$

To exemplify this, ((19a)) is a possible type for (16), assuming the conditions in (19b) hold. Records types are used to model utterance types (aka as *signs*) and to express rules of conversational interaction.

$$(19) \quad \text{a.} \quad \left[ \begin{array}{l} x : \text{Ind} \\ \text{e-time} : \text{Time} \\ \text{e-loc} : \text{Loc} \\ \text{C}_{temp-at-in} : \text{temp\_at\_in}(\text{e-time}, \text{e-location}, x) \end{array} \right]$$

b. -28 : Ind; 3:45AM, Feb 17, 2011 : Time; Nome : Loc; o1 : temp\_at\_in(3:45AM, Feb 17, 2011, Nome, -28)

Armed with these basic logical notions, let us return to characterizing conversational states. On the approach developed in KoS, there is actually no single context—instead of a single context, analysis is formulated at a level of information states, one per conversational participant. The type of such information states is given in (20a), which shows the split into a dialogue gameboard, and a private part of the information state. We leave the structure of the private part unanalyzed here (for details on this, see e.g. Larsson (2002)) and focus on the dialogue gameboard, which represents information that arises from publicized interactions. Its structure is given in (20b):

$$(20) \quad \text{a.} \quad \text{TotalInformationState (TIS):} \quad \left[ \begin{array}{l} \text{dialoguegameboard} : \text{DGB} \\ \text{private} : \text{Private} \end{array} \right]$$

b.  $\text{DGBType} = \left[ \begin{array}{l} \text{spkr} : \text{Ind} \\ \text{addr} : \text{Ind} \\ \text{utt-time} : \text{Time} \\ \text{c-utt} : \text{addressing}(\text{spkr}, \text{addr}, \text{utt-time}) \\ \text{Facts} : \text{Set}(\text{Proposition}) \\ \text{Pending} : \text{list}(\text{locutionary Proposition}) \\ \text{Moves} : \text{list}(\text{locutionary Proposition}) \\ \text{QUD} : \text{poset}(\text{Question}) \end{array} \right]$

In this view of context:

- The spkr/hearer roles serve to keep track of turn ownership.
- **FACTS** represents the shared knowledge conversationalists utilize during a conversation. More operationally, this amounts to information that a conversationalist can use embedded under presuppositional operators.
- **Pending**: represents information about utterances that are as yet *ungrounded*.<sup>11</sup> Each element of Pending is, for reasons explained below, a *locutionary proposition*, a proposition individuated by an utterance event and a grammatical type that classifies that event.
- **Moves**: represents information about utterances that have been grounded. The main motivation is to segregate from the entire repository of presuppositions information on the basis of which coherent reactions to the latest conversational move can be computed.
- **QUD**: (mnemonic for Questions Under Discussion)—questions that constitute a “live issue”. That is, questions that have been *introduced for discussion* at a given point in the conversation and not yet been *downdated*: A query  $q$  updates QUD with  $q$ , whereas an assertion  $p$  updates QUD with  $p?$ . There are additional, indirect ways for questions to get added into QUD, the most prominent of which is during metacommunicative interaction (see below). Being maximal in QUD (MAX-QUD) corresponds to being the current ‘discourse topic’ and is a key component in the theory.

A conversational state  $c1$  will be a record  $r1$  such that (21) holds; in other words,  $r1$  should have the make up in (21)a and the constraints in (21)b need to be met:<sup>12</sup>

$$(21) \quad \text{a.} \quad r1 : \text{DGBType} \left[ \begin{array}{l} \text{spkr} = A \\ \text{addr} = B \\ \text{utt-time} = t1 \\ \text{c-utt} = p_{\text{utt}(A,B,t1)} \\ \text{Facts} = \text{cg1} \\ \text{Moves} = \langle m1, \dots, mk \rangle \\ \text{QUD} = Q \end{array} \right]$$

b.  $A : \text{Ind}, B : \text{IND}, t1 : \text{TIME}, p_{\text{utt}(A,B,t1)} : \text{addressing}(A,B,t1),$   
 $\text{cg1} : \text{Set}(\text{Proposition}), \langle m1, \dots, mk \rangle : \text{list}(\text{illocutionaryProposition}), Q : \text{poset}(\text{Question})$

The basic units of change are mappings between dialogue gameboards that specify how one gameboard configuration can be modified into another on the basis of dialogue moves. We call a mapping between DGB types a *conversational rule*. The types specifying its domain and its range we dub, respectively, the *preconditions* and the *effects*, both of which are supertypes of DGBType. Examples of such rules, needed to analyze querying and assertion interaction and whose use is exemplified in (23) below, are given in (22).<sup>13</sup>

<sup>11</sup>Here ‘grounding’ (in the sense of Clark and Schaefer (1989, Clark (1996)) refers to the process of establishing presuppositions that utterances are mutually understood.

<sup>12</sup>In the sequel we omit utterance times for simplicity.

<sup>13</sup>These rules employ a number of abbreviatory conventions. First, instead of specifying the full value of the list Moves, we record merely its first member, which we call ‘LatestMove’. Second the preconditions can be written as a *merge* of two record types  $\text{DGBType}^- \wedge_{\text{merge}} \text{PreCondSpec}$ , one of which  $\text{DGBType}^-$  is a supertype of DGBType and therefore represents predictable information common to all conversational rules;  $\text{PreCondSpec}$  represents information specific to the preconditions of this particular interaction type. Similarly, the effects can be written as a merge of two record types  $\text{DGBType}^0 \wedge_{\text{merge}} \text{ChangePreCondSpec}$ , where  $\text{DGBType}^0$  is a supertype of the preconditions and  $\text{ChangePreCondSpec}$  represents those aspects of the preconditions that have changed. So we can *abbreviate* conversational rules as in (i); the unabbreviated



- $$\begin{array}{l}
\left[ \begin{array}{l}
\text{pre} : \left[ \begin{array}{l}
q : \text{Question} \\
\text{LatestMove} = \text{Ask}(\text{spkr}, \text{addr}, q) : \text{IllocProp}
\end{array} \right] \\
\text{effects} : \left[ \text{qud} = \langle q, \text{pre. qud} \rangle : \text{poset}(\text{Question}) \right]
\end{array} \right] \\
\text{b. QSpec} \\
\left[ \begin{array}{l}
\text{pre} : \left[ \text{qud} = \langle i, I \rangle : \text{poset}(\text{InfoStruc}) \right] \\
\text{effects} : \text{TurnUnderspec} \wedge_{\text{merge}} \left[ \begin{array}{l}
r : \text{AbSemObj} \\
R: \text{IllocRel} \\
\text{LatestMove} = R(\text{spkr}, \text{addr}, r) : \text{IllocProp} \\
c1 : \text{Qspecific}(r, i, q)
\end{array} \right]
\end{array} \right] \\
\text{c. Assert QUD-incrementation} \\
\left[ \begin{array}{l}
\text{pre} : \left[ \begin{array}{l}
p : \text{Prop} \\
\text{LatestMove} = \text{Assert}(\text{spkr}, \text{addr}, p) : \text{IllocProp}
\end{array} \right] \\
\text{effects} : \left[ \text{qud} = \langle p?, \text{pre. qud} \rangle : \text{poset}(\text{Question}) \right]
\end{array} \right] \\
\text{d. Accept move} \\
\left[ \begin{array}{l}
\text{pre} : \left[ \begin{array}{l}
\text{spkr} : \text{Ind} \\
\text{addr} : \text{Ind} \\
p : \text{Prop} \\
\text{LatestMove} = \text{Assert}(\text{spkr}, \text{addr}, p) : \text{IllocProp} \\
\text{qud} = \langle p?, \text{pre. qud} \rangle : \text{poset}(\text{Question})
\end{array} \right] \\
\text{effects} : \left[ \begin{array}{l}
\text{spkr} = \text{pre. addr} : \text{Ind} \\
\text{addr} = \text{pre. spkr} : \text{Ind} \\
\text{LatestMove} = \text{Accept}(\text{spkr}, \text{addr}, p) : \text{IllocProp}
\end{array} \right]
\end{array} \right] \\
\text{e. Fact Update/ QUD Dwndate} \\
\left[ \begin{array}{l}
\text{pre} : \left[ \begin{array}{l}
p : \text{Prop} \\
\text{LatestMove} = \text{Accept}(\text{spkr}, \text{addr}, p) \vee \\
\text{Confirm}(\text{spkr}, \text{addr}, p) : \text{IllocProp} \\
\text{qud} = \langle p?, \text{pre. qud} \rangle : \text{poset}(\text{Question})
\end{array} \right] \\
\text{effects} : \left[ \begin{array}{l}
\text{facts} = \text{pre. facts} \cup \{p\} : \text{Set}(\text{Prop}) \\
\text{qud} = \text{NonResolve}(\text{pre. qud}, \text{facts}) : \text{poset}(\text{Question})
\end{array} \right]
\end{array} \right]
\end{array}$$

We can exemplify how these rules work in (23), which involves discussion and disagreement at the illocutionary level. A poses a query, which via *Ask QUD-incrementation* updates Moves and via *QSpec* licenses B's assertion, which in turn updates Moves via *Assertion QUD-incrementation*. A rejects B's assertion, and then offers her own proposal, which B accepts. This licenses acceptance, incrementation of FACTS and dwndating of QUD via *Accept* and *Fact update/QUD dwndate*, respectively:

- (23) a. A(1): Who's a good candidate?  
B(2): Peter.  
A(3): (3a) No, (3b) Paul is.  
B(4): OK.

b.

Utt.	DGB Update (Conditions)	Rule
initial	MOVES = $\langle \rangle$ QUD = $\langle \rangle$ FACTS = $cg1$	
1	LatestMove := Ask(A,B,q0) QUD := $\langle q0 \rangle$	Ask QUD-incrementation QSpec
2	LatestMove := Assert(B,A,p1) (About(p1,q0)) QUD := $\langle p1?, q0 \rangle$	Assert QUD-incrementation QSpec
3a	LatestMove := Assert(A,B, $\neg p1$ ) (About( $\neg p1, p1?$ )) QUD := $\langle \neg p1?, p1?, q0 \rangle$	Assert QUD-incrementation QSpec
3b	LatestMove := Assert(A,B, p2) (About( $\neg p2, q0$ )) QUD := $\langle p2?, \neg p1?, p1?, q0 \rangle$	Assert QUD-incrementation QSpec
4	LatestMove := Accept(A,B, p2) QUD := $\langle q0 \rangle$ FACTS := $cg1 \wedge p2 \wedge \neg p1$	Assert QUD-incrementation Accept Fact update/QUD downdate

Two comments on (23b). One minor point is that B’s acceptance is vague: we have assumed it involves accepting (3b) and (3a) and is neutral with respect to whether  $q0$  has been exhaustively discussed. But clearly, it could also be interpreted as only accepting (3b) or could also be understood as closing the discussion completely. A more significant point that will apply to other examples we consider below concerns the ordering on QUD. (23b) illustrates why QUD should not be viewed as a stack, but rather a partially ordered set: (3b) addresses the initial question posed, not (directly) the issue of whether *Peter* is a good candidate, the most recently introduced issue. Data such as these as well as from multi-party dialogue, motivated Ginzburg (2012) to propose that when a question  $q$  is pushed onto QUD it doesn’t subsume *all* existing questions in QUD, but rather only those on which  $q$  does not *depend*:

$$(24) \quad q \text{ is } QUD^{mod(dependence)} \text{ maximal iff for any } q_0 \text{ in QUD such that } \neg Depend(q, q_0): \\ q \succ q_0.$$

This is conceptually attractive because it reinforces the assumption that the order in QUD has an intuitive semantic basis. One effect this has is to ensure that any polar question  $p?$  introduced into QUD, whether by an assertion or by a query, subsequent to a wh-question  $q$  on which  $p?$  depends does not subsume  $q$ .<sup>15</sup>

### 3.2 Grounding and Clarification

Given a setup with DGBs as just described and associated update rules, distributed among the conversationalists, it is relatively straightforward to provide a unified explication of grounding conditions and the potential for Clarification Requests (or CRification) and (metacommunicative) correction. We explain how this can be done, while motivating in particular the information associated with the contextual field PENDING. Schegloff (1987) points out that in principle one can request clarification concerning just about anything in a previous utterance. However, corpus studies of CRs in both a general corpus (Purver, Ginzburg, & Healey, 2001), as well as task oriented ones (Rodriguez & Schlangen, 2004;

<sup>15</sup>For extensive discussion on the nature of the ordering on QUD, see Ginzburg (2012) sections 4.3.3, 4.5, and 8.1.4.

Rieser & Moore, 2005) indicate that there are four main categories of CRs:

- **Repetition:** CRs that request the previous utterance (or parts of it) to be repeated:

(25) Tim (1): Could I have one of those (unclear)?  
Dorothy (2): Can you have what?<sup>16</sup>

- **Confirmation:** CRs that seek to confirm understanding of a prior utterance:

(26) Marsha: yeah that's it, this, she's got three rottweilers now and  
Sarah: three? (=Are you saying she's got THREE rottweilers now?)  
Marsha: yeah, one died so only got three now

- **Intended Content:** CRs that query the intended content of a prior utterance:

(27) Tim (5): Those pink things that af after we had our lunch.  
Dorothy (6): Pink things?  
Tim (7): Yeah. Er those things in that bottle.  
Dorothy (8): Oh **I know what you mean.** For your throat?

- **Intention recognition:** CRs that query the goal underlying a prior utterance.

(28) Norrine: When is the barbecue, the twentieth? (pause) Something of June.  
Chris: Thirtieth.  
Norrine: A Sunday.  
Chris: Sunday.  
Norrine: Mm.  
Chris: Why? (= *Why do you ask when the barbecue is*)  
Norrine: Becau Because I forgot (pause) That was the day I was thinking  
of having a proper lunch party but I won't do it if you're going out.

How to characterize the relevance of such responses? The data we have just seen in (25)–(28) indicates that the search space for potential clarification questions is small. We will suggest that this can be modelled in terms of a small number of schemas of the form: *if  $u$  is an utterance and  $u0$  is a constituent of  $u$ , add the clarification question  $CQ(u0)$  into  $QUD$ .* To understand why, we first need to consider how utterances are integrated into the DGB.

In terms of the Dialogue GameBoard the issue can be formulated as follows—what information needs to be associated with **Pending** to enable the formulation of grounding conditions/CR potential? The requisite information needs to be such that it enables the original speaker to interpret and recognize the coherence of the range of possible clarification queries that the original addressee might make.

Ginzburg (2012) offers detailed arguments on this issue, including considerations of the phonological/syntactic parallelism exhibited between CRs and their antecedents and the existence of CRs whose function is to request repetition of (parts of) an utterance, see (25) above. Taken together with the obvious need for **Pending** to include values for the contextual parameters specified by the utterance type, Ginzburg concludes that the type of **Pending** combines tokens of the utterance, its parts, and of the constituents of the content with the utterance type associated with the utterance. An entity that fits this specification is the *locutionary proposition* defined by the utterance: in the immediate aftermath of a speech event  $u$ , **Pending** gets updated with a record of the form of (29a) of type *locutionary proposition* (LocProp). Here  $T_u$  is a grammatical type for classifying  $u$

---

<sup>16</sup>Examples (25)–(28) are taken from the British National Corpus.

that emerges during the process of parsing  $u$ . In the most general case, given the need to accommodate structural ambiguity, it should be thought of as a *chart* (Cooper, 2012), but in the cases we consider here it can be identified with a *sign* in the sense of Head Driven Phrase Structure Grammar (HPSG). The relationship between  $u$  and  $T_u$ —describable in terms of the proposition  $p_u$  given in (29b)—can be utilized in providing an analysis of grounding/CRification conditions, as shown in (30):<sup>17</sup>

- (29) a.  $\text{LocProp} = \left[ \begin{array}{l} \text{sit} = u \\ \text{sit-type} = T_u \end{array} \right]$   
 b.  $p_u = \left[ \begin{array}{l} \text{sit} = u \\ \text{sit-type} = T_u \end{array} \right]$
- (30) a. Grounding:  $p_u$  is true: the utterance type fully classifies the utterance token.  
 b. CRification:  $p_u$  is false, either because  $T_u$  is weak (e.g. incomplete word recognition) or because  $u$  is incompletely specified (e.g. incomplete contextual resolution—problems with reference resolution or sense disambiguation).

It is useful to conceive of the integration of an utterance in an information state as a potentially cyclic process. Instantiation of some perhaps all contextual parameters will occur as soon as an utterance has taken place, assuming  $T_u$  is uniquely specified; if this is not the case, then CRification can occur on that level. Parameter instantiation can also take place subsequently, as when more information is provided as a consequence of CRification. Given this, utterance integration can be broken into three components:

1. Pending update: in the immediate aftermath of a speech event  $u$ , **Pending** gets updated with a record of the form  $\left[ \begin{array}{l} \text{sit} = u \\ \text{sit-type} = T_u \end{array} \right]$
2. Contextual extension: If  $T_u$  is uniquely specified, try to instantiate the contextual parameters of  $T_u$  relative to the context provided by the DGB: find a record  $w$  that extends  $u$  and such that  $w$  contains a subrecord of the dgb-param anchoring intended by  $u$ 's speaker; integrate  $w$  into *MaxPending*:  $\text{MaxPending} := \left[ \begin{array}{ll} \text{sit} & = w \\ \text{sit-type} & = T_u \end{array} \right]$
3. Move update/Pending downdate: if *MaxPending* is true, update Moves, so that  $\text{LatestMove} := \text{MaxPending}$ , downdate *MaxPending* from **Pending**.

We exemplify this series of contextual updates in (31): (31a) exemplifies an utterance type akin to an HPSG *sign*; we subsequently call this type ‘IGH’. (31b) exemplifies a locutionary proposition whose situational component is  $u_0$  (with four sub-utterances  $u_{is}, u_{Georges}, u_{here}, u_{is\ georges\ here}$ ) and whose type component is IGH. (31c) exemplifies a DGB in the immediate aftermath of an utterance classified by the type IGH; we note for future reference also certain utterance-related presuppositions that must be in place—the fact that  $u_0$  is the most recent utterance and the existence of appropriate witnesses for the contextual parameters ‘l’ and ‘g’, corresponding to the subutterances ‘here’ and ‘Georges’. (31d) exemplifies a witness for the contextual parameters of IGH and (31f) the evolution of the DGB after using the rule of Contextual extension with the witness  $w_0$ :

---

<sup>17</sup>A particularly detailed theory of grounding has been developed in the PTT framework, e.g. Poesio and Traum (1997) and Poesio and Rieser (2010).

(31) a. IGH = 
$$\left[ \begin{array}{l} \text{PHON : is georges here} \\ \text{CAT = V[+fin] : syncat} \\ \text{constits = \{ is, georges, here, is georges here \}: set(sign)} \\ \\ \text{DGB-PARAMS : } \left[ \begin{array}{l} \text{spkr: IND} \\ \text{addr: IND} \\ \text{s0: SIT} \\ \text{l: LOC} \\ \text{g: IND} \end{array} \right] \\ \\ \text{cont = Ask(spkr,addr, ?[sit-type = In(l,g)]) : IllocProp} \end{array} \right]$$

b. 
$$\left[ \begin{array}{l} \text{sit = u0 = } \left[ \begin{array}{l} \text{phon = izjorjhia} \\ \text{cat = V[+fin,+root]} \\ \text{constits = \{ u_{is}, u_{Georges}, u_{here}, u_{is georges here} \}} \\ \\ \text{dgb-params = } \left[ \begin{array}{l} \text{s0 = sit0} \\ \text{spkr = A} \\ \text{addr = B} \\ \text{l = loc0} \\ \text{g = g0} \end{array} \right] \\ \\ \text{cont = ?} \left[ \begin{array}{l} \text{sit = s0} \\ \text{sit-type = Present(g,l)} \end{array} \right] \end{array} \right] \\ \\ \text{sit-type = } \left[ \begin{array}{l} \text{PHON : is georges here} \\ \text{CAT = V[+fin] : syncat} \\ \text{constits = \{ is, georges, here, is georges here \}: set(sign)} \\ \\ \text{C-PARAMS : } \left[ \begin{array}{l} \text{spkr: IND} \\ \text{addr: IND} \\ \text{s0: SIT} \\ \text{l: LOC} \\ \text{g: IND} \end{array} \right] \\ \\ \text{cont = Ask(spkr,addr, ?} \left[ \begin{array}{l} \text{sit = s0} \\ \text{sit-type = In(l,g)} \end{array} \right] \text{) : IllocProp} \end{array} \right] \end{array} \right]$$

c. 
$$\text{dgb0 = } \left[ \begin{array}{l} \text{spkr} = \text{A} \\ \text{addr} = \text{B} \\ \\ \text{pending} = \left\langle \left[ \begin{array}{l} \text{sit} = \text{u0} \\ \text{sit-type} = \text{IGH} \end{array} \right] \right\rangle \\ \\ \text{qud} = \{ \} \\ \\ \text{facts} = \left\{ \text{In}(1, \{ \text{A}, \text{B} \}), \text{Named}('Georges', \text{g}), \right. \\ \left. \text{MostRecentSpeechEvent}(\text{u0}), \dots \right\} \\ \\ \text{moves} = \langle \rangle \end{array} \right]$$

$$\begin{array}{l}
\text{d. } v0 = \left[ \begin{array}{l} \text{spkr} = A0 \\ \text{addr} = B0 \\ \text{utt-time} = t0 \\ \text{s0} = \text{sit1} \\ \text{l} = l0 \\ \text{g} = g0 \\ \text{c3} = \text{pr1} \end{array} \right] \\
\text{e. } w0 = v0 \cup u0 \\
\text{f. } \text{dgb1} = \left[ \begin{array}{l} \text{spkr} = A \\ \text{addr} = B \\ \text{pending} = \left\langle \left[ \begin{array}{l} \text{sit} = w0 \\ \text{sit-type} = \text{IGH} \end{array} \right] \right\rangle \\ \text{qud} = \text{dgb1.qud} \\ \text{facts} = \text{dgb1.facts} \\ \text{moves} = \text{dgb1.moves} \end{array} \right]
\end{array}$$

We concentrate here on characterizing the range of possible CRs, specifically *intended content* CRs (27); analogous remarks apply to other types of CRs. The non-sentential CRs in (32a) and (32b) are interpretable as in the parenthesized readings. This provides justification for the assumption that the context that emerges in clarification interaction involves the accommodation of an issue—one that for A’s utterance in (32), assuming the sub-utterance ‘Bo’ is at issue, could be paraphrased as (32c). The accommodation of this issue into QUD could be taken to license any utterances that are co-propositional with this issue, where *CoPropositionality* is the relation between utterances defined in (33). This will also allow as relevant responses corrections, as in (32d):

- (32) A: Is Bo leaving?  
a. B: Bo? (= Who do you mean ‘Bo’?)  
b. B: Who? (= Who do you mean ‘Bo’?)  
c. *Who do you mean ‘Bo’?*  
d. B: You mean Mo.

- (33) *CoPropositionality*  
a. Two utterances  $u_0$  and  $u_1$  are *co-propositional* iff the questions  $q_0$  and  $q_1$  they contribute to QUD are co-propositional.  
(i)  $\text{qud-contrib}(m0.\text{cont})$  is  $m0.\text{cont}$  if  $m0.\text{cont} : \text{Question}$   
(ii)  $\text{qud-contrib}(m0.\text{cont})$  is  $?m0.\text{cont}$  if  $m0.\text{cont} : \text{Prop}$ <sup>18</sup>  
b.  $q_0$  and  $q_1$  are co-propositional if there exists a record  $r$  such that  $q_0(r) = q_1(r)$ .

*CoPropositionality* for two questions means that, modulo their domain, the questions involve similar answers. For instance ‘Whether Bo left’, ‘Who left’, and ‘Which student left’ (assuming Bo is a student) are all co-propositional. In the current context co-propositionality amounts to: either a CR which differs from MaxQud at most in terms of its domain, or a correction—a proposition that instantiates MaxQud.

We also note one fairly minor technical modification to the DGB field QUD, motivated in detail in Fernández (2006, Ginzburg (2012), assuming one wishes to exploit QUD to specify the resolution of non-sentential utterances such as short answers, sluicing, and various other fragments. QUD tracks not simply questions qua semantic objects, but pairs of

<sup>18</sup>Recall from the assertion protocol that asserting  $p$  introduces  $p?$  into QUD.

entities: a question and an antecedent sub-utterance. This latter entity provides a partial specification of the focal (sub)utterance, and hence it is dubbed the *focus establishing constituent* (FEC) (cf. *parallel element* in higher order unification–based approaches to ellipsis resolution e.g. Gardent and Kohlhase (1997).) Thus, the FEC in the QUD associated with a wh-query will be the wh-phrase utterance, the FEC in the QUD emerging from a quantificational utterance will be the QNP utterance, whereas the FEC in a QUD accommodated in a clarification context will be the sub-utterance under clarification. Hence the type of QUD is *InfoStruc*, as defined in (34):<sup>19</sup>

$$(34) \quad \text{Info-struct} = \left[ \begin{array}{l} \text{q : Questn} \\ \text{fec : set(LocProp)} \end{array} \right]$$

Repetition and meaning-oriented CRs can be specified by means of a uniform class of conversational rules, dubbed *Clarification Context Update Rules* (CCURs) in Ginzburg (2012). Each CCUR specifies an accommodated MaxQud built up from a sub-utterance  $u1$  of the target utterance, the maximal element of Pending (*MaxPending*). Common to all CCURs is a license to follow up *MaxPending* with an utterance which is *co-propositional* with MaxQud. (35) is a simplified formulation of one CCUR, *Parameter identification*, which allows B to raise the issue about A’s sub-utterance  $u0$ : *what did A mean by  $u0$ ?*

$$(35) \quad \text{Parameter identification:} \left[ \begin{array}{l} \text{pre} : \left[ \begin{array}{l} \text{Spkr : Ind} \\ \text{MaxPending : LocProp} \\ \text{u0} \in \text{MaxPending.sit.constits} \end{array} \right] \\ \text{effects} : \left[ \begin{array}{l} \text{MaxQud} = \left[ \begin{array}{l} \text{q} = \lambda x \text{Mean}(A, u0, x) \\ \text{fec} = u0 \end{array} \right] : \text{InfoStruc} \\ \text{LatestMove : LocProp} \\ \text{c1: CoProp(LatestMove.cont, MaxQud.q)} \end{array} \right] \end{array} \right]$$

Parameter Identification (35) underpins CRs such as (36b)–(36c) as follow-ups to (36a). We can also deal with corrections, as in (36d). B’s corrective utterance is co-propositional with  $\lambda x \text{Mean}(A, u0, x)$ , and hence allowed by the specification:

- (36) a. A: Is Bo here?  
 b. B: Who do you mean ‘Bo’?  
 c. B: Bo? (= Who is ‘Bo’?)  
 d. B: You mean Jo.

The examples in (37) exemplify the MaxQud.q specification of other CCURs:

- (37) a. **Parameter focussing:** raises as MaxQud.q  
 $\lambda x \text{MaxPending.content}(u1.\text{content} \mapsto x)$   
 b. **Utterance repetition:** raises as MaxQud.q  
 $\lambda x \text{Utter}(A, u1, x)$  (*What did A utter in  $u1$ ?* “What did you say?”)  
 c. **Utterance prediction:** raises as MaxQud.q  
 $\lambda x \text{UtterAfter}(A, u1, x)$  (*What will A utter after  $u1$ ?* “What were you going to say?”)<sup>20</sup>

<sup>19</sup>In the case of singleton values for the FEC we will typically abuse notation and identify the set by its single member.

<sup>20</sup>This is modelled after the proposal of Purver (2004) for analyzing cases such as (i), which he calls *fillers*:  
 (i)A: Are you ... B: angry? (=Did you mean to say ‘angry’ after ‘you’?)

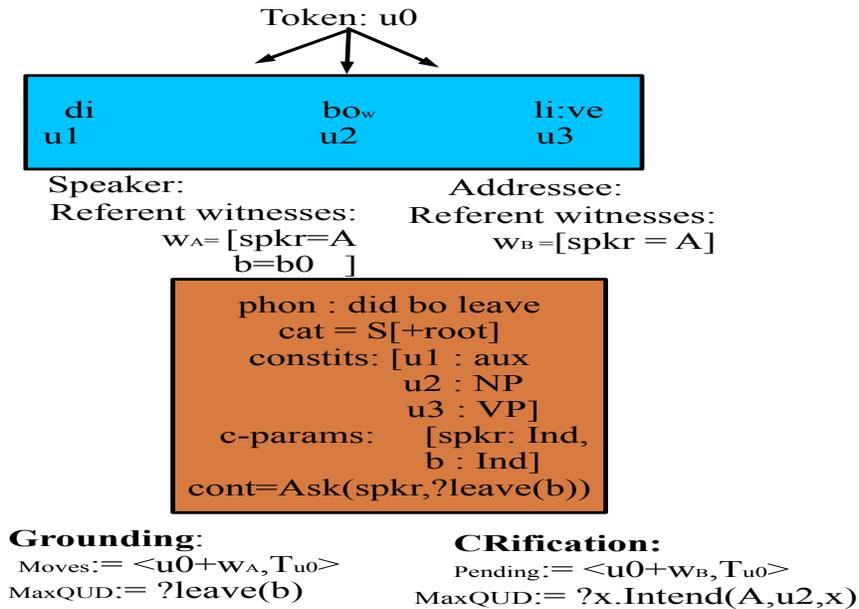


Figure 3: A single utterance gives rise to distinct updates of the DGB for distinct participants

To exemplify our account of how CRs get integrated in context, we exemplify in Figure 3 how the same input leads to distinct outputs on the “public level” of information states. In this case this arises due to differential ability to anchor the contextual parameters. The utterance  $u_0$  has three sub-utterances,  $u_1$ ,  $u_2$ ,  $u_3$ , given in Figure 3 with their approximate pronunciations. A can ground her own utterance since she knows the values of the contextual parameters, which we assume here for simplicity include the speaker and the referent of the sub-utterance ‘Bo’. This means that the locutionary proposition associated with  $u_0$ —the proposition whose situational value is a record that arises by unioning  $u_0$  with the witnesses for the contextual parameters and whose type is given in Figure 3—is true. This enables the “canonical” illocutionary update to be performed: the issue ‘whether b left’ becomes the maximal element of QUD. In contrast, let assume that B lacks a witness for the referent of ‘Bo’. As a result, the locutionary proposition associated with  $u_0$  which B can construct is not true. Given this, B uses the CCUR parameter identification to build a context appropriate for a clarification request: B increments QUD with the issue  $\lambda x \text{Mean}(A, u_2, x)$ , and the locutionary proposition associated with  $u_0$  which B has constructed remains in Pending.

The final generalizations we need to make are along two dimensions. First, whereas for semantically based CRification, it is sufficient to think about updates to MaxPending as resulting from an extension of (records that) witness contextual parameters, for repetition CRs we also need to allow for change on the utterance type dimension. So we generalize Contextual extension to Pending extension, formulated as follows:

(38) Pending extension:

- a. if  $\text{MaxPending} = \begin{bmatrix} \text{sit} = u \\ \text{sit-type} = T_u \end{bmatrix}$  and  $p_w = \begin{bmatrix} \text{sit} = w \\ \text{sit-type} = T_w \end{bmatrix}$  extends  $p_u$  and reflects  $u$ 's speaker's intention, then update *MaxPending*:  $\text{MaxPending} := \begin{bmatrix} \text{sit} = w \\ \text{sit-type} = T_w \end{bmatrix}$
- b.  $\text{PropExtension}(p1 = \begin{bmatrix} \text{sit} = u \\ \text{sit-type} = T_u \end{bmatrix}, p2 = \begin{bmatrix} \text{sit} = v \\ \text{sit-type} = T_v \end{bmatrix})$  iff  $p1, p2$ : Prop and (a) for all fields either  $u.f = v.f$  or  $u.f \sqsubset v.f$  and (b)  $T_v \sqsubset T_u$

So far, the only non-grounding action we have considered is clarification interaction, in which there is a missing witness for a contextual parameter or phonological type. This triggers a query for that information and a unification of the required information into the representation of the utterance. (Metacommunicative) corrections are a variant on this theme: instead of a *missing* witness, they involve (pointing out) an *incorrect* witness, which needs to be replaced by the correct value. As we pointed out above, we have an account for the *coherence* of content-oriented corrections (see (23)) and metacommunicative ones (see (36d)); what remains to specify for the latter is the effect on the DGB.

One possible means for unifying the update and downdate/replacement associated with clarification interaction and corrections, respectively, would be to use an operation such as asymmetric unification in which later information takes precedence. Such a logical operation, named *priority union*, is specified in Grover, Brew, Manandhar, Moens, and Schoter (1994), who exemplify a number of its uses. Given the complexity of this operation, however, we postulate an additional update operation, which effects replacements of the desired kind:

(39) Pending replacement:

- a. if  $\text{MaxPending} = \begin{bmatrix} \text{sit} = u \\ \text{sit-type} = T_u \end{bmatrix}$  and  $p_w = \begin{bmatrix} \text{sit} = w \\ \text{sit-type} = T_w \end{bmatrix}$  is a substitution instance of  $p_u$  and reflects  $u$ 's speaker's intention, then update *MaxPending*:  $\text{MaxPending} := \begin{bmatrix} \text{sit} = w \\ \text{sit-type} = T_w \end{bmatrix}$
- b.  $\text{SubstInst}(p1 = \begin{bmatrix} \text{sit} = u \\ \text{sit-type} = T_u \end{bmatrix}, p2 = \begin{bmatrix} \text{sit} = v \\ \text{sit-type} = T_v \end{bmatrix})$  iff  $p1, p2$ : Prop and (a) for all fields either  $u.f = v.f$  or for some  $T$ ,  $u.f : T$  and  $v.f : T$

To exemplify this, we consider the cross-turn self-correction example in ((40)). A utters 'Is Georges here?'. parameter identification licenses the accommodation of 'What did A mean by uttering 'Georges?'' as MaxQUD, which in turn licenses 'I meant Jacques' as an utterance co-propositional with MaxQUD. Subsequent to this Pending Replacement applies:

(40) a. A: Is [<sub>u0</sub>Georges] here? I meant Jacques.

In more detail: after the utterance of 'Is Georges here', A's FACTS will include the presuppositions that the most recent speech event is  $u_0$  ('Is Georges here'), which includes as sub-utterance  $u_{\text{georges}}$ , and that  $u_0$  is classified by the type IGH; the DGB is essentially the following:

$$(41) \quad \text{a.} \quad \text{A.dgb1} = \left[ \begin{array}{l} \text{spkr} = \text{A} \\ \text{addr} = \text{B} \\ \text{pending} = \text{p0} = \left\langle \left[ \begin{array}{l} \text{sit} = \text{u0} \\ \text{sit-type} = \text{IGH} \end{array} \right] \right\rangle \\ \text{qud} = \langle \rangle \\ \text{facts} = \left\{ \begin{array}{l} \text{In}(1, \{ \text{A}, \text{B} \}), \text{Named}(\text{'Georges'}, \text{g}), \\ \text{MostRecentSpeechEvent}(\text{u0}), \\ \text{Classify}(\text{IGH}, \text{u0}) \dots \end{array} \right\} \\ \text{moves} = \langle \rangle \end{array} \right]$$

This allows for **parameter identification** to be used—the issue ‘What did A mean by  $u_{\text{georges}}$ ’ becomes MaxQUD with ‘Jacques’ as FEC. This licences as LatestMove ‘I meant Jacques’, which in turn leads to an update of QUD:

$$(42) \quad \text{A.dgb2} = \left[ \begin{array}{l} \text{spkr} = \text{A} \\ \text{addr} = \text{B} \\ \text{pending} = \left\langle \left[ \begin{array}{l} \text{sit} = \text{u0} \\ \text{sit-type} = \text{IGH} \end{array} \right] \right\rangle \\ \text{qud} = \left\langle \left[ \begin{array}{l} \text{q} = \text{?Mean}(\text{A}, u_{\text{georges}}, \text{jacques}'), \text{fec} = \{ \} \end{array} \right], \left[ \begin{array}{l} \text{q} = \lambda x \text{Mean}(\text{A}, \text{u0}, x), \text{fec} = \text{'Jacques'} \end{array} \right] \right\rangle \\ \text{facts} = \left\{ \begin{array}{l} \text{Named}(\text{'Georges'}, \text{georges}'), \text{Named}(\text{'Jacques'}, \text{jacques}'), \\ \text{2ndMostRecentSpeechEvent}(\text{u0}), \\ \text{Classify}(\text{IGH}, \text{u0}), \\ \text{MostRecentSpeechEvent}(\text{u1}), \\ \text{Classify}(\text{'I meant Jacques'}, \text{u1}) \dots \end{array} \right\} \\ \text{moves} = \langle \text{Assert}(\text{A}, \text{Mean}(\text{A}, u_{\text{georges}}, \text{jacques}')) \rangle \end{array} \right]$$

Accepting this gives rise to an application of **pending replacement**, which modifies the original locutionary proposition:  $u0$  is modified to a record  $v0$  with the referent *jacques*’ replacing *georges*’ and the utterance type is now IJH (‘Is Jacques here?’) whose *phon* includes the form **jacques**; MAXPENDING is modified accordingly:

$$(43) \quad \text{a.} \quad \text{A.dgb3} = \left[ \begin{array}{l} \text{spkr} = \text{A} \\ \text{addr} = \text{B} \\ \text{pending} = \left\langle \left[ \begin{array}{l} \text{sit} = \text{v0} \\ \text{sit-type} = \text{IJH} \end{array} \right] \right\rangle \\ \text{qud} = \langle \rangle \\ \text{facts} = \left\{ \begin{array}{l} \text{2ndMostRecentSpeechEvent}(\text{u0}), \\ \text{Classify}(\text{IGH}, \text{u0}), \\ \text{MostRecentSpeechEvent}(\text{u1}), \\ \text{Classify}(\text{'I meant Jacques'}, \text{u1}), \text{Named}(\text{'Jacques'}, \text{jacques}), \end{array} \right\} \\ \text{moves} = \langle \text{Assert}(\text{A}, \text{Mean}(\text{A}, \text{u0}, \text{jacques}')) \rangle \end{array} \right]$$

As can be readily observed, the utterance  $u_0$  is still a component of facts in FACTS, and hence also its sub-utterance  $u_{\text{georges}}$ . Neither utterance is a component of PENDING, whose content will be subject to uptake in the next utterance. Given that they are in FACTS, referential possibilities to those two utterances (‘Is Georges here’ and ‘Georges’)—and to the referent of ‘Georges’ are not eliminated.

## 4 From CRs to Dysfluency: Informal Sketch

The approach described above for CRs and self/other-corrections at a cross-turn level extends relatively seamlessly to self-corrections, hesitations, and other types of intra-turn dysfluencies. Before going into the technical details, we sketch the account at an informal level, indicating some of its main consequences.

As we pointed out above, the main idea underlying KOS’ theory of CRs is that in the aftermath of an utterance  $u$  a variety of questions concerning  $u$  and definable from  $u$  and its grammatical type become available to the addressee of the utterance. These questions regulate the subject matter and ellipsis potential of CRs concerning  $u$  and generally have a short lifespan in context.

We propose that a very similar account applies to dysfluencies. As the utterance unfolds incrementally there arise questions about what has happened so far (e.g. *what did the speaker mean with sub-utterance  $u_1$ ?*) or what is still to come (e.g. *what word does the speaker mean to utter after sub-utterance  $u_2$ ?*). Or slightly more technically, we suggest that incrementally certain *utterance monitoring* and *utterance planning* questions can be pushed on to QUD.

By making this assumption we obtain a number of pleasing consequences. We can:

1. **explain similarities to other-corrections:** the same mechanism is at work, differentiated only by the QUDs that get accommodated.
2. **explain how the other can take over & do the second part of the dysfluency:** if ‘what did I want to say’ / ‘what do I want to say next’ is indeed a question under discussion, then it should in principle also be possible for the interpreter to address that.
3. **explain how inferences can be drawn from the dysfluency:** Once the question ‘what do I want to say next’ has been pushed on the QUD, the listener can ask ‘why did he raise that question?’, just like he can do with any other question that someone raises. And often a good answer is ‘because he really doesn’t know, and a good reason for that could be that it is indeed difficult to know that, which makes sense for this thing here which doesn’t really have a good name, as opposed to that thing over there, which can be named easily’. This would actually also explain the finding of Arnold et al. (2007), namely that if you explain to subjects that the speaker has a pathology that makes it hard for them to remember names for things, the inference that ‘uh oh’ means that they are trying to describe the thing that is hard to describe goes (largely) away (see Section 2.2.2). In our approach, this would then just not be a good answer anymore to the question ‘why did he raise that question’.
4. **explain internal coherence of dysfluencies:** ‘#I was a little bit + swimming’ is an odd dysfluency, it can never mean ‘I was swimming’ in the way that ‘I was a little bit + actually, quite a bit shocked by that’ means ‘I was quite a bit shocked by that’. Why? Because ‘swimming’ is not a good answer to ‘What did I mean to say when I said ‘a little bit’?’.
5. **explain why appropriateness changes implicate that original use unreasonable:** examples like (44) involve quantity implicatures. These can be explicated based on reasoning such as the following: *I could have said (reperandum), but on*

*reflection I said (alteration), which differs only in filtering away the requisite entailment.*

- (44) a. it's basically (the f- + a front) leg [implicature: no unique front leg]  
b. Ehm . imagine that's like (the + a) leg . [implicature: no unique leg]

## 5 Dysfluency Rules

### 5.1 An Incremental Perspective

As we have seen, there are quite a number of benefits that arrive by integrating CRs and dysfluencies within one explanatory framework. Still, attractive as it might be, there is some technical work to be done.

In fact, the only modification we make is to extend Pending to incorporate utterances that are *in progress*, and hence, incompletely specified semantically and phonologically. This presupposes the use of a grammar which can associate syntactic types and contents on a word by word basis. For dialogue this is a move that has extensive motivation (for a review see e.g. Rieser and Schlangen (2011a) and for detailed evidence the papers in Rieser and Schlangen (2011b).). There is by now a long tradition within certain grammatical frameworks of specifying grammars to ensure incremental processing, emanating from Categorical Grammar, Lexicalized Tree Adjoining Grammar, and various subsequent frameworks such as Dynamic Dependency Grammar (Milward, 1994), and Dynamic Syntax (Kempson et al., 2000). From a semantic point of view, as emphasized by Milward (1994), one of the main requirements is that ‘a non-trivial semantic representation is built word by word ... What constitutes a non-trivial representation is debatable. The position taken here is that it must use all the information given so far. Thus, an acceptable representation for the sentence fragment ‘John likes’ would be  $\lambda x \text{like}(\text{john}', x)$ , but not a semantic product such as  $\text{john}' * \lambda(x, y)$ .’

Specifying a grammatical framework of the required kind constitutes a paper in its own right. Nonetheless, the closest in spirit is recent work on incremental semantic construction for dialogue by Peldszus, Buß, Baumann, and Schlangen (2012) and Peldszus and Schlangen (2012), based on the framework of Robust Minimal Recursion Semantics (RMRS) (Copestake, 2007), which enables predicate–argument structure to be under-specified. Peldszus and Schlangen formulate and implement an algorithm for interpreting an incrementally provided syntactic representation in a top–down left–to–right fashion. They argue for this strategy (as opposed to e.g. a bottom up one) as it provides monotonic semantic interpretation that gets further specified as each word gets encountered. Concretely for us, this means that the elements of CONSTITS, the potential objects of repair, have their syntactic and semantic classifications constructed monotonically, as long as no repair act occurs.

Here we illustrate their account with one of their examples reformulated using TTR, simplifying and modifying it in various respects, in particular abstracting away from one of their main contributions—the semantic combinatorics.<sup>21</sup> In the example that follows (Figure 4) semantic material added by a given word after the initial word is in bold face. The imperative verb ‘take’ introduces both illocutionary force and a predicate with two roles, one of which is identified with the addressee; the demonstrative determiner introduces a contextual parameter which is identified with the role of object taken (the label  $y$ ); ‘book’ introduces a restriction on that contextual parameter; ‘in’ introduces a descriptive predicate with two roles, one of which is identified with  $y$ .

---

<sup>21</sup>For another account which proposes the use of TTR in incremental processing see Purver, Eshghi, and Hough (2011).

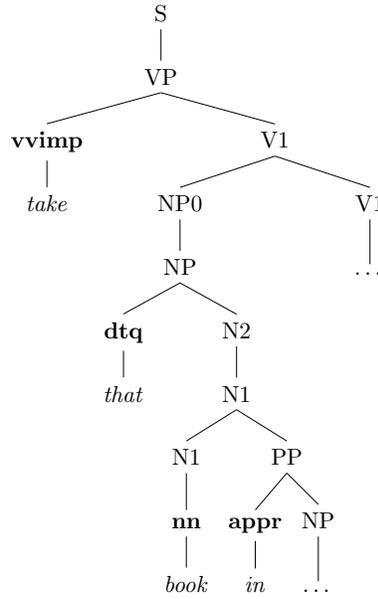


Figure 4: Incremental syntactic derivation of a simple example sentence. (Peldszus and Schlangen’s Figure 2)

(45) a. Take ...

$$\left[ \begin{array}{l} \text{dgb-params : } \left[ \begin{array}{l} A : \text{Ind} \\ B : \text{Ind} \\ c0 : \text{addr}(A,B) \\ s0 : \text{Rec} \end{array} \right] \\ \text{cont} = \left[ \begin{array}{l} \text{sit} = s0 \\ \text{sit-type} = \left[ \begin{array}{l} y : \text{Ind} \\ x = B : \text{Ind} \\ c1 : \text{Order}(A,B,\text{Take}(x,y)) \end{array} \right] \end{array} \right] : \text{Prop} \end{array} \right]$$

b. Take that ...

$$\left[ \begin{array}{l} \text{dgb-params : } \left[ \begin{array}{l} A : \text{Ind} \\ B : \text{Ind} \\ c0 : \text{addr}(A,B) \\ s0 : \text{Rec} \\ \mathbf{d} : \text{Ind} \end{array} \right] \\ \text{cont} = \left[ \begin{array}{l} \text{sit} = s0 \\ \text{sit-type} = \left[ \begin{array}{l} \mathbf{y} = \mathbf{d} : \text{Ind} \\ x = B : \text{Ind} \\ c1 : \text{Order}(A,B,\text{Take}(x,y)) \end{array} \right] \end{array} \right] : \text{Prop} \end{array} \right]$$

c. Take that book ...

$$\left[ \begin{array}{l} \text{dgb-params : } \left[ \begin{array}{l} A : \text{Ind} \\ B : \text{Ind} \\ c0 : \text{addr}(A,B) \\ s0 : \text{Rec} \\ d : \text{Ind} \\ \mathbf{c2} : \text{book}(d) \end{array} \right] \\ \text{cont} = \left[ \begin{array}{l} \text{sit} = s0 \\ \text{sit-type} = \left[ \begin{array}{l} y = d : \text{Ind} \\ x = B : \text{Ind} \\ c1 : \text{Order}(A,B,\text{Take}(x,y)) \end{array} \right] \end{array} \right] : \text{Prop} \end{array} \right]$$

$$\text{d. Take that book in ...} \left[ \begin{array}{l} \text{dgb-params :} \left[ \begin{array}{l} A : \text{Ind} \\ B : \text{Ind} \\ c0 : \text{addr}(A,B) \\ s0 : \text{Rec} \\ d : \text{Ind} \\ c2 : \text{book}(d) \end{array} \right] \\ \\ \text{cont =} \left[ \begin{array}{l} \text{sit} = s0 \\ \text{sit-type} = \left[ \begin{array}{l} y = d : \text{Ind} \\ x = B : \text{Ind} \\ \mathbf{z} : \mathbf{Ind} \\ \mathbf{v} = \mathbf{y} : \mathbf{Ind} \\ \mathbf{c3} : \mathbf{In}(y,z) \\ c1 : \text{Order}(A,B,\text{Take}(x,y)) \end{array} \right] \end{array} \right] : \text{Prop} \end{array} \right]$$

For our current purposes, the decisions we need to make can be stated independently of the specific grammatical formalism used. The main assumptions we are forced to make concern **Pending instantiation** and **contextual instantiation** and more generally, the testing of the fit between the speech events and the types assigned to them. We assume that this takes place incrementally. For concreteness we will assume further that this takes place word by word, though examples like (46), which demonstrate the existence of word-internal monitoring, show that this is occasionally an overly strong assumption.

(46) Looking at the *tex*-technical functions. (From Besser and Alexandersson (2007).)

## 5.2 Backward Looking Dysfluencies

Our analysis now distinguishes between backward-looking dysfluencies (BLDs) and forward-looking dysfluencies (FLDs). BLDs we assume are possible essentially at any point where there is ‘correctable material’. Technically this amounts to Pending not being empty. We assume that editing phrases are, at least in some cases, content-ful constituents of the repair. This is implemented by the rule in (47) **Backwards Looking Appropriateness Repair**. Given that  $u0$  is a constituent in MaxPending, it is possible to accommodate as MaxQud the following InfoStruc: the issue is ‘what did A mean by  $u0$ ’, whereas the FEC is  $u0$ ; this specifies that the follow up utterance needs to be co-propositional with MaxQud.

$$(47) \text{ Backwards Looking Appropriateness Repair:} \left[ \begin{array}{l} \text{pre :} \left[ \begin{array}{l} \text{spkr} : \text{Ind} \\ \text{addr} : \text{Ind} \\ \text{pending} = \langle p0, \text{rest} \rangle : \text{list}(\text{LocProp}) \\ u0 : \text{LocProp} \\ c1 : \text{member}(u0, p0.\text{sit}.\text{constits}) \end{array} \right] \\ \\ \text{effects :} \text{TurnUnderspec} \wedge_{\text{merge}} \left[ \begin{array}{l} \text{MaxQud} = \\ \left[ \begin{array}{l} q = \lambda x \text{ Mean}(\text{pre}.\text{spkr}, \text{pre}.\text{u0}, x) \\ \text{fec} = u0 \end{array} \right] : \text{InfoStruc} \\ \text{LatestMove} : \text{LocProp} \\ c2 : \text{CoPropositional}(\text{LatestMove}^{\text{content}}, \text{MaxQud}) \end{array} \right] \end{array} \right]$$

In short, this rule, which is equivalent to **Parameter Identification** (35)—apart from underspecifying the turn holder, allows us to analyse the alteration (and the editing terms, if present) of a BLD as providing an answer to an issue that has been accommodated as MaxQud and whose FEC corresponds to the reparandum of the dysfluency. Since the rule leaves the next turn-taker underspecified, it can also deal with other-corrections and content CRs, such as those in (36b)-(36d).

To make all this clearer, we consider an example in detail. We emphasize that this treatment is almost identical to example (40) we discussed in section 3.2; the sole difference here is that the self-correction occurs mid-utterance and, hence, necessitates using an incremental content (the one from (45d)).

(48) Take that book in I mean from the shelf

A utters ‘Take that book in’. **Backwards Looking Appropriateness Repair** licenses the accommodation of ‘What did A mean by uttering ‘in’?’ as MaxQud, which in turn licenses ‘I meant from’ as an utterance co-propositional with MaxQud. Subsequent to this **Pending Replacement** applies and the utterance continues.

In detail: after the utterance of ‘Take that book in’, A’s FACTS will include the presuppositions that the most recent speech event is  $u_0$  (‘Take that book in’), which includes as sub-utterance  $u_{in}$ ; The DGB is essentially the one in (49):

$$(49) \quad A.dgb1 = \left[ \begin{array}{l} \text{spkr} = A \\ \text{addr} = B \\ \text{pending} = p_0 = \left\langle \left[ \begin{array}{l} \text{sit} = u_0 \\ \text{sit-type} = T_{\text{Take that book in...}} \end{array} \right] \right\rangle \\ \text{qud} = \langle \rangle \\ \text{facts} = \left\{ \begin{array}{l} \text{MostRecentSpeechEvent}(u_0), \\ \text{Classify}(T_{\text{Take that book in...}}, u_0) \end{array} \right\} \\ \text{moves} = \langle \rangle \end{array} \right]$$

$$(50) \quad T_{\text{Take that book in ...}} = \left[ \begin{array}{l} \text{phon} : \text{take that book in} \\ \text{cat} = v : \text{syncat} \\ \text{constits} = \left\{ \begin{array}{l} \text{Take, that, book, in, book in,} \\ \text{that book in} \end{array} \right\} : \text{set}(\text{sign}) \\ \text{dgb-params} : \left[ \begin{array}{l} A : \text{Ind} \\ B : \text{Ind} \\ c_0 : \text{addr}(A,B) \\ s_0 : \text{Rec} \\ d : \text{Ind} \\ c_2 : \text{book}(d) \end{array} \right] \\ \text{cont} = \left[ \begin{array}{l} \text{sit} = s_0 \\ \text{sit-type} = \left[ \begin{array}{l} y = d : \text{Ind} \\ x = B : \text{Ind} \\ z : \text{Ind} \\ v = y : \text{Ind} \\ c_2 : \text{In}(y,z) \\ c_1 : \text{Order}(A,B,\text{Take}(x,y)) \end{array} \right] : \text{Prop} \end{array} \right] \end{array} \right]$$

This allows for **Backwards Looking Appropriateness Repair** to be used. Its effects are shown in (51): the issue ‘What did A mean by  $u_{in}$ ’ becomes MaxQud, with the reparandum ‘in’ as FEC. This licences as LatestMove ‘I meant from’:

$$(51) \quad \text{A.dgb2} = \left[ \begin{array}{l} \text{spkr} = \text{A} \\ \text{addr} = \text{B} \\ \text{pending} = \left\langle \left[ \begin{array}{l} \text{sit} = \text{u0} \\ \text{sit-type} = \text{T}_{\text{Take that book in...}} \end{array} \right] \right\rangle \\ \text{MaxQud} = \left[ \begin{array}{l} \text{q} = \lambda x \text{ Mean}(\text{A}, \text{u}_{in}, x) \\ \text{fec} = \text{u}_{in} \end{array} \right] \\ \text{facts} = \left\{ \begin{array}{l} \text{2ndMostRecentSpeechEvent}(\text{u0}), \\ \text{Classify}(\text{T}_{\text{Take that book in}}, \text{u0}) \\ \text{MostRecentSpeechEvent}(\text{u1}), \\ \text{Classify}(\text{T}_{\text{I meant from}}, \text{u1}) \end{array} \right\} \\ \text{moves} = \left\langle \text{Assert}(\text{A}, \text{Mean}(\text{A}, \text{u}_{in}, \text{from})) \right\rangle \end{array} \right]$$

Accepting this gives rise to an application of **pending replacement**, which modifies the original locutionary proposition: *u0* is modified to a record *v0* with the relation *from* replacing *in* and the utterance type is now ‘Take that book from’ whose *phon* includes the form *from*; MAXPENDING is modified accordingly:

$$(52) \quad \text{A.dgb3} = \left[ \begin{array}{l} \text{spkr} = \text{A} \\ \text{addr} = \text{B} \\ \text{pending} = \left\langle \left[ \begin{array}{l} \text{sit} = \text{v0} \\ \text{sit-type} = \text{T}_{\text{Take that book from}} \end{array} \right] \right\rangle \\ \text{qud} = \langle \rangle \\ \text{facts} = \left\{ \begin{array}{l} \text{2ndMostRecentSpeechEvent}(\text{u0}), \\ \text{Classify}(\text{T}_{\text{Take that book in}}, \text{u0}) \\ \text{MostRecentSpeechEvent}(\text{u1}), \\ \text{Classify}(\text{T}_{\text{I meant from}}, \text{u1}) \dots \end{array} \right\} \\ \text{moves} = \left\langle \text{Assert}(\text{A}, \text{Mean}(\text{A}, \text{u}_{in}, \text{from})) \right\rangle \end{array} \right]$$

We now turn to a slightly different example that can be analysed in essentially the same way as (48). Whereas in (48) the editing terms ‘I mean’ plus the alteration ‘from the shelf’ form a canonical sentential structure, in (53) the alteration ‘headphones’ is non-sentential. We assume this non-sentential utterance is interpreted in precisely the same way as a short answer like (54) (see e.g. Ginzburg and Sag (2000, Fernández (2006), Ginzburg (2012)). After the application of **Backward Looking Appropriateness Repair**, the issue ‘What did A mean with the utterance ‘earphones’ with ‘earphones’ as FEC becomes QUD-maximal. This licenses the bare fragment ‘headphones’, which gets the reading ‘I mean headphones’.

(53) From BNC (file: KP0 369-370):  
Have you seen Mark’s erm earphones? Headphones.

(54) A: Who left? B: Bill.

This analysis would extend to the following example due to Levelt (1989), with MaxQud.q = ‘what did A mean by FEC?’ and the FEC = ‘to the right’ (the occurrence after ‘and’):

(55) To the right is yellow, and to the right – further to the right is blue.

Our analysis presupposes that the addressee is able to compute the QUD to be accommodated and its FEC once she has processed the reparandum on the basis of (syntactic) parallelism between reparandum and alteration. The rule-governed nature of this process has been argued for previously by (Levitt, 1989), who posited a well-formedness (coordination) rule which he argued dysfluencies need to observe;<sup>22</sup> see also (Hindle, 1983; Morrill, 2000). That this task facing the addressee is computable is clear given that one can automatically filter dysfluencies with rule-based dysfluency parsers that essentially rely on identifying (and removing) the reparandum (see e.g. Johnson and Charniak (2004) and Miller and Schuler (2008)).

### 5.3 Some More BLD Examples

We consider some more examples, which do not, we think, require any modification to our basic analysis, but point to some other interesting empirical issues. The first example we consider is (56). This differs from (48) in one significant way—a different editing phrase is used, namely ‘no’, which has distinct properties from ‘I mean’.

- (56) From Levitt (1989):  
From yellow down to brown - no - that’s red.

Whereas ‘I mean’ is naturally viewed as a syntactic constituent of the alteration, ‘no’ cannot be so analyzed. There are two obvious ways to analyze ‘no’’s role. The most parsimonious way would be to assimilate it to uses like (57), where the resolution is based on a contextually available polar question or proposition.<sup>23</sup>

- (57) a. A: Is Bill coming? B: No, Mary is.  
b. A: Bill is coming. B: No, Mary is.

In order to adopt such an analysis we would need to motivate the emergence of the requisite polar question or proposition, e.g. ‘Is u0 what I meant to say?’. And the most obvious way of doing that would be to postulate a variant of (47), where this was the MaxQud. There is nothing clearly wrong with such an approach, which would have the benefit of capturing the widespread use of negative discourse particles across languages for this function too. Nonetheless, apart from being somewhat *ad hoc*, this approach would also require some additional machinery to explain the coherence of the part of alteration following ‘no’. In the case of (57a), one can appeal to two explanations for why ‘Mary is’ is uttered: for some cases ‘Bill’ is accented and this justifies the independent assumption that the issue of ‘Who is coming’ is MaxQud; there are also (complementary) considerations of cooperativity relative to A’s original query. The former consideration does not apply in the case of (57b), whereas the latter does with cooperativeness being replaced by goal persistence—persisting in producing the utterance for whatever reason that motivated it in the first place.

An alternative analysis, which would avoid postulating an additional *conversational rule*, would involve instead positing an additional meaning for ‘no’, which is arguably needed for other uses such as:

- (58) a. [A opens freezer to discover smashed beer bottle] A: No! (‘I do not want *this* (the beer bottle smashing) to happen’)  
b. [Little Billie approaches socket holding nail] Parent: No Billie (‘I do not want *this* (Billie putting the nail in the socket) to happen’)

<sup>22</sup>Though see (Van Wijk & Kempen, 1987; Cori, De Fornel, & Marandin, 1997) for evidence that this rule can be overridden, as well as our own discussion of this issue below.

<sup>23</sup>Recall the conversational rules (22a) and (22c). These have the effect of introducing *p?* as MaxQud, both after a polar query *p?* and an assertion *p*. See Farkas and Bruce (2010) for a distinct, but related analysis.

This use of ‘no’ involves the expression of a negative attitude towards an event. A possible lexical entry for this use is given in (59), in which *sit1* is the contextual parameter for the undesired event:

$$(59) \quad \left[ \begin{array}{l} \text{phon} : \text{no} \\ \text{cat.head} = \text{adv}[+\text{ic}] : \text{syncat} \\ \text{dgb-params} = \left[ \begin{array}{l} \text{sit1} : \text{Rec} \\ \text{spkr} : \text{Ind} \end{array} \right] : \text{RecType} \\ \text{cont} = \neg \text{Want}(\text{spkr}, \text{sit1}) : \text{Prop} \end{array} \right]$$

This would, in particular, allow ‘no’ to be used to express a negative attitude towards an unintended utterance event. We could analyze (56) as involving the utterance ‘brown’. Following this, the rule (47) is triggered with the specification  $\text{MaxQud.q} = \text{what did A mean by FEC?}$  and the  $\text{FEC} = \text{‘brown.’}$  The analysis then proceeds like the earlier cases. Nonetheless, there is an additional issue which this case does bring out: the alternation (‘that’s red’) is sentential rather than directly parallel to the reparandum. This fits nicely with viewing the alteration as an answer to a question. It is indeed a counterexample to an overly syntactic view of self-correction, as embodied in Levelt’s rule. And this also means that the repaired utterance is not, in fact, a grammatical utterance if one filters away the reparandum (\*From yellow down to that’s red).<sup>24</sup> And, hence, just as with a clarification interaction case such as (60), one has to assume an additional inference process that leads from the provision of the answer to the triggering of pending replacement (pending extension in the case of (60)).<sup>25</sup>

(60) A: Is Jill coming? B: Jill? A: Surely you’re acquainted with my cousin. B: Right, no she’s not.

A similar analysis can be offered to a constructed example suggested to us by an anonymous reviewer, (61), which exemplifies an embedded correction.

(61) [<sub>u0</sub> Can you give me a flight [<sub>u1</sub> from Boston to New York.]] [<sub>u2</sub> No not from Boston, not to New York, but [<sub>u3</sub> from New York to Boston.]] [<sub>u4</sub> No I was right in the first place.]

Subsequent to the initial utterance *u0*, as a consequence of the use of ‘No’, a negative attitude is expressed toward *u0*, and the rule (47) is triggered with the specification  $\text{MaxQud.q} = q_{u1} = \text{what did A mean by FEC?}$  and the  $\text{FEC} = \text{‘from Boston to New York.’}$  *u2* is a (non sentential) utterance providing both a negative and positive answer concerning this question.<sup>26</sup> Subsequent to this utterance, another use of ‘No’ triggering the use of the rule (47) with the specification  $\text{MaxQud.q} = q_{u3} = \text{what did A mean by FEC?}$  and the  $\text{FEC} = \text{‘from New York to Boston.’}$  *u4* addresses *q<sub>u3</sub>*, while using

<sup>24</sup>Such cases—the breakdown of parallelism—are of course well known in ellipsis resolution; opinions vary as to what conclusions to draw from them.

<sup>25</sup>In fact, Levelt classifies examples such as (56) as *fresh starts*. But that seems to be just an easy way out for any exceptions to his rule—(56) is not a fresh start like, say, another of his examples “Straight to, or the entrance is brown.” where the reparandum is an initial segment of the utterance eliminated from subsequent processing of the utterance. And indeed calling (56) a fresh start doesn’t solve the problem because semantically we need to achieve the effect of modifying the utterance to one whose import is equivalent to “From yellow down to red.”. In fairness to Levelt, he is quite clear about the vagueness of the notion of “fresh start” (see p. 85, Levelt (1983)); nonetheless, the decision how to classify a repair is for him not of huge import since he is concerned with syntax and offers no formal semantic analysis.

<sup>26</sup>In this sense, this example is parallel to utterance (3) in example (23), which we discussed earlier.

the definite ‘the first place’ which can be understood as referring to  $u1$ .<sup>27</sup> This answer therefore resolves the issue  $q_{u3}$  and, therefore, also the issue  $q_{u1}$ : what did A mean by  $u1$ ? She was right (in saying)  $u1$ . This result will arise if subsequent to  $u0$ ,  $u0$  remains in Pending, unaltered until  $u2$  gets removed from Pending *after* the processing of  $u4$ . But what if A is hasty and immediately after  $u2$  applies Pending replacement yielding a modified  $u0'$  where  $u1$  has been replaced by  $u3$ ? Even if A adopted this strategy,  $u1$  taking place remains an element of FACTS, as discussed with respect to example (40), and hence a potential referent of ‘the first place’, in which case  $u0'$  could be modified back to  $u0$ .

We mention three more examples, given in (62).

- (62) a. We go straight on, or– we enter via red, then go straight on to green. ( From Levelt (1989))  
 b. The design of or– the point of putting two sensors on each side. (From Besser and Alexandersson (2007)).  
 c. Why it is – why is it that nobody makes a decent toilet seat? (From Fay (1980), cited by Levelt (1989))

Examples (62a)-(62b) are similar to (56) apart from the occurrence of the disjunction/discourse particle ‘or’. An analysis of such cases involves providing an analysis of ‘or’. We assume that these uses probably relate to other “corrective” uses of ‘or’, as in:

- (63) a. I’m going to be free. Or uh Bill is.  
 b. Who left yesterday? Or actually who left during the last week?

Whatever precise import we give to ‘or’, then we can analyze (62a) using the same analysis as was provided for (56) *mutatis mutandis*, with: MaxQud.q = what did A mean by FEC?, FEC = ‘We go straight on ’ (the occurrence after ‘or’) interpreting the alteration as a short answer.

Example (62c) can receive a similar analysis, although there is no editing phrase, in this case MaxQud.q = what did A mean by FEC?, FEC = ‘it is’ (the occurrence after ‘and’), and interpreting the alteration as a short answer. What is interesting about this case is that the reparamandum ‘it is’ is not a constituent. This exemplifies our earlier suggestion that the elements of PENDING need not always be viewed as constituents, but rather as elements of a *chart*.

## 5.4 Forward Looking Dysfluencies

Forward Looking Dysfluencies are distinct from their backward cousins in one significant way, on our view—they *require* an editing phrase, one whose import is the existence of a soon-to-be-uttered word. We will presently offer a lexical entry for ‘um’, inspired in part by Clark and FoxTree (2002) and Horne (2008) who argue that filled pauses are *conventionally* used interjections.

We specify FLDs with the update rule in (64)—given a context where the LatestMove is a forward looking editing phrase by A, the next speaker—underspecified between the current one and the addressee—may address the issue of what A intended to say next by providing a co-propositional utterance:<sup>28</sup>

<sup>27</sup>Arguably, it can also be understood as referring to  $u0$ , but in that case there would seem to be an indirect reference to  $u1$  as well, so we avoid this more complex scenario.

<sup>28</sup>This rule is inspired in part by Purver’s rule for *fillers*, (91), p. 92, (Purver, 2004). Given that our rule leaves the turn ownership unspecified we unify FLDs with fillers.

(64) Forward Looking Utterance Rule:

$$\left[ \begin{array}{l}
 \text{spkr} : \text{Ind} \\
 \text{addr} : \text{Ind} \\
 \text{preconds} : \left[ \begin{array}{l}
 \text{pending} = \langle \text{p0,rest} \rangle : \text{list}(\text{LocProp}) \\
 \text{u0} : \text{LocProp} \\
 \text{c1} : \text{member}(\text{u0}, \text{p0.sit.constits}) \\
 \text{LatestMove}^{\text{content}} = \text{FLDEdit}(\text{spkr}, \text{u0}) : \text{IlocProp}
 \end{array} \right] \\
 \text{effects} : \text{TurnUnderspec} \wedge_{\text{merge}} \\
 \left[ \begin{array}{l}
 \text{MaxQud} = \\
 \left[ \begin{array}{l}
 \text{q} = \lambda x \text{ MeanNextUtt}(\text{pre.spkr}, \text{pre.u0}, x) \\
 \text{fec} = \{ \}
 \end{array} \right] : \text{InfoStruc} \\
 \text{LatestMove} : \text{LocProp} \\
 \text{c2} : \text{Copropositional}(\text{LatestMove}^{\text{content}}, \text{MaxQud})
 \end{array} \right]
 \end{array} \right]$$

(64) differs from its BLD analogue, in two ways. First, in that the preconditions involves the LatestMove having as its content what we describe as an *FLDEdit* move, which we elucidate somewhat shortly. Words like ‘uh’, ‘thee’ will be assumed to have such a force, hence the utterance of such a word is a prerequisite for an FLD. A second difference concerns parallelism: for BLDs it is intuitive that parallelism exists between reparandum and alteration (with caveats, as with the example (56) etc) given that one is replacing one sub-utterance with another that is essentially of the same type. However, for FLDs there is no such intuition—what is taking place is a search for the word after the reparandum, which has no reason to be parallel to the reparandum. Hence in our rule (64), the FEC is specified as the empty set.

To make things explicit, we assume that ‘uh’ could be analyzed by means of the lexical entry in (65):<sup>29</sup>

$$(65) \left[ \begin{array}{l}
 \text{phon} : \text{uh} \\
 \text{cat} = \textit{interjection} : \text{syncat} \\
 \text{dgb-params} : \left[ \begin{array}{l}
 \text{spkr} : \text{IND} \\
 \text{addr} : \text{IND} \\
 \text{MaxPending} : \text{LocProp} \\
 \text{u0} : \text{LocProp} \\
 \text{c1} : \text{member}(\text{u0}, \text{MaxPending.sit.constits}) \\
 \text{rest} : \text{address}(\text{spkr}, \text{addr}, \text{MaxPending})
 \end{array} \right] \\
 \text{cont} = \left[ \text{c1} : \text{FLDEdit}(\text{spkr}, \text{addr}, \text{MaxPending}) \right] : \text{Prop}
 \end{array} \right]$$

We demonstrate how to analyze (66):

(66) A: Show flights arriving in uh Boston. (Shriberg, 1994)

After A utters  $u0 = \text{‘in’}$ , she interjects ‘uh’, thereby expressing  $\text{FLDEdit}(A, B, \text{‘in’})$ . This triggers the Forward Looking Utterance rule with  $\text{MaxQud.q} = \lambda x \text{ MeanNextUtt}(A, \text{‘in’}, x)$ . ‘Boston’ can then be interpreted as answering this question, with resolution based on the short answer rule.

<sup>29</sup>This lexical entry needs to be refined somewhat since it does not, as it stand, allow for turn initial utterances of ‘uh’, which are clearly possible.

Similar analyses can be provided for (67). Here instead of ‘uh’ we have lengthened versions of ‘the’ and ‘a’ respectively, which express FLDEdit moves:

- (67) a. And also the- the dog was old. (Besser & Alexandersson, 2007)  
 b. A vertical line to a- to a black disk (Levelt, 1989)

Let us return to consider what the predicate ‘FLDEdit’ amounts to from a semantic point of view. Intuitively, (68) should be understood as ‘A wants to say something to B *after* u0, but is having difficulty (so this will take a bit of time)’:

- (68) FLDEdit(A,B,u0)

This means we could unpack (68) in a number of ways, most obviously by making explicit the utterance-to-be-produced *u1*, representing this roughly as in (69):

- (69)  $\exists u1[\text{After}(u1,u0) \wedge \text{Want}(A,\text{Utter}(A,B,u1))]$

This opens the way for a more ‘pragmatic’ account of FLDs, one in which (64) could be *derived* rather than stipulated. Once a word is uttered that introduces FLDEdit(A,B,u0) into the context, in other words has an import like (69). This leads to a context akin to ones like (70), that license *inter alia* elliptical constructions like sluicing and pronominal anaphora, tied as they are to an existential quantifier in the semantic representation:

- (70) a. A: A woman phoned. (Potential follow ups: A/B: She ... ; B: Who?)  
 b. A: Max drank some wine. (Potential follow ups: A/B: It ... ; B: What kind of wine? )

Indeed a nice consequence of (64), whether we view it as basic or derived, is that it offers the potential to explain cases like (71) where in the aftermath of a filled pause an issue along the lines of the one we have posited as the *effect* of the conversational rule (64) actually gets uttered:

- (71) a. Carol: Well it’s (pause) it’s (pause) er (pause) what’s his name? Bernard Matthews’ turkey roast. (BNC, KBJ)  
 b. Here we are in this place, what’s its name? Australia.  
 c. They’re pretty ... um, how can I describe the Finns? They’re quite an unusual crowd actually.<sup>30</sup>  
 d. I understand you have to do your job, but sometimes you can maybe do it a little bit more ... I don’t have the right word, I don’t want to be mean.<sup>31</sup>

On our account such utterances are licensed because these questions are co-propositional with the issue ‘what did A mean to say after u0’. This suggests that a different range of such questions will occur depending on the identity of (the syntactic/semantic type of) u0.<sup>32</sup>

To test whether this is indeed the case, we ran a corpus study on the spoken language section of the BNC, using the search engine SCoRE (Purver, 2001) to search for all self-addressed queries.<sup>33</sup>

<sup>30</sup><http://www.guardian.co.uk/sport/2010/sep/10/small-talk-steve-backley-interview>.

<sup>31</sup><http://www.guardian.co.uk/sport/2013/jan/27/victoria-azarenka-australian-open-victory>

<sup>32</sup>We are grateful to an anonymous reviewer for alerting us to this issue and the related issue of whether any question, in principle, would do, as long as it would ultimately lead to the right answer. The reviewer’s example was (i):

(i) Well its er (pause) what’s the fifth root of 32? 2 turkey roasts

<sup>33</sup>We searched using the pattern

Representative examples are in (72) and the distribution is summarized in Table 1.

- (72) a. (**anticipating an N'**:) on top of the erm (pause) what do you call it?  
 b. (**anticipating a locative NP**:) No, we went out on Sat , er Sunday to erm (pause) where did we go?  
 c. (**anticipating an NP complement**:) He can't get any money se (pause) so so he can't get erm (pause) what do you call it?  
 d. (**anticipating a person-denoting NP**:) But you see somebody I think it was erm what's his name?  
 e. (**anticipating a person-denoting NP**: with erm, who was it who went bust?  
 f. (**anticipating a predicative phrase**: she's erm (pause) what is she, Indian or something?

category	questions found	Total
pre NP: prep - or verb - or NP and -	what's his/her name? what do they/you call him/her/it? who was it/the woman? what's the other one? what did you/I say? what did it mention	19 13 3 3 2 2 <b>42</b>
det -	what do/did they/you call it/that/them what's it called what is it what am I looking for	14 2 3 1 <b>20</b>
locative prep -	Where is it Where do they call that What's the name of the street/address what do they call X Where do we go Where did it say now what is it	3 2 2 2 1 1 1 <b>12</b>
be -	what is she/it what's the word I want? what do you call it?	3 1 1 <b>5</b>
say -	what did X say where did I get the number?	3 1 <b>4</b>
Total self addressed questions		<b>83</b>

Table 1: Self addressed questions in dysfluencies in the British National Corpus

Table 1 indicates that self-addressed queries occur in a highly restricted set of con-  
 'noun preceding 'er' or 'erm' preceding a wh word, adjacent to a verb.'

texts, above all where an NP is anticipated and after ‘the’. Moreover, the distribution of such queries across these contexts varies manifestly: the anticipated NP contexts involve predominantly a search for a name or for how the person/thing is called with some ‘who’-questions as well, whereas the post ‘the’ contexts only allow ‘what’ questions, predominantly of the form ‘what does X call Y’; anticipated location NP contexts predominantly involve ‘where’ questions. The final two classes identified are somewhat smaller, so generalizations there are less robust—nonetheless, the anticipated predicative phrase and post ‘say’ context seem to involve quite distinct distributions from the other classes mentioned above.

With respect to self addressed queries we have so far suggested that their coherence is accounted for directly on the basis of the conversational rule that licenses utterances that are co-propositional with the question ‘what did A mean to say after u0’. Capturing in this way an analogy with the coherence of clarification questions by B after a (completed) utterance by A.

Self addressed queries also highlight another feature of KoS’s dialogue semantics: the fact that a speaker can straightforwardly answer their own question, indeed in these cases the speaker is the “addressee” of the query. Such cases get handled easily in KoS because turn taking is abstracted away from querying: the conversational rule *QSpec*, introduced earlier as (22b), allows either conversationalist to take the turn given the QUD-maximality of *q*. This contrasts with a view of querying derived from Speech Act Theory (e.g. Searle (1969)) still widely assumed (see e.g. Asher and Lascarides (2003)), where there is very tight link to intentional categories of 2-person dialogue (‘...Speaker wants Hearer to provide an answer ... Speaker does not know the answer ...’).

## 6 Conclusions

In this paper we have developed an account of the semantics of dysfluencies. Our account distinguishes Backwards Looking Dysfluencies (BLDs), dysfluencies where the moment of interruption is followed by an alteration that refers back to an already uttered reparandum from Forwards Looking Dysfluencies (FLDs), where the moment of interruption is followed by a completion of the utterance which is delayed by a filled or unfilled pause (hesitation) or a repetition of a previously uttered part of the utterance (repetition). In both cases the mechanisms involved are minor refinements of rules proposed in earlier work to deal with clarificational interaction. The only substantive assumption we take on board relative to this earlier work is the assumption of incremental interpretation, the assumption that the grammar provides types which enable word-by-word parsing and interpretation. In fact, for cross-turn dysfluencies, we demonstrate that our account applies without any assumptions of intrasentential incremental processing. The assumption of the need for incremental processing is one that is supported by a wealth of recent work in psycholinguistics and is incorporated in a number of current grammatical frameworks.

Our account offers a precise explication of the roles of all key components of a dysfluency, including editing phrases and filled pauses, capturing the parallelism between reparandum and alteration, while also allowing for instances where it is relaxed, as in sentential alterations. It directly predicts the possibility of self-addressed questions, a class of queries that occurs in a very restricted range of syntactic/semantic contexts and that has not been described or analyzed in previous work. More generally, it provides a unified analysis of repair and correction that incorporates disagreement at illocutionary and metacommunicative levels, as well as self-correction across and within turns. There is no existing account with this coverage, to the best of our knowledge.

The current work is clearly ‘proof of concept’. What remains to be done is to develop a detailed incremental semantics, as well as to consider in detail the range of dysfluencies evinced in actual and potential conversations. It is important to do this across a wide

range of languages given the range of cross-linguistic variation with regards to dysfluency constructions surveyed in section 2.2.4. Finding a principled explanation for the syntactic/semantic contexts in which self-addressed questions occur, one which is presumably tied to common areas of difficulty in the utterance planning process, is also important. Indeed in line with the aforementioned work on cross-linguistic variation, we hypothesize that the syntactic/semantic contexts in which self-addressed questions occur should vary significantly across languages. We hope to pursue all this in future work.

The account we provide has significant methodological import and forces a number of foundational issues to be addressed. As we have seen, dysfluencies are an utterly ubiquitous phenomenon in language use that interacts with a variety of linguistic phenomena (including anaphora, ellipsis, implicature, discourse particles) and are subject to phonological, syntactic, and semantic constraints internal to individual languages. Nonetheless, they can only be analyzed in frameworks where metacommunicative interaction is integrated into the linguistic context. This partitions frameworks where such integration is effected (e.g. KoS, PTT (Poesio & Rieser, 2010) ) or at least addressed (e.g. Dynamic Syntax (Purver, Gregoromichelaki, Meyer-Viol, & Cann, 2010)) from work in most current formal semantic accounts of context where such integration is missing (e.g. standard DRT (van Eijck & Kamp, 1997), SDRT (Asher & Lascarides, 2003), Roberts' formal pragmatics (Roberts, 1996)/2012, (Farkas & Bruce, 2010), Inquisitive Semantics (Groenendijk & Roelofsen, 2009).) and which cannot, therefore, in principle, analyze dysfluency phenomena. A more fundamental point can be made: editing phrases like 'no', 'or', and 'I mean' select *inter alia* for speech events that include the discompetent products of performance. This means that the latter are also integrated within the realm of semantic competence. Just like friction is routinely abstracted away from analysis by physicists, though is straightforwardly integrated into their models, the same should hold for dysfluencies in models of linguistic knowledge and use. This suggests the need to rethink the traditional competence/performance dichotomy in a way that avoids casting aside pervasively produced classes of utterances.

## Acknowledgements

Raquel Fernández acknowledges support from NWO (MEERVOUD grant 632.002.001). David Schlangen acknowledges support from DFG (Emmy Noether Programme). Some portions of this paper were presented at Constraints in Discourse 2011 in Agay, at the 2011 Amsterdam Colloquium and at a 2012 ESSLLI evening lecture in Opole. We thank the audiences there as well as reviewers for Amsterdam Colloquium and for *Semantics and Pragmatics* for their very helpful comments.

(Ginzburg)

CLILLAC-ARP (EA 3967) & Laboratoire Linguistique Formelle (LLF) (UMR 7110)  
& Laboratoire d'Excellence (LabEx)–Empirical Foundations of Linguistics (EFL)  
Université Paris-Diderot, Sorbonne Paris Cité, Paris, France

(Fernández)

Institute for Logic, Language & Computation  
University of Amsterdam  
P.O. Box 94242, 1090 GE Amsterdam, The Netherlands

(Schlangen)

Faculty of Linguistics and Literary Studies  
Bielefeld University, P.O. Box 10 01 31, 33615 Bielefeld, Germany

## References

- Allen, J. & Core, M. (1997). Draft of damsl: dialog act markup in several layers. Discourse Research Initiative.
- Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914–930.
- Asher, N. & Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press, Cambridge.
- Bailey, K. G. D. & Ferreira, F. (2007). The processing of filled pause disfluencies in the visual world. In van Gompel, R. P. G., Fischer, M. H., Murray, W. S., & Hill, R. (Eds.), *Eye Movements: A Window on Mind and Brain*, pp. 485–500. Elsevier.
- Barwise, J. & Perry, J. (1983). *Situations and Attitudes*. Bradford Books. MIT Press, Cambridge.
- Besser, J. & Andersson, J. (2007). A comprehensive disfluency model for multi-party interaction. In *Proceedings of SigDial 8*, pp. 182–189.
- Brennan, S. E. & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*, 274–296.
- Callaway, C. (2003). Do we need deep generation of disfluent dialogue?. In *Proceedings of the AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, pp. 6–11 Palo Alto, CA.
- Candea, M., Vasilescu, I., Adda-Decker, M., et al. (2005). Inter-and intra-language acoustic analysis of autonomous fillers. In *Proceedings of DISS 05, Disfluency in Spontaneous Speech Workshop*, pp. 47–52.
- Charniak, E. & Johnson, M. (2001). Edit detection and parsing for transcribed speech. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics and Language technologies*, pp. 1–9. Association for Computational Linguistics.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Chomsky, N. (1972). *Language and Mind (extended edition)*. Harcourt, Brace & Jovanovich.
- Clark, H. & FoxTree, J. (2002). Using uh and um in spontaneous speech. *Cognition*, *84*, 73–111.
- Clark, H. (1996). *Using Language*. Cambridge University Press, Cambridge.
- Clark, H. & Schaefer, E. (1989). Contributing to discourse. *Cognitive science*, *13*(2), 259–294.
- Cooper, R. (2005). Austinian truth, attitudes and type theory. *Research on Language and Computation*, *3*(4), 333–362.
- Cooper, R. (2012). Type theory and semantics in flux. In Kempson, R., Asher, N., & Fernando, T. (Eds.), *Handbook of the Philosophy of Science*, Vol. 14: Philosophy of Linguistics. Elsevier, Amsterdam.
- Cooper, R. & Ginzburg, J. ((forthcoming)). Type theory with records for natural language semantics. In Fox, C. & Lappin, S. (Eds.), *Handbook of Contemporary Semantic Theory, second edition* Oxford. Blackwell.
- Cooper, R. & Poesio, M. (1994). Situation theory. In *Fracas Deliverable D8*. The Fracas Consortium, Centre for Cognitive Science, Edinburgh.

- Copestake, A. (2007). Semantic composition with (robust) minimal recursion semantics. In *Proceedings of the Workshop on Deep Linguistic Processing*, pp. 73–80. Association for Computational Linguistics.
- Core, M. (1999). *Dialog Parsing: From Speech Repairs to Speech Acts*. Ph.D. thesis, Department of Computer Science, University of Rochester, Rochester, New York.
- Core, M. G. & Schubert, L. K. (1999). Speech repairs: a parsing perspective. In *Satellite meeting ICPHS 99*, pp. 47–50.
- Cori, M., De Fornel, M., & Marandin, J.-M. (1997). Parsing repairs. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 47–58.
- De Fornel, M. & Marandin, J. (1996). L’analyse grammaticale des auto-réparations. *Le gré des langues*, 10, 8–68.
- Farkas, D. & Bruce, K. (2010). On reacting to assertions and polar questions. *Journal of Semantics*, 27:1, 81–118.
- Fay, D. (1980). Transformational errors. In Fromkin, V. A. (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. Academic Press, New York.
- Fernández, R. (2006). *Non-Sentential Utterances in Dialogue: Classification, Resolution and Use*. Ph.D. thesis, King’s College, London.
- Fernando, T. (2007). Observing events and situations in time. *Linguistics and Philosophy*, 30(5), 527–550.
- Ferreira, F. & Bailey, K. G. D. (2004). Disfluencies and human language comprehension. *TRENDS in Cognitive Science*, 8(5), 231–237.
- Ferreira, F., Lau, E. F., & Bailey, K. G. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, 28(5), 721–749.
- Finlayson, I. R. & Corley, M. (2012). Disfluency in dialogue: an intentional signal from the speaker?. *Psychonomic Bulletin and Review*, 19, 921–928.
- Fox, B., Hayashi, M., & Jaspersen, R. (1996). Resources and repair: a cross-linguistic study of syntax and repair. *Studies in Interactional Sociolinguistics*, 13, 185–237.
- Fox, B. A., Maschler, Y., & Uhmman, S. (2010). A cross-linguistic study of self-repair: evidence from english, german, and hebrew. *Journal of Pragmatics*, 42(9), 2487–2505.
- Gardent, C. & Kohlhase, M. (1997). Computing parallelism in discourse. In *IJCAI*, pp. 1016–1021.
- Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford.
- Ginzburg, J. & Fernández, R. (2010). Computational models of dialogue. In Clark, A., Fox, C., & Lappin, S. (Eds.), *Handbook of Computational Linguistics and Natural Language* Oxford. Blackwell.
- Ginzburg, J. & Sag, I. A. (2000). *Interrogative Investigations: the form, meaning and use of English Interrogatives*. No. 123 in CSLI Lecture Notes. CSLI Publications, Stanford: California.
- Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). Switchboard: telephone speech corpus for research and development. In *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, pp. 517–520 San Francisco, USA.
- Groenendijk, J. & Roelofsen, F. (2009). Inquisitive semantics and pragmatics. Presented at the Stanford workshop on Language, Communication and Rational Agency, May 30-31, 2009 .

- Grover, C., Brew, C., Manandhar, S., Moens, M., & Schoter, A. (1994). Priority union and generalization in discourse grammar. In Schoter, A. & Vogel, C. (Eds.), *Edinburgh Working Papers in Cognitive Science: Nonclassical feature systems*. Centre for Cognitive Science, University of Edinburgh, Edinburgh.
- Heeman, P. A. & Allen, J. F. (1999). Speech repairs, intonational phrases and discourse markers: modeling speakers' utterances in spoken dialogue. *Computational Linguistics*, 25(4), 527–571.
- Hindle, D. (1983). Deterministic parsing of syntactic non-fluencies. In *Proceedings of the 21st annual meeting on Association for Computational Linguistics*, ACL '83, pp. 123–128.
- Horne, M. (2008). Attitude reports in spontaneous dialogue: uncertainty, politeness and filled pauses. In Borin, L. & Larsson, S. (Eds.), *From Quantification to Conversation*, pp. 309–318. Gothenburg University, Gothenburg.
- Horne, M. (2012). Attitude reports in spontaneous dialogue: uncertainty, politeness and filled pauses. In Borin, L. & Larsson, S. (Eds.), *From Quantification to Conversation: Festschrift for Robin Cooper on the occasion of his 65th Birthday*, pp. 297–308. College Publications, Lars Borin and Staffan Larsson. This paper appeared in an online version of this collection in 2008.
- Hough, J. & Purver, M. (2012). Processing self-repairs in an incremental type-theoretic dialogue system. In Schmidt, S. B., Ginzburg, J., & Larsson, S. (Eds.), *Proceedings of SemDial 2012 (Seinedial)*, pp. 136–144 Paris, France.
- Jackendoff, R. (2005). Alternative minimalist visions of language. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, Vol. 41, pp. 189–226.
- Johnson, M. & Charniak, E. (2004). A tag-based noisy channel model of speech repairs. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, p. 33. Association for Computational Linguistics.
- Keller, F. (2000). *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. Ph.D. thesis, University of Edinburgh.
- Kempson, R., Meyer-Viol, W., & Gabbay, D. (2000). *Dynamic Syntax: The Flow of Language Understanding*. Blackwell, Oxford.
- Clark, A., Giorgolo, G., & Lappin, S. (2013). Statistical representation of grammaticality judgements: the limits of n-gram models. ACL Workshop on Cognitive Modeling and Computational Linguistics, Sofia.
- Larsson, S. (2002). *Issue based Dialogue Management*. Ph.D. thesis, Gothenburg University.
- Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14(4), 41–104.
- Levelt, W. J. (1989). *Speaking: From intention to articulation*. The MIT Press.
- Lickley, R. J. (1994). *Detecting Disfluency in Spontaneous Speech*. Ph.D. thesis, University of Edinburgh.
- McKelvie, D. (1998). The syntax of disfluency in spontaneous spoken english. Human Communication Research Center, University of Edinburgh, Human Communication Research Center, University of Edinburgh.
- Miller, T. & Schuler, W. (2008). A unified syntactic model for parsing fluent and disfluent speech. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pp. 105–108. Association for Computational Linguistics.
- Milward, D. (1994). Dynamic dependency grammar. *Linguistics and Philosophy*, 17, 561–405.

- Milward, D. & Cooper, R. (1994). Incremental interpretation: applications, theory, and relationship to dynamic semantics. In *Proceedings of the 15th conference on Computational Linguistics*, pp. 748–754. Association for Computational Linguistics.
- Morrill, G. (2000). Incremental processing and acceptability. *Computational Linguistics*, 26(3), 319–338.
- Peldszus, A., Buß, O., Baumann, T., & Schlangen, D. (2012). Joint satisfaction of syntactic and pragmatic constraints improves incremental spoken language understanding. In *Proceedings of the Conference of the European Association for Computational Linguistics (EACL 2012)*, pp. 514–523 Avignon, France.
- Peldszus, A. & Schlangen, D. (2012). Incremental construction of robust but deep semantic representations for use in responsive dialogue systems. In Hajičová, E. (Ed.), *Proceedings of the Workshop on Advances in discourse analysis and its computational aspects at Coling 2012* Mumbai, India.
- Poesio, M. (1995). A model of conversation processing based on micro conversational events. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Poesio, M. & Rieser, H. (2010). (prolegomena to a theory of) completions, continuations, and coordination in dialogue. *Dialogue and Discourse*, 1, 1–89.
- Poesio, M. & Traum, D. (1997). Conversational actions and discourse situations. *Computational Intelligence*, 13, 309–347.
- Purver, M. (2001). Score: a tool for searching the bnc. Tech. rep. TR-01-07, King’s College, London.
- Purver, M. (2004). *The Theory and Use of Clarification in Dialogue*. Ph.D. thesis, King’s College, London.
- Purver, M., Eshghi, A., & Hough, J. (2011). Incremental semantic construction in a dialogue system. In Bos, J. & Pulman, S. (Eds.), *Proceedings of the 9th International Conference on Computational Semantics (IWCS)*, pp. 365–369. Oxford, UK.
- Purver, M., Ginzburg, J., & Healey, P. (2001). On the means for clarification in dialogue. In van Kuppevelt, J. & Smith, R. (Eds.), *Current and New Directions in Discourse and Dialogue*, pp. 235–256. Kluwer.
- Purver, M., Gregoromichelaki, E., Meyer-Viol, W., & Cann, R. (2010). Splitting the ‘i’s and crossing the ‘you’s: context, speech acts and grammar. In Lupkowski, P. & Purver, M. (Eds.), *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*, pp. 43–50 Poznań.
- Ranta, A. (1994). *Type Theoretical Grammar*. Oxford University Press, Oxford.
- Rieser, H. & Schlangen, D. (2011a). Introduction to the special issue on incremental processing in dialogue. *Dialogue and Discourse*, 2, 1–10.
- Rieser, H. & Schlangen, D. (2011b). Special issue on incremental processing in dialogue. *Dialogue and Discourse*, 2.
- Rieser, V. & Moore, J. (2005). Implications for generating clarification requests in task-oriented dialogues. In *Proceedings of the 43rd Meeting of the Association for Computational Linguistics* Michigan.
- Roberts, C. (1996). Information structure in discourse: towards an integrated formal theory of pragmatics. *Working Papers in Linguistics-Ohio State University Department of Linguistics*, 91–136. Reprinted in *Semantics and Pragmatics*, 2012.
- Rodriguez, K. & Schlangen, D. (2004). Form, intonation and function of clarification requests in german task-oriented spoken dialogues. In Ginzburg, J. & Vallduvi, E. (Eds.), *Proceedings of Catalog’04, The 8th Workshop on the Semantics and Pragmatics of Dialogue* Universitat Pompeu Fabra, Barcelona.

- Schegloff, E. (1987). Some sources of misunderstanding in talk-in-interaction. *Linguistics*, 25, 201–218.
- Schegloff, E., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361–382.
- Schlangen, D., Baumann, T., & Atterer, M. (2009). Incremental reference resolution: the task, metrics for evaluation, and a bayesian filtering model that is sensitive to disfluencies. In *Proceedings of SIGdial 2009, the 10th Annual SIGDIAL Meeting on Discourse and Dialogue* London, UK.
- Searle, J. (1969). *Speech Acts*. Cambridge University Press, Cambridge.
- Seidenberg, M. (1997). Language acquisition and use: learning and applying probabilistic constraints. *Science*, 275(5306), 1599.
- Shriberg, E. (1996). Disfluencies in switchboard. In *Proceedings of International Conference on Spoken Language Processing*, Vol. 96, pp. 11–14.
- Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of California at Berkeley, Berkeley, USA.
- Skantze, G. & Hjalmarsson, A. (2010). Towards incremental speech generation in dialogue systems. In *Proceedings of the 11th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 1–8 Tokyo, Japan. Association for Computational Linguistics.
- Stolcke, A. & Shriberg, E. (1996). Statistical language modeling for speech disfluencies. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 405–409.
- van Eijck, J. & Kamp, H. (1997). Representing discourse in context..
- Van Wijk, C. & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: evidence from experimentally elicited corrections. *Cognitive Psychology*, 19(4), 403–440.
- Wouk, F., Fox, B., Hayashi, M., Fincke, S., Tao, L., Sorjonen, M., Laakso, M., & Flores-Hernandez, W. (2009). A cross-linguistic investigation of the site of initiation of same turn self repair. *Conversation analysis: Comparative perspectives*.
- Yoshida, E. & Lickley, R. (2010). Disfluency patterns in dialogue processing. In *Proceedings of DiSS-LPSS Joint Workshop*, pp. 115–118.
- Zhao, Y. & Jurafsky, D. (2005). A preliminary study of mandarin filled pauses. In *Disfluency in Spontaneous Speech*.