JONATHAN GINZBURG and ROBIN COOPER

CLARIFICATION, ELLIPSIS, AND THE NATURE OF CONTEXTUAL UPDATES IN DIALOGUE

ABSTRACT. The paper investigates an elliptical construction, Clarification Ellipsis, that occurs in dialogue. We suggest that this provides data that demonstrates that updates resulting from utterances cannot be defined in purely semantic terms, contrary to the prevailing assumptions of existing approaches to dynamic semantics. We offer a computationally oriented analysis of the resolution of ellipsis in certain cases of dialogue clarification. We show that this goes beyond standard techniques used in anaphora and ellipsis resolution and requires operations on highly structured, linguistically heterogeneous representations. We characterize these operations and the representations on which they operate. We offer an analysis couched in a version of Head-Driven Phrase Structure Grammar combined with a theory of information states (IS) in dialogue. We sketch an algorithm for the process of utterance integration in IS which leads to grounding or clarification. The account proposed here has direct applications to the theory of attitude reports, an issue which is explored briefly in the concluding remarks of the paper.

1. INTRODUCTION

1.1. *Dynamic Semantics: Representationalism and (Weak) Montogovianism*

The last two decades of the twentieth century saw the rise of dynamic semantics, a paradigm that sought to replace a semantics based on truth conditions with one based on context change. One of the main issues that provoked much interest within dynamic semantics was the issue of *representationalism*. Initial dynamic semantic work within the frameworks of Discourse Representation Theory (DRT) (Kamp 1981; Kamp and Reyle 1993) and File Change Semantics (FCS) (Heim 1982) argued that (discourse) representations were indispensable components of the dynamic semantics programme, needed to explicate anaphoric potential, presupposition (van der Sandt 1992), and the attitudes (see e.g., work within S(egmented)DRT (Asher 1993; Asher and Lascarides 1998)). In reaction to this emerged a body of work, particularly within the frameworks of Dynamic Predicate Logic (Groenendijk and Stokhof 1991b) and Dynamic Montague Grammar (Groenendijk and Stokhof 1991a; Chierchia 1995)

Linguistics and Philosophy **27:** 297–365, 2004.

which sought, in the spirit of the work of Montague (73), to demonstrate the dispensibility of representations in semantic theorizing, at least as far as anaphora and presupposition were concerned.[1] Despite the division over representationalism, in their explication of context change both DRT/FCS/SDRT and DPL/DMG implicitly subscribe to a view we will dub the Pure content hypothesis or weak Montogovianism:

(1) Pure content hypothesis (Weak Montogovianism): the content which is updated in dynamic semantics consists of (a representation of) the *content* of utterances (excluding formal linguistic properties such as syntactic or phonological properties.).

The sobriquet Weak Montogovianism derives from a tendentious comment made by Montague, namely that he failed 'to see any great interest in syntax except as a preliminary to semantics'. Although this remark occasioned much opprobium for Montague, the formulation in (1) seems commonly adhered to in the dynamic semantics approaches cited above. In this paper, we investigate an elliptical construction, Clarification Ellipsis, that occurs in dialogue. We will suggest that this provides data that refutes Weak Montogovianism. In a dynamic approach, where meaning is explicated in terms of context change, the negation of Weak Montogovianism, which can be stated positively as (2), implies representationalism.

(2) Hybrid content hypothesis: the content which is updated in dynamic semantics consists of structure expressing detailed relationships between the content and formal properties (syntax, phonology etc) of the various parts of an utterance.

Beyond this, we will suggest that, at least as far as the semantics of dialogue goes, context change needs to be explicated with reference to utterances. That is, one needs to adopt the following:

(3) Utterances as events hypothesis: Utterances are spatio-temporally located events involving the sequential enunciation of one or more word(s).

Utterances as events originates in early work in Situation Semantics (Barwise and Perry 1983). In various guises, it plays a significant role in recent work on underspecification – explicitly argued for in Poesio (1998), more implicitly assumed in Copestake et al., (n.d.), Milward (2000). The

[1] The issue of attitude reports received relatively little attention in these approaches.

account we develop also has direct applications to the theory of attitude reports, even though we will address this issue here only in passing.

1.2. *Clarification Ellipsis: Key Properties*

Clarification ellipsis (CE), nonsentential elliptical queries such as (4a(i),(ii)) that most prototypically involve repetition of a word/constituent from the most recent utterance, are commonplace in human conversation.[2,3,4,5] Despite their syntactic simplicity, CEs can convey complex and widely distinct contents. Two common readings/understandings of CE are exemplified in (4b, c):[6]

(4) a. A: Did Bo finagle a raise? B: (i) Bo?/ (ii) Finagle?

[2] An anonymous reviewer for *Linguistics and Philosophy* has suggested to us that it is problematic to use the term 'ellipsis' for CE, given that neither reading we are concerned with involves ellision of material that occurs previously in the dialogue in a straightforward syntactic way. For whatever it's worth though, the utterances in question are elliptical in the standard accepted sense of the word, i.e. the meaning they get could be more completely expressed by adding additional constituents.

[3] See Section 5.3 for a discussion of uses of this construction that do not involve requests for clarification.

[4] See Purver et al. (2002) for a detailed analysis of the frequency of CEs among clarification requests in the British National Corpus (BNC). This is based on a random sampling of the 10 million word dialogue subcorpus of the BNC consisting of c. 150,000 words. 4% of sentences were found to be clarification requests. Of these 29% were found to be CEs, an indication of the productivity of the construction.

[5] Although CE has not received much attention in theoretical or computational linguistics, its occurrence has been widely noted in the literature on language acquisition, see e.g. (Garvey 1979; McTear 1987; Ninio and Snow 1996).

[6] The examples of CE we provide throughout the paper involve as antecedent an interrogative context. This is purely for expository convenience, given that in such contexts, a bare fragment enunciated with rising intonation can usually only be understood as CE. Of course CE can equally occur with declarative antecedents as in (i), though in such a case, arguably, there also exists a non-CE reading that involves expressing doubt as to the claim made by the other speaker, paraphrased as in (iv):

(i) A: Bo finagled a raise. B: (i) Bo?/ (ii) finagled?
(ii) **Clausal reading**: Are you claiming that BO (of all people) finagled a raise/Bo FINAGLED a raise (of all actions)
(iii) **Constituent reading**: Who is Bo?/What does it mean to finagle?
(iv) Did BO (of all people) finagle a raise/Did Bo FINAGLE a raise (of all actions)

The difference between a declarative and interrogative antecedent, then, is only of import as far as the clausal reading goes. As will become clear, our account can accommodate CE readings with declarative antecedents (or imperative antecedents, for that matter) with equal ease as interrogative ones; for this, see footnote 50.

b. **Clausal reading**: Are you asking if BO (of all people) finagled a raise/Bo FINAGLED a raise (of all actions)

c. **Constituent reading**: Who is Bo?/What does it mean to finagle?

The *clausal* reading is most readily paraphrased by a polar interrogative – it is commonly used simply to confirm the content of a particular subutterance. The *constituent* reading, in contrast, is most readily paraphrased by a *wh*-interrogative – its main function is to elicit an alternative description or ostension to the content (referent or predicate etc) intended by the original speaker of the reprised subutterance.

There is one important issue related to CE that we need to bring up at the outset, given that the way (we believe) it is resolved plays an important role in the rest of the paper. This is the issue of whether CE involves an ambiguity or is simply vague. A number of people have suggested to us that all CEs could be analyzed in terms of a single reading, so that e.g. (3a(i)) could be paraphrased as follows:[7]

(5) I thought I heard you say *Bo*, and I don't know why you would do so?

The range of understandings associated with (4a(i)) would then be deduced from (5) using pragmatic reasoning of various sorts. Clearly, we agree, such reasoning does play a role in the understanding of CE, as it does in various other cases. However, with respect to the distinction between clausal and constituent understandings there are some considerations that do favour the existence of an ambiguity. First, the BNC provides numerous examples of misunderstandings concerning CE interpretation,[8] where a speaker intends one interpretation, is misunderstood, and clarifies his original interpretation:

(6) a. George ... you always had er er say every foot he had with a piece of spunyarn in the wire/Anon1: Spunyarn?/George: Spunyarn, yes/ Anon1: What's spunyarn?
George: Well that's like er tarred rope.
BNC file H5G, sentences 193–196

b. A: Have a laugh and joke with Dick./ B: Dick?/A: Have a laugh and joke with Dick./B: Who's Dick?

[7] This paraphrase was suggested by an anonymous ACL reviewer.

[8] This confirms our (non-instrumentally tested) impression that these understandings are *not* on the whole disambiguated intonationally. All our CE data from the BNC was found using SCoRE, Matt Purver's dialogue oriented BNC search engine (Purver 2001).

c. Andy: Bloody lucky though./Monica: Pikey! Typical! / Andy: Pikey? /Nick: Pikey! /Andy: What's pikey? What does pikey mean? / Monica: I dunno. Crusty.

Actually, with respect to the examples in (6), it is hard to rule out an additional interpretation: Anon1/A/Andy, lacking a referent for *Spunyarn/Dick/pikey*, initially checks to see that he has heard correctly, and only after that asks a question equivalent to a constituent CE reading. But even if this were the case, this would not strengthen the case for a paraphrase along the lines of (5). It is hard to see how using the latter enables one to predict George/B/Nick's response. Rather, for (6a), for instance, (5) would lead us to expect George to respond:

(7) George: Spunyarn, because I thought it would be important to establish that spunyarn was in the wire.

More generally, it is hard to come up with a question that generalizes the clausal and constituent readings while predicting the correct range of responses CE elicits.

Perhaps even more crucially, the clausal and constituent readings involve distinct syntactic and phonological parallelism conditions. Neither clausal, nor constituent readings actually require phonological identity between target and source:

(8) a. A: Did Bo leave? B: My cousin? (clausal reading: **Are you asking if my cousin of all people left?**; constituent reading: **When you say *Bo*, are you referring to my cousin?**)

b. A: Did she annoy Bo? B: Sue? (clausal reading: **Are you asking if Sue of all people annoyed Bo?**; constituent reading: **When you say *she*, are you referring to Sue?**)

c. A: Did you bike to work yesterday? B: Cycle? (clausal reading: **Are you asking if I, of all things, cycled to work yesterday?**; constituent reading: **When you say *bike*, are you referring to the activity of cycling?**)

Indeed, a similar range of readings emerges if a bare *wh*-phrase is used, as exemplified in (9). Note that for bare wh clarifications the clausal/constituent ambiguity is of less semantic significance than with non-wh CE. For (9a), for instance, both readings signal that B cannot resolve the reference of *Bo*. The difference is that the clausal reading involves a presupposition that all other constituents of meaning were unproblematic, whereas the constituent reading involves no such commitment:

(9) a. A: Did Bo leave? B: Who? (clausal reading: **Who$_i$ is it you're asking whether s/he$_i$ left?**; constituent reading: **When you say *Bo*, who are you referring to?**)

b. A: Did she annoy Bo? B: Who? (clausal reading: **Who$_i$ is it you're asking whether she$_i$ annoyed Bo?**; constituent reading: **When you say *she*, who are you referring to?**)

For both types of readings, nonetheless, partial syntactic parallelism does obtain: an XP used to clarify an antecedent sub-utterance u_1 must match u_1 categorially:[9]

(10) a. A: I phoned him. B: Him? / #he?

[9] Given the relatively poor morphological marking of English, syntactic parallelism is not always straightforward to demonstrate conclusively. An anonymous reviewer for *Linguistics and Philosophy* points out that (i) is possible as a response to (10d):

(i) B: Exercise? Me?

We have not claimed in this paper to provide an exhaustive characterization of all possible non-sentential utterance types, intended as clarifications or otherwise. We believe that (i) does not get either of the CE readings we have proposed. Indeed on its most obvious understanding, (i) is a response which conveys a negative response to the polar question posed by A. It is not a CE at all. Hence, in any case it cannot be taken to counterexemplify our claim about parallelism as a requirement on CE. We note though that in subsequent correspondence the reviewer disputes our judgement concerning (i), suggesting that it possesses a clausal reading.

The reviewer also points out that 'him' as a response to (10b) improves considerably if B is pointing at the intended referent. We believe this is connected with the fact that in English whereas accusative pronouns appear routinely in elliptical utterances, nominative pronouns are frequently infelicitous when standing alone, as illustrated in dialogues like the following, where the pronoun fragment has a nominative antecedent:

(i) A: Didn't KIM write that letter?
B: [pause] Nope. Me/#I/Her/#She/Him/#He.

(ii) A: Who stole the beer?
B: Bo/#I/Me/#He/Him/#She/Her. (examples (10a,b) in Chapter 8 of Ginzburg and Sag (2000))

Cross-linguistically, in languages with strong pronouns and particularly languages which mark case on all NPs, such contexts unambiguously call for a pronoun manifesting the case of the antecedent, as the following examples from German show:

(iii) A(1): Hat Kim nicht den Brief geschrieben?
B(2): Nein, Ich/#Mich/Er/#Ihm/Sie/#Ihr
A(1): Had Kim not the-acc letter sent?
B(2): No I-nom/acc/dat/He-nom/dat/She-nom/acc
A: Didn't KIM write that letter?
B: [pause] Nope. Me/Him/Her.

b. A: Did he phone you? B: He? / #him?

c. A: Did he adore the book? B: Adore? / #adored?

d. A: Were you cycling yesterday? B: Cycling?/biking?/#biked?

A systematic way to disambiguate clausal readings from constituent ones arises from their distinct semantic nature: clausal readings require a presupposition that both speech participants share (a belief about) the content of the to-be-clarified sub-utterance; there is no such requirement for constituent readings. This is demonstrated most clearly with indexicals. Thus, in (11) given that A and B are located at distinct locations, the content of *here* gets resolved distinctly for A and B. As a result, the only reading possible for the CE is a constituent reading:

(11) (Context: A is located in Gothenburg, B is located in Hyderabad) A: Let's hold the conference here. B: Here? (=what location are you talking about; ≠Are you asking if we should hold the conference in Hyderabad of all places?).

In a context where A and B are located in the same place, the possibility of a clausal reading is significantly enhanced, if not preferable:

(12) (Context: A and B are located in Gothenburg) A: Let's hold the conference here. B: Here? (Either: what location are you talking about; Or: Are you asking if we should hold the conference in Gothenburg of all places?).

(iv) A: Wem will sie schmeicheln? B: Ihm/#ihn.
Who-dat wants she flatter, B: He-dat/#acc.
A: Who does she want to flatter? B: Him.

(v) A: Wen will sie loben? B: Ihn/#ihm.
A: Who-acc wants she praise, B: He-acc/#dat.
A: Who does she want to praise? B: Him.

Indeed in such languages parallelism in CE is far easier to demonstrate, as illustrated here for German:

(vi) A: Ist dieser Platz noch frei? B: Dieser/#Diesem/#Diesen Platz?
Is this-nom place-nom still free this-nom/B:#this-dat/#this-acc place-nom
A: Is this place free? B: This place?

(vii) A: Darf ich Ihrem Freund noch ein Stück Torte geben? B: Ihm/#Er?
May I your-dat friend-dat another a piece cake give? B: him-dat/#he-nom
A: May I give your friend another piece of cake? B: Him?

In particular, the pronoun *I* whose reference changes across speakers can in CE only be understood as a constituent reading:

(13) a. A: Can I come in? B: I? (= **who is *I*, who are you** etc; cannot mean: **Am I asking if I of all people can come in**.)

This suggests that constituent readings can but need not involve quotation, whereas clausal readings do not. Thus, an adequate description of constituent readings will need to refer to phonological information of the source, whereas an adequate description of clausal readings can be blind to such information.

Note also that in contrast to the much discussed case of verb phrase ellipsis (VPE) and related constructions such as gapping,[10] antecedents for CE are not restricted by island constraints. This is illustrated by (14), where the antecedents for CE come from within a *wh*-interrogative (dependencies into which are subject to the so-called *wh*-island constraint.) and from a conjunct:

(14) a. A: Mary told me who Bill kowtowed to at Jill's party. B: Bill?/kowtowed?

b. A: Mo and Bo left. B: Bo?

The fact that conversationalists can consistently pick an unintended CE understanding and the existence of distinct parallelism conditions on the clausal and constituent understandings strongly suggests that these two understandings arise out of distinct linguistic mechanisms. In other words, ambiguity is involved, not vagueness.[11]

[10] It should be noted that whether VPE exhibits a syntactic parallelism condition is controversial and subject to complex conditions, see e.g. (Kehler 1993; Hardt 1993).

[11] Closely related to this issue is the issue of what other readings/understandings, if any, CE exhibits. Based on introspective evidence, we had originally assumed that CE can also be used to elicit confirmations or alternative descriptions at the level of lexical identification ('Did you utter the word *Bo*?'). However, fairly extensive investigation of the BNC provides little support for this intuition, as discussed by Purver et al. (2002) – lexical identification tends to be queried by means of non-elliptical forms such as *what did you just say* or *eh?/pardon?/sorry?*. Purver et al. (2002) conclude that the clausal and constituent readings seem to be by far the two most robust understandings of CE. As will become clear below, given the contextual information used in the resolution method we will be proposing, there will be no difficulty in principle of accommodating a reading such as lexical identification.

1.3. *Theoretical Desiderata for an Account of CE*

Our aim in this paper is to develop a formal grammatical description of CE. This involves two main components – (a) an account of the contextual background in which the requisite ellipsis resolution takes place, and (b) a grammatical framework into which such resolution can be integrated. However, providing a descriptively adequate account of CE is a difficult challenge for existing approaches to ellipsis and to theories of dialogue context and its evolution. It also forces one to make a number of significant decisions about the architecture and ontology of grammar. Let us take these points in turn.

As we saw above, CE exhibits a tricky combination of phonological/partial syntactic parallelism and intricate contents. This seems to require a different approach from existing treatments of ellipsis in the theoretical and the computational literature. Thus, attempting to adapt existing copying/reconstruction approaches to bare NP ellipsis (e.g. (Reinhart 1991; Lappin and Gregory 1997)),[12] which try to build in sensitivity to islands into the ellipsis resolution process is problematic. Problems at the semantic level are perhaps more serious: the required readings for CE cannot be derived. Reinhart's approach couched in a Government Binding framework is based on LF movement of the XP construed as a generalized quantifier which predicates of a predicate formed by λ-abstracting over the antecedent clause. Such an approach cannot generate clausal readings without assuming that illocutionary force is syntactically represented in the antecedent LF, an obviously problematic assumption. For constituent readings of CE the situation is even worse – there does not seem to be any obvious way for the account to generate readings remotely approximating the desired content. Lappin and Gregory's HPSG approach involves copying the head daughter of the VP heading the clause in the previous conjunct and constructing an assignment of the elements of the fragment site to the copied verb's SUBCAT list. Generating a clausal reading would involve minimally assuming illocutionary information is represented somehow in the verb's semantics. As with Reinhart's approach, there seems no way to get at constituent readings.

An alternative, more semantically oriented approach to VPE and other ellipsis phenomena, is based on using Higher Order Unification (HOU)

[12] These treatments were developed to handle cases such as (i)–(iii):

(i) John threw flowers at Mary, and chocolates too. (Lappin and Gregory's (12))
(ii) Bill wrote reviews for the journal last year and articles this year. (Lappin and Gregory's (20))
(iii) John sings, and beautifully too. (Lappin and Gregory's (21)).

on logical forms (in the computational, not GB sense) to solve equations whose solutions yield the (unexpressed) ellipsis predicates (see e.g. (Dalrymple et al. 1991; Pulman 1997)). Given a theory of dialogue context which could set up the requisite equations, HOU could be used to derive the clausal reading of CE.[13] However, there is no obvious way to extend such a system to provide constituent readings. Moreover, since HOU operates on logical forms, that do not contain syntactic and phonological information, a HOU-based account would face difficulties in accommodating the phonological and syntactic parallelism exhibited by CE.

Given these difficulties, how then to develop an analysis of CE? There seem to be four fundamental ingredients which are required:

- **Fractal heterogeneity**: Utterance representations need to encode phonological, syntactic, semantic, and contextual information *fractally*.[14] That is, the requisite representation format needs to contain heterogenous (viz. phonological, syntactic, semantic, and contextual) information and, moreover, this applies uniformly as the parts get smaller and smaller.
- **Sub-utterance accessibility**: Access to all sub-utterances of the utterance is ensured. This given that, in principle, any semantically meaningful sub-utterance can be clarified using CE under conditions of phonological or partial syntactic parallelism.
- **Utterance reference**: CE involves reference to utterance *events* not just utterance *types*. In (14b), for instance, a constituent reading means *who are you referring to by Bo in the previous utterance*, NOT *who does Bo refer to in general* – the name *Bo* is borne by many individuals. Consequently, the analysis of utterances must include references to (previously occurring) utterance events.
- **Clarification potential**: A theory of dialogue processing should provide a characterization of the space of possible requests for clarification of a given utterance. In particular, it should allow the contents

[13] HOU for an ellipsis $Q(a)$ involves locating an antecedent logical form C and an antecedent 'parallel' term b occurring in C, solving the equation $Q(b) = C(b)$ (and filtering certain trivial or repetitive solutions). Given a theory of dialogue context which sets up the equations, HOU can yield the check reading of CE as follows:

(i) A: Did Bo leave? B: Bo?

(ii) $Q(b) = Ask(A, leave(b))$

(iii) $Q = \lambda x.Ask(A, leave(x))$ ("who$_x$ are you asking if x left").

[14] Our use of 'fractal' in this context follows Pollard (forthcoming).

that arise in CE to become available during a potential request for clarification.

All four ingredients are either theoretically controversial or not currently available. **Fractal heterogeneity** would seem to favour sign-based grammar architectures, a generalization of Montague's 'rule to rule' strategy. For instance, architectures proposed in theories like Head Driven Phrase Structure Grammar (HPSG) (see e.g. (Pollard and Sag 1994)), Lexical Functional Grammar (LFG) (e.g. (Bresnan 2000)), and in Categorial Grammar (see e.g. (Moortgat 1997)). In such approaches, representations of all linguistic types from phrases down to the word level involve parallel specification of phonological, syntactic, semantic, and contextual information. This contrasts with modular, derivational grammar architectures such as those proposed in current transformational approaches such as GB and the Minimalist Program (see e.g. (Chomsky 1986a, 1995)), in which syntactic operations are insulated from access to the PF and LF levels of representation.[15]

As far as we are aware, **Sub-utterance accessibility** is not directly implemented in any existing grammatical theory. As we will see, it is relatively straightforward to modify a phrase structure grammar in which constituency is internalized in the grammar (e.g. HPSG) to satisfy this. It is far harder to do this in a radically lexicalist framework such as Categorial Grammar.

Herbert Clark and a number of his collaborators (see the papers collected in Clark (1993) and the book Clark (1996)) initiated a significant turn in the study of dialogue. Clark et al. suggest, primarily on the basis of psycholinguistic experimentation and corpus analysis, that an important structuring factor in conversation is the continual monitoring by conversational participants (CPs) of each other for evidence that the latest utterance has been understood. If such evidence is provided, the utterance is *grounded* and things can proceed unhindered. Otherwise, clarification is called for. There has been much work on the grounding process, from experimental, corpus-based, formal and implementational perspectives (for the latter two see e.g. (Traum 1994; Poesio and Traum 1997)). One pivotal assumption of the model of grounding put forward by Poesio and Traum is the need to refer to utterance events, the assumption we have labelled above as **Utterance Reference**. Poesio in particular, following the lead of (Barwise and Perry 1983), has demonstrated the role **Utterance Reference**

[15] For detailed comparison of these two strategies see (Johnson and Lappin 1999).

plays in accounting for the functioning of locutions like 'the former'/'the latter' utterance-denoting demonstratives ('Could you repeat *that*').[16]

This assumption is controversial from a number of theoretical perspectives:

1. In a number of writings (see e.g. (Chomsky, 1986b)), Chomsky has argued that the proper domain of study for linguistics is a discipline he refers to as I(nternal)-language, a view of language as a biological endowment as 'some notion of structure in the mind of the speaker'. This he contrasts with a view of language (E(xternal)-language), as 'a collection of actions, or utterances, or linguistic forms (words, sentences) paired with meanings, or as a system of linguistic forms or events.' ((Chomsky, 1986b), p. 21). The implication is that utterances qua speech events have no role in the theory of grammar.
2. The classical view of meaning as represented by Montague tends to be of an object related to an occurrence in a formal language. As such meanings can be said to be related to the utterances which they are the meanings of and these utterances could be viewed as speech events. This is perhaps clearest in the work of Kaplan (1989) on indexicals. However, on the classical view meanings are normally considered to have an independent existence and not to be defined in terms of utterances. Cases where utterances actually refer to utterances are avoided as they are associated with paradoxes. The meaning of CEs are not only associated with the CE utterances but depend for their characterization on reference to previous utterances and this is not allowed according to the classical view where the domains of utterances and semantic reference are held separate.

Despite the wealth of work on grounding, there has been a dearth of work on what happens when grounding cannot take place. In particular, formal work on grounding such as (Poesio and Traum 1997) spells out in great detail what updates arise in an information state as a result of grounding and develop an extension of DRT in which analogues of **Sub-utterance accessibility** and **Utterance reference** hold.[17] However, such work has not offered a characterization of the clarification possibilities spawned

[16] See (Poesio and Muskens 1997) and references cited therein. In similar vein, Manfred Krifka points out to us the existence of utterances such as (i), which target the pronunciation of a previous utterance:

(i) Eliza: Give me the bo"les.
Prof. Higgins: Bo"les?

[17] It is a bit harder to assess whether **Fractal heterogeneity** holds. It does seem within the spirit of the approach, nonetheless.

by a given utterance. Providing the beginnings of such a characterization constitutes an important task, which we undertake in this paper.

The structure of the paper is as follows: we start by sketching in fairly pretheoretical terms an account of how an utterance gets integrated in an information state or leads to clarification. The following two sections, Sections 3 and 4, provide background material for the formal account we subsequently develop: our formalization is based on a relatively minor modification to a version of HPSG developed in Ginzburg and Sag (2000) (G&S-00), combined with a theory of dialogue interaction KOS[18] (Ginzburg 1996; Ginzburg (forthcoming); Bohlin et al. 1999). We couch the analysis in HPSG for two main reasons: first, because it already contains a number of the characteristics we argue are required from a grammatical framework (e.g. a single level in which phonology, content, and context are integrated.). Second, since the version of HPSG we will be utilizing already possesses fairly extensive analyses of dialogue ellipsis.[19] In Section 5 we present our account of CE: this will involve both a spelling out of certain context operations associated with clarification, as well as grammatical analysis. In Section 6 we sketch an algorithm for the process of utterance integration by CPs, an algorithm that underpins our analysis of CE. Finally, in Section 7 we offer some conclusions, including the tying together of the problem of CE with work on the representation of attitudes.

A formal account of CE, then, requires one to spell out both the mechanisms of context change which make available the entities by means of which ellipsis gets resolved, as well as how these mechanisms interface with the principles by means of which words and phrases get assigned a conventional import. This sort of account seems to us to fall squarely into what a pioneer of the field characterized generative grammar as being about:

> A grammar constructed in accord with the principles postulated in such a theory [of generative grammar] gives an explicit characterization of a language and its structure – and within the broader semiotic theory envisioned but not developed here, an explicit characterization as well of the meaning and reference of expressions and conditions of appropriate use. [(Chomsky 1955), p. 9.]

[18] Kos is the name of an island, not an acronym; 'k' pertains to konversation and 's' to semantics.

[19] However, see (Cooper and Ginzburg 2002) for an alternative formulation of portions of the present account using Martin-Löf Type Theory (MLTT). Apart from increased computational tractability and simplicity of logical apparatus, this provides one additional key advantage: MLTT offers perspicuous means of capturing DRT-style anaphora dynamics (see e.g. (Ranta 1994)).

2. Utterance Representation: Grounding and Clarification

We start by offering an informal description of how an utterance *u* such as (15) can get grounded or spawn a clarification by an addressee Bora:

(15) Ariadne: Did Jo leave?

Ariadne is attempting to convey to Bora her question whether the property she has referred to with her utterance of *leave* holds of the person she has referred to with the name *Jo*. Bora is required to try and find values for these references.[20] Finding values is, with an important caveat we return to shortly, a necessary condition for Bora to ground Ariadne's utterance, thereby signalling that its content has been integrated in Bora's Information State (IS). Modelling this condition for successful grounding provides one obvious constraint on the representation of utterance types: such a representation must involve a function from or λ-abstract over a set of certain parameters (the *contextual parameters*) to contents. This much is familiar already from early work on context dependence by Montague (1974a) and Kaplan (1989) to more recent work in situation semantics (Gawron and Peters 1990; Cooper and Poesio 1994).

The caveat we alluded to above is that the necessity to find values for contextual parameters is goal driven. Relative to certain goals one might decide simply to *existentially quantify* the problematic referent and work with this somewhat weakened content. Although in the current work we do not attempt to formalize the link between the goals underlying an utterance and the necessity to instantiate contextual parameters, we will, following (Israel and Perry 1991; Cooper 1998), propose a modelling of the operation of existential quantification of contextual parameters.

What happens when Bora cannot or is at least uncertain as to how he should instantiate in his IS a contextual parameter i? In such a case Bora needs to do at least the following: (1) perform a partial update of the existing context with the successfully processed components of the utterance, (2) pose a clarification question that involves reference to the sub-utterance u_i from which i emanates. Since the original speaker, Ariadne, can coherently integrate a clarification question once she hears it, it follows that,

[20] It is well-known that proper names can be used to refer to different individuals on different occasions. We tend to be less aware that other parts of speech have similar behaviour. The word *leave* can be used to refer to different actions on different occasions, e.g. temporary departure for a prearranged meeting, going home, resignation etc. Moreover, when considering dialogue one needs to be aware of different knowledge levels among the conversationalists which applies equally to their lexical knowledge.

for a given utterance, there is a predictable range of < partial updates + consequent clarification questions>. These we take to be specified by a set of *coercion operations* on utterance representations.[21] Indeed we assume that a component of dialogue competence is knowledge of these coercion operations.

CE gives us some indication concerning both the input and required output of these operations. One such operation, which we will refer to as parameter identification, essentially involves as output a question paraphrasable as *what is the intended reference of sub-utterance* u_i*?*. The partially updated context in which such a clarification takes place is such that simply repeating the segmental phonology of u_i using a distinctive intonation (e.g. focus-associated rise with spreading high tone or focus-associated fall with spreading low tone) enables that question to be expressed. Another existent coercion operation is one which we will refer to as parameter focussing. This involves a (partially updated) context in which the issue under discussion is a question that arises by instantiating all contextual parameters except for i and abstracting over i. In such a context, one can seek *confirmation* that i gets the value Bora suspects it has by uttering with intonational patterns mentioned before any apparently co-referential phrase whose syntactic category is identical to u_1's.

From this discussion, it becomes clear that coercion operations and by extension the grounding process cannot be defined simply on contents or even on meanings, regardless of how liberally we allow these to be structured. This is a somewhat surprising and certainly controversial conclusion – it implies that **updates resulting from utterances cannot be defined in purely semantic terms.** In other words, this restates the assumption we stated in Section 1 as the Hybrid content hypothesis, which we repeat here as (16)

(16) Hybrid content hypothesis: the content which is updated in dynamic semantics consists of structure expressing detailed relationships between the content and formal properties (syntax, phonology etc) of the various parts of an utterance.

It is worth spelling out the steps that lead to this conclusion. Consider first the perspective of the speaker of a given utterance. As soon as she has completed her utterance, we *need* to assume that she updates her information state with the content of her utterance. She can after all keep the turn and follow up on the initial utterance using material from the initial utterance:

[21] The term *coercion operation* is inspired by work on utterance representation within a type theoretic framework reported in Cooper (1998).

(17) Ariadne: (1) Did Jo leave? (2) (If so,) When?

However, whereas Ariadne is, at least to a first approximation, omniscient about her own utterance,[22] there is of course no guarantee that Bora, the addressee, will manage to combine the information he obtained in processing the utterance with his existing linguistic and background assumptions to yield a coherent content. As we stressed above, Ariadne would find (a range of) requests for clarification from Bora to be entirely coherent and interpretable:

(18) Ariadne: Did Jo leave?
Bora: Jo?

In light of this, we need to assume that Ariadne's information state contains some representation which enables her to interpret and recognize the coherence of a class of possible clarification queries that Bora might make. Conversely, in so far as possible, one would like this representation to be 'independently motivated', i.e. a representation that Ariadne might plausibly be expected to have associated with her information state for her own processing needs. What entities are plausible candidates?

The most parsimonious candidate is of course the content of the utterance since we know that this has to be in Ariadne's information state regardless. The problem is, however, that on most conceptions of content, be they relatively coarse-grained (e.g. possible worlds based) or more fine grained (e.g. as explicated in situation semantics), the content is simply insufficiently structured to enable the requisite 'backtracking' to be defined. Thus, (uttered in a single context) the contents expressed by the sentences in (19a-c) are all taken to be identical, say (19d), in a situation semantics treatment:[23]

(19) a. Ariadne: Jill is the president.

b. Ariadne: [Pointing at Jill] She is the president.

c. Ariadne: That tall woman over there is the president.

d. $\mathsf{prop}(s, \langle\langle \text{President}; \text{j}\rangle\rangle)$

Clearly, (19a–c) can lead to different and pairwise inappropriate clarification queries:

[22] Malapropic speakers undermine this assumption. Their behaviour is, nonetheless, viewed as comical at best or even bizarre.

[23] (19d) denotes the Austinian proposition individuated by a situation s and the SOA $\langle\langle \text{President}; \text{j}\rangle\rangle$.

(20) a. Bora: Who is Jill?

b. Ariadne: She?

c. Bora: Which tall woman?

A potentially more promising candidate than the content of an utterance is its meaning. Meanings in the intended sense were originally introduced by Montague and Kaplan to help explicate the logical properties of indexicals. Meanings were identified with functions from contexts, which provide values for certain parameters (the *contextual parameters*), to contents. A more structured implementation of this notion was provided by work in situation semantics that identifies meanings with restriction bearing abstracts: the variables abstracted over correspond to the contextual parameters; the restrictions provide conditions the parameters must satisfy (e.g. naming information, being the objects of a demonstrative act etc.). We will assume this latter view of meanings, but regardless of the implementation, as we pointed out above, meanings provide a useful notion for conceptualizing grounding/clarification potential. This is because the range of contextual parameters offers a possible characterization of the contextually variable and hence potentially problematic constituents of utterance content:

(21) a. I hear you.

b. f: c $\mapsto$ Hear(s,a,t), where s is the speaker in c, a is the addressee and t overlaps with the time of c.

c. $\lambda c, s, a, t\, Hear(s, a, t)$, where s is the speaker in c, a is the addressee and t overlaps with the time of c.

It is important to stress, in line with our observation above about the verb *leave* (see footnote 20), that if we conceive of meanings as entities which characterize potential sources of misunderstanding, then predicates associated with verbs, common nouns, adjectives, and prepositions will also need to be assumed to project parameters requiring instantiation in context. This of course leads to meanings becoming highly structured objects, far more structured than meanings were conceived on the original Montague/Kaplan view. There are, however, problems with using meanings, even as highly structured as suggested here, to characterize clarification potential. One problem is the familiar one of grain. In terms of the concept or property that they represent, one would be hard pressed to distinguish the meanings of words such as *attorney, lawyer,* and *advocate*.

And yet, since knowledge of language is not uniform, it is clear that the clarification potential of the sentences in (22) is not identical. One can be acquainted with the word *lawyer* but not with *attorney*. Moreover, since the words are distinct phonologically, which word was used initially makes a difference as to how the clarification can be formulated:

(22) a. Ariadne: Jo is a lawyer. Bora: A lawyer?/What do you mean a lawyer?/#What do you mean an advocate?/#What do you mean an attorney?

b. Ariadne: Jo is an advocate. Bora: #What do you mean a lawyer?/An advocate?/What do you mean an advocate?/#What do you mean an attorney?

c. Ariadne: Jo is an attorney. Bora: #What do you mean a lawyer?/#What do you mean an advocate?/What do you mean an attorney?/An attorney?

A related point arises from considering the following punful exchange:

(23) Ariadne: It rained horribly yesterday.
Bora: It?
Ariadne (laughs): Oh the sky, I suppose.

Although Bora's clarification query is sheer smart-aleckry, it is interpretable, despite the fact that the expletive does not, on most accounts contribute an argument filler. This interpretability depends on reference to the utterance pronounced *it*, not simply on whether a meaning parameter is projected.

These data, together with our earlier data on syntactic parallelism exhibited by CE (examples (10)), make it clear that meanings, however liberally we structure them, are not the right entity from which to characterize clarification potential. Pretheoretically, what is needed is a representation which satisfies the characteristics we discussed in Section 1:

- **Fractal heterogeneity**: for each sub-utterance the representation must encode phonological, syntactic, semantic, and contextual information.
- **Sub-utterance accessibility**: Access to all sub-utterances of the utterance is ensured.

Let us dub a representational entity with these properties an *utterance skeleton* – it is skeletal because we assume it still needs to be fleshed out by instantiating its semantic parameters. We will see below that HPSG

signs can be modified to function as utterance skeletons. It still remains to provide grounds that utterance skeletons are 'cognitively motivated', i.e. that the CPs have access to and reason to preserve such representational entities. Let us consider the speaker first. Obviously, given that she has made the utterance, she must in the immediate pre- and aftermath of the utterance possess the information given by the utterance skeleton associated with an utterance. Moreover, quite apart from clarification potential, there is also evidence from presupposition that suggests that the information encoded by a sign is needed. Thus, if Ariadne makes the utterance in (24a(1)), a variety of facts about the utterance becomes presupposed (in bold face in (24(i)–(iii))), as evinced by the possibility of embedding them under a factive-presupposition predicate such as 'interesting'. (24b, c) are similar:

(24) a. Ariadne(1): Did Mark send you a love letter?
Bora(2): No, though it's interesting
(i) **that you refer to Mark/my brother**
(ii) **that you bring up the sending of love letters**
(iii) **that you ask about Mark's epistolary habits** (example taken from (Ginzburg, 1998)).

b. Ariadne(1): Kien left yesterday.
Bora(2a): Aha. **The fact that the second word in your previous utterance starts with 'l'** is interesting.
Bora(2b): **The fact that the last word you uttered is 'yesterday'** is strange – just like the song we're listening to now.

c. "The twinkling of what?", said the King.
"It began with the tea," the Hatter replied.
"Of course twinkling begins with a T" said the King sharply. ((Carroll 1865), p. 108).

Of course many of these potential presuppositions sink without trace, unless explicitly brought out into the open.[24] But these data do suggest that the utterance skeleton associated with an utterance has utility for the speaker, whether the utterance gets grounded immediately or otherwise. What of the addressee? There is a potential problem illustrated by our somewhat careless use of the definite article in talking about *the* utterance skeleton associated with an utterance. Uniqueness is normally correct as far as the speaker goes, but is frequently an unrealistic assumption concerning the addressee. Indeed, even an existential presupposition about a

[24] For a mechanism that guarantees this, see the proposal concerning the structure of presupposed facts in (Ginzburg (forthcoming)).

complete utterance skeleton is not always satisfied for addressees, who might not be able to come up with a complete parse, most obviously due to lexical ignorance or a noisy environment. Since these are common causes for clarification, we need to ensure that we do not define clarification potential in terms of an entity we cannot ensure is available to the addressee.

The upshot of this is that we need to go 'one level higher' and employ some means of underspecifying the utterance skeleton. There are various ways of doing this. One way is to work with descriptions of utterance skeletons – the speaker possesses a definite such description uniquely satisfied by her utterance, whereas the best an addressee might be able to do given his perceptory input and background information is to construct one or more descriptions which are not uniquely satisfiable. A related strategy is described by Milward (2000). He shows how to associate with utterances a *semantic chart*: in close analogy to the charts built up in parsing using context free grammars,[25] a semantic chart is a graph that represents the various states which arise in the parsing and meaning construction of an utterance. (25) illustrates a semantic chart produced for an utterance which a recognizer hypothesised could be either *from Boston to London Heathrow* or *from Bolton to London Heathrow*. Here usage of a single index is used as meta-level disjunction:

(25) 0–1–p: from, 1–2–np:Boston,
1–2–np: Bolton, 2–3–p:to,
3–5–np:London$_{Heathrow}$,
0–2–pp: 0–1–p(1–2–np),
2–5–pp:2–3–p(3–5–np)

In the sequel, we will define the coercion operations and more generally the grounding process on descriptions of utterance skeletons. We believe though that our account could be recast in terms of semantic charts.

Let us summarize this section: we started by informally describing the grounding process in which an utterance gets integrated in an IS or spawns a clarification. One conclusion to emerge from this discussion was that CPs possess as part of their dialogical competence knowledge of *coercion operations*. These are operations by means of which the partial understanding of an utterance *u* can effect an update of an IS in which a clarification question about *u* can be posed. Consideration of the range of the required inputs and outputs of such operations, as evinced by the phenomenon of CE, leads to the conclusion that **updates resulting from**

[25] See (Gazdar and Mellish 1988) for an elementary exposition of this.

utterances cannot be defined in purely semantic terms. We considered what representational entity could serve for the definition of coercion operations. Our conclusion is that such an entity must be one that satisfies the following two properties: (1) **Sub-utterance accessibility**: access to all sub-utterances of the utterance is ensured, and (2) **Fractal heterogeneity**: for each sub-utterance the representation must encode phonological, syntactic, semantic, and contextual information. We have dubbed the requisite representational entity an utterance skeleton. We now turn to provide one concrete modelling of utterance skeletons within the framework of HPSG.

3. Representing Utterances in HPSG

We introduced utterance skeletons as representations that encode in parallel for each sub-utterance down to the word level phonological, syntactic, semantic, and contextual information. The notion of a *sign*, as developed within the framework of HPSG, provides one means of modelling an utterance skeleton, once we make certain relatively minor modifications.[26]

In HPSG words and phrases are modelled as feature structures of type *sign*. Since the features associated with structures of this type include PHON(OLOGY), SYN(TAX)SEM(ANTICS), and C(ON)T(E)XT, the constraints imposed on signs correspond to the general conventions governing the sound-syntax-meaning relation in a given language.[27] (26) illustrates (a simplified version of) the constraints associated with the sentence *Leslie drinks milk*, analyzed as instantiating the type *hd-subj-ph*:

[26] For a fairly elementary introduction to HPSG see (Sag and Wasow, 1999). For a recent version of HPSG see (G&S-00), which we shall be assuming in the sequel, modulo certain modifications we explain below.

[27] It is common in many versions of HPSG to posit a type *synsem* which has associated with it *inter alia* the features *cat(egory)* and *cont(ent)*. The main motivation for this type is to provide for complement selection which typically involves both semantic and syntactic information. We will usually omit this additional feature 'layer' and deal directly with CAT and CONT.

(26)
$$\left[\begin{array}{ll}
\textit{hd-subj-ph} & \\
\textsc{phon} & \langle \boxed{1}, \boxed{3}, \boxed{4} \rangle \\
\textsc{cat} & \textsc{s} \\
\textsc{cont} & \left[\begin{array}{ll} \textit{proposition} & \\ \textsc{sit} & s \\ \textsc{soa} & \boxed{6}\left[\begin{array}{ll} \textit{drink-rel} & \\ \textsc{drinker} & \boxed{l} \\ \textsc{drunk} & \boxed{m} \end{array}\right] \end{array}\right] \\
\textsc{dtrs} & \langle \boxed{0}\left[\begin{array}{lll} \textsc{phon} & \langle \boxed{1}\textsc{Leslie} \rangle & \\ \textsc{cat} & \textsc{np} & \\ \textsc{cont} & \boxed{l} & \end{array}\right], \boxed{2} \rangle \\
\textsc{hd-dtr} & \boxed{2}\left[\begin{array}{ll} \textit{hd-comp-ph} & \\ \textsc{phon} & \langle \boxed{3}, \boxed{4} \rangle \\ \textsc{cat} & \textsc{vp} \\ \textsc{cont} & \boxed{6} \\ \textsc{hd-dtr} & \boxed{5}\left[\begin{array}{lll} \textit{word} & & \\ \textsc{phon} & \langle \boxed{3}\textsc{drinks} \rangle & \\ \textsc{cat} & v & \\ \textsc{cont} & \boxed{6} & \end{array}\right] \\ \textsc{dtrs} & \left\langle \boxed{5}, \left[\begin{array}{lll} \textsc{phon} & \langle \boxed{4}\textsc{milk} \rangle & \\ \textsc{cat} & \textsc{np} & \\ \textsc{cont} & \boxed{m} & \end{array}\right] \right\rangle \end{array}\right]
\end{array}\right]$$

This analysis illustrates in particular the property of fractal heterogeneity we suggested utterance representations need to possess. At the top level of analysis, the type specified as *hd-subj-ph* is specified for PHON (a list length 3 of speech sounds), for CAT (value being S), and for CONT (the proposition that the situation s is of type $\langle\langle Drink; l, m\rangle\rangle$). In addition, since this is a phrasal sign it is also specified for the list-valued feature DTRS, which provides information about the immediate constituents of the phrase. In this case there are two, one of which constitutes the value of the feature HD-DTR, which headed phrases are specified for. Each of the daughters in turn are specified for PHON, CAT, and CONT, the phrasal ones are specified also for DTRS and (the headed ones) for HD-DTR.

In the HPSG version of (G&S-00) one of the fundamental properties of the type *h(eade)d-ph* is the Generalized Head Feature Principle (GHFP):[28]

(27) Generalized Head Feature Principle (GHFP)

hd-ph:

$$\left[\text{SYNSEM } /\boxed{1}\right] \rightarrow \ldots \mathbf{H}\left[\text{SYNSEM } /\boxed{1}\right] \ldots$$

The '/' notation, following Lascarides and Copestake (1999), here indicates a default constraint – specifically one requiring that the SYNSEM value of the mother of a headed phrase and that of its head daughter are identical by default. Specific subtypes of *hd-ph* may override the GHFP, but by formulating (27) in defeasible terms, a statement of interacting constraints on subtypes is all that is required when one wishes to circumvent inheritance from the head daughter.[29]

There is a lot more that can be said about the make-up of HPSG signs, some of which we will indeed say below, when we discuss an HPSG approach to non-sentential utterances. However, for now we wish to stress two aspects about HPSG signs, which require modification if they are to serve the role we envisage for them as utterance skeletons. The first concerns semantics. The content associated with signs, phrasal or lexical, is drawn from a situation theoretic ontology. The ontology distinguishes *inter alia* questions, propositions, facts, situations/events, and outcomes (the denotata of imperatives and subjunctives). These, in turn, are built up as structured objects from situations and SOAs, the latter of which are built from relations and assignments of individuals to the argument roles of the relations. However, existing versions of HPSG make no allowances for a communicative process, i.e. for the need to reason in context with a meaning to obtain a content. Semantic values directly encode contents – via values for the feature CONTENT – that an idealized agent would associate with an utterance. One allowance that is made to contextual dependence is enshrined in the feature C(ONTEXTUAL)-INDICES. Standardly, this encodes information about speaker/hearer/utterance-time. We propose to revamp the treatment of C-INDICES, renaming it in the process C(ONTEXTUAL)-PARAM(ETER)S. This will now encode the entire inventory of contextual parameters of an utterance (proper names, deictic

[28] We adopt a notational convention according to which the head daughter of a construction is denoted with a large bold faced **H**.

[29] Note that one could replace the GHFP with a set of nondefault constraints, each of which specifies the relevant identities on particular subtypes of *hd-ph*. This use of defaults is thus in principle abbreviatory.

pronouns, indexicals).[30,31] This modification of C-INDICES will allow signs to play a role akin to the role associated with 'meanings', i.e. to function as abstracts with roles that need to be instantiated. Thus, the typed feature structure (TFS) in (28a) can be construed as a meaning given in terms of the restricted simultaneous abstract in (28b):

(28) a.
$$\left[\begin{array}{l}
\textsc{c-params} \left\{\begin{array}{l}
\left[\begin{array}{ll}\textsc{index} & \mathrm{b}\\ \textsc{restr} & \{\mathit{named}(\mathrm{Bo})(\mathrm{b})\}\end{array}\right],\\
\left[\begin{array}{ll}\textsc{index} & \mathrm{t}\\ \textsc{restr} & \{\mathit{precedes}(\mathrm{t},\mathrm{k})\}\end{array}\right],\\
\left[\begin{array}{ll}\textsc{index} & \mathrm{i}\\ \textsc{restr} & \{\mathit{spkr}(\mathrm{i})\}\end{array}\right], \left[\begin{array}{ll}\textsc{index} & \mathrm{j}\\ \textsc{restr} & \{\mathit{addr}(\mathrm{j})\}\end{array}\right],\\
\left[\begin{array}{ll}\textsc{index} & \mathrm{k}\\ \textsc{restr} & \{\mathit{utt\text{-}time}(\mathrm{k})\}\end{array}\right], \left[\begin{array}{ll}\textsc{index} & \mathrm{s}\\ \textsc{restr} & \{\}\end{array}\right]
\end{array}\right\}\\
\textsc{cont} \left[\begin{array}{ll}
\mathit{ask\text{-}rel} & \\
\textsc{asker} & \mathrm{i}\\
\textsc{asked} & \mathrm{j}\\
\textsc{msg-arg} & \left[\begin{array}{ll}
\mathit{question} & \\
\textsc{params} & \{\}\\
\textsc{prop} & \left[\begin{array}{ll}\textsc{sit} & \mathrm{s}\\ \textsc{soa} & \left[\begin{array}{ll}\mathit{leave\text{-}rel} & \\ \textsc{agt} & \mathrm{b}\\ \textsc{time} & \mathrm{t}\end{array}\right]\end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]$$

[30] Indeed, in principle, relation names should also be included, since they vary with context and are subject to clarification as well. Such a step involves a significant change to how argument roles are handled in existing HPSG. Hence, we do not make such a move here.

[31] There are many contextual factors implicated in a variety of semantic phenomena that we have to abstract away from here for reasons of space. Most prominent among these concerns some notion akin to 'discourse referent' needed for the treatment of pronominal anaphora. We return to this latter issue in Section 7.

b. $\lambda\{b, t, i, j, s\}$
$[named(Bo)(b), precede(t, k), spkr(i),$
$addr(j), utt - time(k)]$
$\mathsf{prop}(s, \langle\langle Leave; b, t\rangle\rangle)$

Restricted simultaneous abstracts such as (28b) can be construed in a number of ways. They can be construed as structured objects within a situation theoretic universe (see e.g. (Seligman and Moss 1997; Ginzburg and Sag 2000)). Alternatively, they can be construed in a type theoretic framework as functions from records to record types (see e.g. (Cooper 1998)). In both formalizations the intuition is that the restrictions/record provide(s) a characterization of the values entities instantiating the meaning must satisfy.

The values of C-PARAMS get amalgamated via lexical heads and are propagated in entirely analogous fashion to non-local features such as SLASH, WH, BCKGRD (see (G&S-00)). Thus, we assume that C-PARAMS is a feature subject to the (defeasible) constraint in (29): this constraint specifies that a given word's value for a non-local feature F is the union of the values that word's arguments – specified by the feature ARG(UMENT)-ST(RUCTURE) take for F:

(29) Non-LOCAL Amalgamation Constraint

For every non-LOCAL feature **F**:

$$word \Rightarrow / \left[\begin{array}{l} \mathbf{F}\ \boxed{\Sigma_1} \cup \ldots \cup \boxed{\Sigma_n} \\ \text{ARG-ST}\ \left\langle [\mathbf{F}\ \boxed{\Sigma_1}], \ldots, [\mathbf{F}\ \boxed{\Sigma_n}] \right\rangle \end{array}\right]$$

The inheritance of C-PARAMS specifications (a set of parameters), will be successively inherited from head daughter to mother within headed structures by the GHFP. Moreover, since the GHFP is a default constraint, we leave open the possibility that some construction might override the default, adding appropriateness conditions of its own to the set of pooled contextual parameter conditions. One instance of this will be seen below in our treatment of root clauses.

The second modification we make concerns the encoding of phrasal constituency. Standardly, as we have seen above, the feature DTRS is used to encode immediate phrasal constituency. To facilitate statement of coercion operations, we need access to all phrasal constituents – given that a contextual parameter emanating from deeply embedding constituents are as clarifiable as immediate constituents. We posit a set valued feature CONSTIT(UENT)S whose value is the set of all constituents immediate or

otherwise of a given sign.[32,33] The value of CONSTITS is calculated by means of the following constraint, which is analogous, if not identical to the *Non-local Amalgamation Constraint*:[34]

(30) CONSTITS Amalgamation Constraint:

$$\textit{phrase} \Rightarrow \left[\begin{array}{l} \text{CONSTITS } \left\{\boxed{1},\ldots,\boxed{n}\right\} \cup \boxed{\Sigma_1} \cup \ldots \cup \boxed{\Sigma_n} \\ \text{DTRS } \left\langle \boxed{1}[\text{CONSTITS } \boxed{\Sigma_1}], \ldots, \boxed{n}[\text{CONSTITS } \boxed{\Sigma_n}] \right\rangle \end{array}\right]$$

In an attempt both to save space and to make our examples more readable for those not familiar with HPSG attribute-value matrix (AVM) notation, we will frequently use certain abbreviations throughout. These are shown in (31):[35]

[32] In fact, having posited CONSTITS one could eliminate DTRS: this by making the value of CONSTITS be a set of sets whose first level elements are the immediate constituents. Such a move could have interesting theoretical consequences, for instance for the treatment of non-local dependencies, as suggested to us by Berthold Chrysmann. For current purposes, we stick with tradition and tolerate the redundancy of both DTRS and CONSTITS.

[33] A related strategy, which inspired ours, is that proposed in Gregory and Lappin (1999). As part of a search procedure utilized by a post-parsing module which effects ellipsis resolution, Gregory and Lappin define a procedure whose essence is to calculate the value of CONSTITS for a given sign. In contrast, we encode this information directly in the utterance representation. Given that signs represent information about utterances which constitute part of a CP's linguistic competence, then in so far as information about (not solely immediate) constituents is manipulated as part of this competence, it seems preferable to encode this information in the utterance representations, rather than appeal to a post-parsing module whose theoretical status is unclear.

[34] Signs of type *word* are assumed to have an empty value for CONSTITS.

[35] In our account questions are treated in terms of feature structures like (i), where the components are a set of PARAMETERS and a (parametric) proposition; the PARAMS set is empty for a polar question and non-empty for a *wh*-question:

(i)
$$\left[\begin{array}{ll} \textit{question} & \\ \text{PARAMS} & \{\pi, \ldots\} \\ & \left[\begin{array}{ll} \textit{proposition} & \\ \text{SIT} & s \\ \text{SOA} & soa \end{array}\right] \\ \text{PROP} & \end{array}\right]$$

Thus, the careful reader will notice that the abbreviations we introduce involve a number of abuses, including blurring the distinction between SOAs and propositions, and between an index and a parameter (i.e. a restriction bearing index). It is easy to restore these distinctions in context.

(31) HPSG AVM Abbreviations:

AVM	Abbreviation
$\begin{bmatrix} \textit{parameter} \\ \text{INDEX} \quad \text{x} \\ \text{RESTR} \quad \left\{ \begin{bmatrix} \text{INSTANCE} & \text{x} \\ \text{PROPERTY} & \text{P} \end{bmatrix} \right\} \end{bmatrix}$	$x : property(x, P)$
$\begin{bmatrix} \textit{proposition} \\ \text{SOA} \mid \text{NUCLEUS} \quad \begin{bmatrix} \textit{verb\_rel} \\ \text{ROLE\_1} \quad \text{x} \\ \text{ROLE\_2} \quad \text{y} \end{bmatrix} \end{bmatrix}$	$verb(x, y)$
$\begin{bmatrix} \textit{question} \\ \text{PARAMS} \quad \{\,\} \\ \text{PROP} \quad verb(x, y) \end{bmatrix}$	$?.verb(x, y)$
$\begin{bmatrix} \textit{question} \\ \text{PARAMS} \quad \{x : property(x, P)\} \\ \text{PROP} \quad verb(x, y) \end{bmatrix}$	$?x.verb(x, y)$ or $?x : property(x, P).verb(x, y)$

An example of the format for signs we employ is given in (32): Within the phrasal type system of (G&S-00) *root-cl* constitutes the 'start' symbol of the grammar. In particular, phrases of this type have as their content an illocutionary operator embedding the appropriate semantic object (an assertion embedding a proposition, a query embedding a question etc.). Note that *root-cl* specifies an 'idiosyncratic' inheritance pattern for C-PARAMS that overrides the GHFP: the C-PARAMS value of the mother is identical to the union of the C-PARAMS value of the daughter with a set consisting of parameters for the speaker, the addressee, and the utterance time.[36]

[36] Here and throughout we omit various features (e.g. STORE, SLASH etc.) that have no bearing on current issues wherever possible.

(32)
$$\begin{bmatrix} \textit{root-cl} \\ \textsc{phon}\ \text{did bo leave} \\ \textsc{cat}\ \text{V[+fin]} \\ \textsc{c-params}\ \left\{ \begin{array}{l} \text{b:}\textit{named}\text{(Bo)(b), t:}\textit{precedes}\text{(t,k),i:}\textit{spkr}\text{(i),} \\ \text{j:}\textit{addr}\text{(j), k:}\textit{utt-time}\text{(k)} \end{array} \right\} \\ \textsc{cont}\ \textit{ask}\text{(i,j,?.}\textit{leave}\text{(b,t))} \\ \textsc{constits}\ \left\{ \begin{array}{l} \boxed{4}\begin{bmatrix}\textsc{phon}\ \text{Did}\end{bmatrix}, \boxed{5}\begin{bmatrix}\textsc{phon}\ \text{Bo}\end{bmatrix}, \\ \boxed{6}\begin{bmatrix}\textsc{phon}\ \text{leave}\end{bmatrix}, \boxed{7}\begin{bmatrix}\textsc{phon}\ \text{Did Bo leave}\end{bmatrix} \end{array} \right\} \end{bmatrix}$$

Before we can explain how these representations can feature in dialogue reasoning and the resolution of CE, we need to sketch briefly the approach to dialogue ellipsis that we assume.

4. Contextual Evolution, Clauses, and Ellipsis Resolution

4.1. *Context in Dialogue*

We adopt the situation semantics based theory of dialogue context developed in the KOS framework (Ginzburg 1996; Ginzburg forthcoming; Bohlin et al. 1999; Larsson 2002). In Ginzburg (1997a, b, 2001b) Ginzburg demonstrates the existence of intrinsic asymmetries in context between speaker and addressee w/r to ellipsis resolution of bare wh-phrases. In (33a), 'why' must pick up on a fact that positively resolves the initial question A poses, whereas when 'why' is uttered by a new speaker, as in (33b), the resolution is to a fact characterizing A's initial utterance. Note that these data cannot be explained merely as a consequence of the differing coherence of an utterance depending on who makes the utterance: the resolution unavailable to A in (33a) *is* coherent and entirely plausible when it arises from a non-elliptical utterance, whose resolution is not so heavily reliant on context, as in (33c):

(33) (a) A: Where was your Grandmother's sister born? Why? (Unambiguously: 'Why was she born *there*?')
(b) A: Where was your Grandmother's sister born? B: Why? ('Why do you ask where she was born?')
(c) A: Where was your Grandmother's sister born? (and) Why am I asking this question?

Phenomena such as this, in which one CP's contextual possibilities are distinct from another CP's, suggest that a single "context" is not fully adequate to describe dialogue, even when talking about "public" context, which results from overtly registered conversational actions. The approach to context common in formal semantics following Stalnaker (1978) needs to be recast somewhat so that the state of the dialogue at a given point is given in terms of the collection of individual information states of the CPs. This does not necessitate a solipsistic approach, given the considerable evidence, both semantic and psycholinguistic, that CPs try to maintain a common view of the conversation and its background. This, along with other important insights, captured by Stalnaker-inspired presupposition theory (Stalnaker 1978) and Clark-inspired grounding theory (Clark 1996). Hence, KOS posits, following work in the tradition of dialogue games that conversational rules involve updates by each CP of her own *dialogue-gameboard* (DGB), a quasi-public informational repository (cf. Hamblin's *individual commitment slate*, (Hamblin 1970)). This allows conversational action to be viewed as operating on a publically accessible domain which is relative to each CP, and so parametrizable by unpublicized factors such as individual goals and intentions.

In KOS the DGB is construed as a data structure comprising the following attributes: **FACTS**: a set of facts corresponding to the information taken for granted by the CPs,[37]; **QUD** ('questions under discussion'): a set consisting of the currently discussable questions, partially ordered by $\prec$ ('takes conversational precedence'); **LATEST-MOVE**: content of *latest move* made:[38]

(34)
$$\begin{bmatrix} \text{FACTS} & \textit{set of facts} \\ \text{LATEST-MOVE} & \textit{(illocutionary) fact} \\ \text{QUD} & \textit{p.o. set of questions} \end{bmatrix}$$

Both querying and assertion involve a question becoming maximal in the querier/asserter's QUD: the posed question q for a query where q is posed, the polar question $p?$ for an assertion where p is asserted. Given this, we can define adjacency pair relations for dialogue moves: an adjacency pair to a query which poses q is a q-specific utterance; an adjacency pair to an assertion p is either a $p?$-specific utterance or an acceptance

[37] Given presupposition data discussed in Asher (1993) one might postulate that this set is closed under meets and joins.

[38] The type of value LATEST-MOVE will be modified in Section 6 in line with the need for updates to consist of heterogeneous information, as discussed in Section 2.

move, in which a CP updates her FACTS structure with $fact(p)$.[39] A dialogue participant can downdate $q/p?$ from QUD when, as far as her (not necessarily public) goals dictate, sufficient information has been accumulated in FACTS. As an example, we can offer the following (simplified) analysis to the dialogue in (35):

(35) A(1): Who's coming tomorrow?

B(2): Several colleagues of mine.

A(3): I see.

B(4): Mike is coming too.

A's initial query in which she poses the question q_1 causes an update in her QUD. B takes up the question and, hence, updates his QUD. This gives him the opportunity to respond by asserting a proposition p_1 which provides information about q_1. QUD is thus updated with the issue $p_1?$, which becomes the maximal element in QUD. A accepts the assertion, thereby updating her FACTS with the fact corresponding to p_1 and downdating $p_1?$ from QUD.[40] q_1 becomes maximal in QUD again, which licenses providing more information about this question, as B does in (4):[41]

(36) (1): A | QUD:= q_1
(2): B | QUD: q_1; asserts p_1 About q_1: B | QUD:= $q_1 \prec p_1?$
(3): A | QUD:= $q_1 \prec p_1?$
accepts p_1:
A | FACTS:= fact(p_1);
Downdates p? from QUD:= A | QUD: q_1
(4): given A's acceptance:
B | FACTS:= fact(p_1);
downdates $p_1?$ from QUD: B | QUD:= q_1;
asserts p_3 About q_1;
B | QUD:= $q_1 \prec p_3?$

[39] If q is a question, a q-specific utterance is defined as follows: it is an utterance which either provides information σ About q or poses a question q_1 on which q Depends. Here About is partial answerhood and Depends is a relation between questions, which intuitively corresponds to the notion of 'is a subquestion of'. For more on these relations see GS-00.

[40] This discussion is based on the assumption that facts and propositions are ontologically distinct. This is inessential for current purposes, in which case fact(p_1) would simply be p_1.

[41] We omit throughout here the specifications for LATEST-MOVE. Note also that A | QUD represents the value QUD takes in A's Dialogue Gameboard, whereas B | QUD represents the value QUD takes in B's; similarly A | FACTS etc.

In (G&S-00) this framework is integrated into HPSG.[42] (G&S-00) define two new attributes within the CTXT feature structure: Maximal Question Under Discussion (MAX-QUD), whose value is of sort *question*, and Salient Utterance (SAL-UTT), whose value is a set (singleton or empty) of elements of type *sign*. In information structure terms, SAL-UTT can be thought of as a means of underspecifying the subsequent focal (sub)utterance or as a potential *parallel element* (in the sense of (Dalrymple et al. 1991)). MAX-QUD corresponds to the ground of the dialogue at a given point. Since SAL-UTT is a *sign*, it enables one to encode syntactic *categorial* parallelism and, as we will see below, also phonological parallelism. SAL-UTT is computed as the (sub)utterance associated with the role bearing widest scope within MAX-QUD.[43] Before we can explain and exemplify how ellipsis is described and indeed how to extend this account of parallelism to clarification queries, we need to explain how clauses are characterized in the framework of (G&S-00).

4.2. *Declarative and Interrogative Clauses*

Sag (1999) pioneered an approach to characterizing phrases in which information about phrases is encoded by cross-classifying them in a multi-dimensional type hierarchy. Phrases are classified not only in terms of their phrase structure schema or X-bar type, but also with respect to a further informational dimension of CLAUSALITY. Clauses are divided into *inter alia* declarative clauses (*decl-cl*), which denote propositions, and interrogative clauses (*inter-cl*) denoting questions. These are specified as in (37, 38):

(37) *decl-cl*:

$$\left[\text{CONT}\ \left[\begin{array}{l}\textit{proposition}\\ \text{SOA}\ \ /\boxed{1}\end{array}\right]\right] \rightarrow \ldots\ \mathbf{H}\left[\text{CONT}\ \ /\boxed{1}\right]\ldots$$

(38) *inter-cl*:

$$\left[\begin{array}{ll}\text{STORE} & \boxed{\Sigma_1}\\ \text{CONT} & \left[\begin{array}{ll}\textit{question} & \\ \text{PARAMS} & \boxed{\Sigma_2}\end{array}\right]\end{array}\right] \rightarrow \ldots\ \mathbf{H}\left[\text{STORE}\ \ \boxed{\Sigma_1} \uplus \boxed{\Sigma_2}\right]\ldots$$

[42] See (Ginzburg et al. 2001) for a description of SHARDS, a computational implementation of this grammar.

[43] For unary *wh*-interrogatives, SAL-UTT is the *wh*-phrase associated with the PARAMS set of the question; otherwise, its possible values are either the empty set or the utterance associated with the widest scoping quantifier in MAX-QUD.

(37) implies that all declarative clauses denote a proposition such that (by default) its SOA value is identical to the content of its HD-DTR. (38) implies that all interrogative clauses denote a question and requires that the STORE value of a *inter-cl* be the head daughter's STORE value, minus some set of parameters that are included in the clause's PARAMS set. Note that the set of retrieved parameters is intentionally allowed to be the empty set, a move whose consequences will be apparent below. Each maximal phrasal type inherits from both these dimensions. This classification allows specification of systematic correlations between clausal construction types and types of semantic content.

We mention three subtypes of *inter-cl*: *ns-wh-int-cl* is used to generate the familiar extracted (non-subject) *wh*-interrogatives.[44]

(39) *ns-wh-int-cl*:

$$\left[\begin{array}{ll} \text{CONT} & \left[\begin{array}{ll} \textit{question} & \\ \text{PARAMS} & \boxed{\Sigma_2}(\{\boxed{1}\}) \\ \text{PROP} & \boxed{2} \end{array}\right] \\ \text{STORE} & \boxed{\Sigma_1} \\ \text{SLASH} & \{\} \\ \text{WH} & \{\} \end{array}\right] \rightarrow \left[\begin{array}{ll} \text{LOC} & \boxed{4} \\ \text{WH} & \{\boxed{1}\} \end{array}\right], \mathbf{H}\left[\begin{array}{ll} \text{CONT} & \boxed{2} \\ \text{STORE} & \boxed{\Sigma_1} \uplus \boxed{\Sigma_2} \\ \text{SLASH} & \{\boxed{4}\} \\ \text{WH} & \{\} \end{array}\right]$$

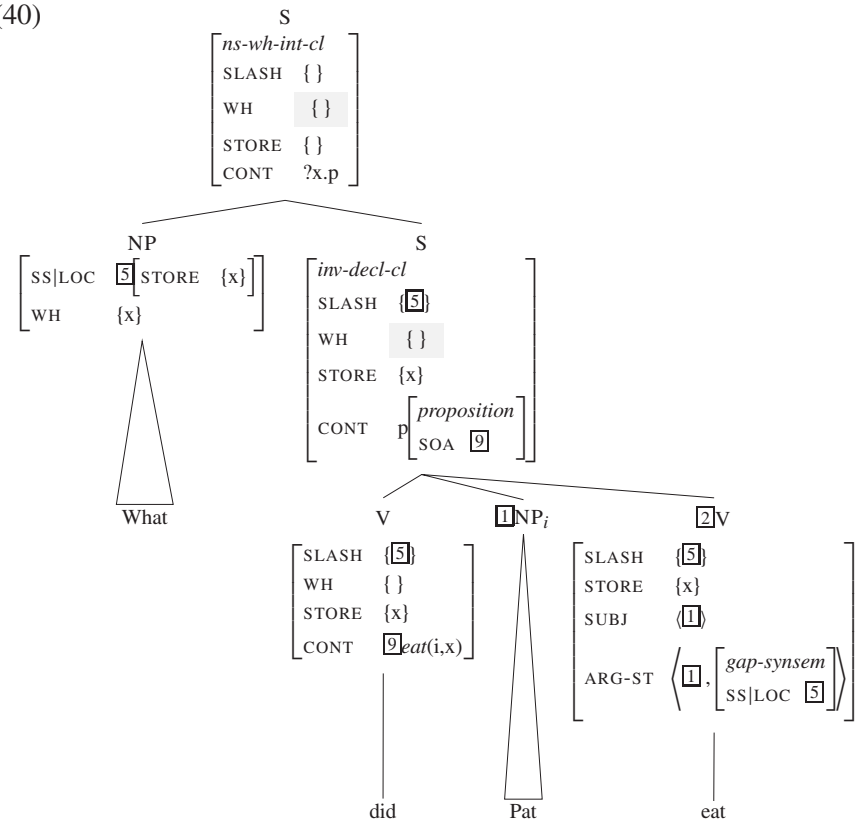
In common with earlier GPSG and HPSG analyses, (39) analyzes extraction in terms of a SLASH dependency: the head daughter's SLASH set is identified with the LOCAL value of the filler daughter. In addition, the constraint embodies the effects of the constraint (38) on *inter-cl*: the PARAMS value of the clause arises by retrieving from storage a subset of the stored parameters. The retrieved parameters must include at least the parameter

[44] The presentation here is quite simplified relative to (G&S-00). We abstract away here from various details that pertain to primarily syntactic issues e.g. auxiliary inversion and pied piping. We also state constraints as stipulations on maximal types rather than deriving them through type inference, as (G&S-00) do.

associated with the filler daughter, which constitutes that daughter's value for the feature WH.

As an illustration of this constraint, consider (40): the filler daughter of this construction is the *wh*-phrase *what*. By the constraint in (39), the parameter *x* associated with this phrase and which is stored at the filler daughter must be a member of the PARAMS set of the question which is the content of (40).[45] The (open) proposition of this question is, by the constraint in (39), identical to the content of the S node *did Pat eat*. The value *what* gets assigned here for the feature SYNSEM|LOCAL (SS|LOC), 5, (the feature bundle comprising its syntactic category, content and store) is identified with the value the S node *did Pat eat* gets for SLASH. This value is ultimately constrained to unify with the unrealized object of the verb *eat*, the second element of the ARG(UMENT)-ST(RUCTURE) of *eat*. This latter is unrealized because it bears the synsem type *gap-synsem*.

(40)



[45] *x* is also the value of the feature WH used *inter alia* to explicate pied piping.

English also allows for 'in situ' interrogatives such as in (41). (G&S-00) argue in detail for a type distinction between *reprise* or echo interrogatives used to seek clarification about a previous utterance (e.g. (41a)) and interrogatives such as (41b) which function as regular information queries:

(41) a. A: What did Pat eat? B: What did WHO eat?

b. A: I'm sending the cakes to the Savoy. B: And the croissants you're sending where?

Common to all 'in-situ' constructions is the fact that they are headed by a finite indicative verb and cannot build questions which serve as the complement of an embedding predicate, i.e. questions where the relevant *in situ wh*-phrases are scoped with non-matrix scope:

(42) a. *We wondered [Dana saw who]. (cf. We wondered [who Dana saw].)

b. *[Sandy visited who] wasn't clear. (cf. [Who Sandy visited] wasn't clear.)

A constraint capturing this characterizes the type *is-int-cl* – also a subtype of *hd-only-ph* (a non-branching headed phrasal type). Reprise interrogatives are appropriate for contexts that arise as a result of the coercion parameter focussing, which we discussed in Section 2 and will be formally presented in Section 5. In this operation MAX-QUD is a question which emerges from a contextual parameter being incorporated into the PARAMS set of a question whose open proposition is identical to the content of the previous utterance. Non-elliptical reprise interrogatives are analyzed in terms of a type *repr-int-cl*, given in (43):[46]

[46] Specifying that the HD-DTR is

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{IC} & + \\ \text{VFORM} & \textit{fin} \end{bmatrix} \end{bmatrix}$$

restricts reprise constructions to unembedded clauses headed by a finite indicative verb, a characteristic of *in situ* constructions in English, as we noted above. I(NDEPENDENT)C(LAUSE) is a Boolean feature positive specification for which means that the clause cannot function as a complement.

(43)
$$\begin{bmatrix} \textit{repr-int-cl} & \\ \text{CONT} & \begin{bmatrix} \textit{question} & \\ \text{PARAMS} & \boxed{\Sigma_3} \\ \text{PROP} & \boxed{1}\begin{bmatrix} \text{SOA|NUCL|MSG-ARG} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{STORE} & \boxed{\Sigma_1} \\ \text{BCKGRD} & \left\{ \textit{prev-utt}(\boxed{0}),\ \textit{cont}(\boxed{0},\boxed{1}) \right\} \uplus \boxed{\Sigma_2} \end{bmatrix}$$
$$\rightarrow$$
$$\mathbf{H}\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{IC} & + \\ \text{VFORM} & \textit{fin} \end{bmatrix} \end{bmatrix} \\ \text{CONT} & \boxed{2} \\ \text{STORE} & \boxed{\Sigma_1} \uplus \boxed{\Sigma_3} \\ \text{BCKGRD} & \boxed{\Sigma_2} \end{bmatrix}$$

To illustrate this: a reprise of (41a) can be performed using (44a). This can be assigned the content in (44b) on the basis of the schema in (43). Thus, to interpret *What did WHO eat?*, one constructs an interrogative of type *ns-wh-int-cl* with the parameter *k* associated with *WHO* remaining in storage. This phrase serves as the head daughter from which the reprise clause is built. Hence, a content emerges corresponding to the content that would be assigned to a non-reprise interrogative *Who$_k$ did you ask what did k eat* (if this were grammatical):

(44)

S

$$\begin{bmatrix} \textit{repr-int-cl} & \\ \text{STORE} & \{\,\} \\ \text{CONT} & ?\text{k:PERSON(k)}.\text{p}_0\text{:}\textit{ask-rel}(\textit{spkr}_0,\text{q}_1\text{:?i:}\textit{thing}(i).\textit{eat-rel}(\text{k,i})) \\ \text{BCKGRND} & \left\{ \textit{prev-utt}(\boxed{0}),\ \textit{cont}(\boxed{0},\text{p}_0) \right\} \end{bmatrix}$$

S

$$\begin{bmatrix} \textit{ns-wh-int-cl} & \\ \text{STORE} & \{\,\text{k}\,\} \\ \text{CONT} & \text{q}_1 \end{bmatrix}$$

NP V NP V

What Did WHO eat

The type *dir-is-int-cl* is used to generate non-reprise in-situ constructions. The only constraint on this type which does not arise by inheritance is that the open proposition of the question content is identified with the content of the head daughter, which is hence required to be propositional:

(45) *dir-is-int-cl*:

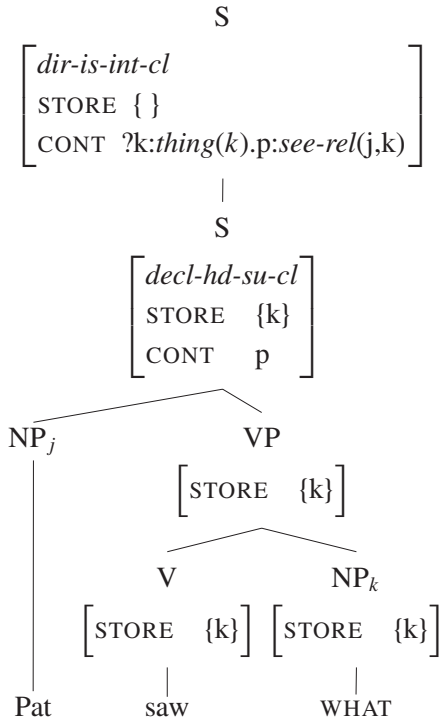
$$\begin{bmatrix} \text{CONT} & \begin{bmatrix} \textit{question} & \\ \text{PARAMS} & \boxed{\Sigma_2} \\ \text{PROP} & \boxed{1} \end{bmatrix} \\ \text{STORE} & \boxed{\Sigma_1} \end{bmatrix}$$

$$\rightarrow$$

$$\mathbf{H}\begin{bmatrix} \text{CONT} & \boxed{1} \\ \text{STORE} & \boxed{\Sigma_1} \uplus \boxed{\Sigma_2} \end{bmatrix}$$

To illustrate:

(46)



4.3. *Elliptical Clauses*

In line with much recent work in HPSG and Categorial Grammar, (G&S-00) do not treat ellipsis by positing a phonologically null head. To account for elliptical constructions such as short answers and sluicing, Ginzburg and Sag posit a phrasal type *decl-frag-cl* – governed by the constraint in (47). With one exception, the various fragments analyzed here will be subtypes of *decl-frag-cl* or else will contain such a phrase as a head daughter.[47]

(47)
$$\left[\begin{array}{ll}
\text{CAT}|\ \text{HEAD} & \left[\begin{array}{ll} v & \\ \text{IC} & + \\ \text{VFORM} & \mathit{fin} \end{array}\right] \\
\text{CONT} & \left[\begin{array}{ll} \mathit{proposition} & \\ \text{SIT} & \boxed{2} \\ \text{SOA} & \left[\begin{array}{ll} \text{QUANTS} & \texttt{order}(\boxed{\Sigma_3}) \oplus \boxed{A} \\ \text{NUCL} & \boxed{5} \end{array}\right] \end{array}\right] \\
\text{STORE} & \boxed{\Sigma_4} \\
\text{MAX-QUD} & \left[\begin{array}{ll} \mathit{question} & \\ \text{PARAMS} & \mathit{neset} \\ \text{PROP} & \left[\begin{array}{ll} \mathit{proposition} & \\ \text{SIT} & \boxed{2} \\ \text{SOA} & \left[\begin{array}{ll} \text{QUANTS} & \boxed{A} \\ \text{NUCL} & \boxed{5} \end{array}\right] \end{array}\right] \end{array}\right] \\
\text{SAL-UTT} & \left\{\left[\begin{array}{ll} \text{CAT} & \boxed{1} \\ \text{CONT}|\text{INDEX} & \boxed{2} \end{array}\right]\right\}
\end{array}\right]$$

$$\rightarrow$$

$$\mathbf{H}\left[\begin{array}{ll}
\text{CAT}|\text{HEAD} & \mathit{nominal} \\
\text{CONT}|\ \text{INDEX} & \boxed{2} \\
\text{STORE} & (\boxed{\Sigma_3} \cup \boxed{\Sigma_4} \mathit{set}(\mathit{param}))
\end{array}\right]$$

This constraint enforces categorial parallelism between the head daughter and the SAL-UTT, as well as coindexing the two. This will have the effect of 'unifying in' the content of the head daughter into a contextually

[47] Our analyses here, as mentioned previously, are applicable only to NP fragments.

provided content. The content of this phrasal type is a proposition: whereas in most headed clauses the content is entirely (or primarily) derived from the head daughter, here it is constructed for the most part from the contextually salient question. This provides the concerned situation and the nucleus, whereas if the fragment is (or contains) a quantifier, that quantifier must outscope any quantifiers already present in the contextually salient question.

To illustrate this, consider the following example:

(48) A: Who left?
B: Jo.

Here A's (sub)utterance of *who* provides the SAL-UTT:

(49)
$$\begin{bmatrix} \text{CAT} & \text{NP} \\ \text{CONT} & \text{x} \end{bmatrix}$$

And the full utterance makes the following question MAX-QUD:

(50) ?x:*person*(x).*leave-rel*(x)

B's utterance thus gets the following analysis:

(51)
$$\begin{bmatrix} \textit{decl-frag-cl} \\ \text{C-PARAMS} \left\{ \text{x: } \textit{named}(\text{x,Jo}) \right\} \\ \text{CONT p:}\textit{leave-rel}(\text{x}) \\ \text{STORE } \{\,\} \\ \text{CTXT} \begin{bmatrix} \text{MAX-QUD } ?\text{x:}\textit{person}(\text{x}).\text{p} \\ \text{SAL-UTT} \left\{ \begin{bmatrix} \text{CAT} & \boxed{6} \\ \text{CONT|INDEX} & \text{x} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

$$\Big|$$

$$\begin{bmatrix} \text{SS|LOC} & \begin{bmatrix} \text{CAT } \boxed{6}\text{NP} \\ \text{CONT|INDEX } \text{x} \end{bmatrix} \end{bmatrix}$$

$$\Big|$$

Jo

5. SIGN COERCION

We now turn to formalizing the coercion operations available to a CP in cases where the available contextual assignment is partial. These were introduced informally in Section 2. We will define here two basic coercion operations on signs. The output of the operations is a partial specification for an utterance in which clarification is sought. We will show subsequently how they can be used as a basis for partially characterizing the clarification potential of utterances and in the grammatical description of the various different readings of CE.

The operations we define have the general form in (52):

(52)
$$\begin{bmatrix} \textit{root-cl} & \\ \text{C-PARAMS} & \{\ldots \text{i} \ldots\} \\ \text{CONSTITS} & \{\ldots \boxed{2}\begin{bmatrix}\text{CONT} & \text{i}\end{bmatrix}\ldots\} \\ \text{CONT} & \textit{illoc-rel}(\text{j,k,msg-arg}) \\ \ldots & \end{bmatrix}$$
$$\Rightarrow$$
$$\begin{bmatrix} \textit{root-cl} & \\ \text{CONT} & \textit{ask-rel}(\text{k,j,?A.p}) \\ \text{CTXT} & \begin{bmatrix}\text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & \text{?B.p}\end{bmatrix} \end{bmatrix}$$

This is to be understood as the following recipe for a clarification request by k of utterance u: given u uttered by CP j (whose associated sign is one) which satisfies the specification in the LHS of the rule, the other CP, k, may respond with any utterance which satisfies the specification in the RHS of the rule.[48] More specifically, the input of the rules singles out a contextual parameter i, which is the content of an element of the daughter set of the utterance $\boxed{2}$. Intuitively, i is a parameter for which the CP either lacks or is dubious about its value. The sub-utterance $\boxed{2}$ is specified to constitute the value of the feature SAL-UTT associated with the context of the clarification utterance cu_0. The descriptive content of cu_0 is a question and it is constrained to share its open proposition with the question which is specified by the rule to constitute MAX-QUD; the

[48] The fact that both the RHS and the LHS of the rule are of type *root-cl* ensures that the rule applies only to signs associated with complete utterances.

sets of parameters abstracted in the two questions, A and B respectively, can but need not be identical.[49] Where the rules differ is with respect to how MAX-QUD gets calculated on the basis of the input.

5.1. *Parameter Focussing*

The first operation we define we dub *parameter focussing*: the essence of the operation involves a problematic contextual parameter becoming a bound parameter of a question about the utterance:

(53) parameter focussing$_i$:

$$\left[\begin{array}{ll} \textit{root-cl} & \\ \text{CTXT-INDICES} & \boxed{1}\{\ldots \text{i} \ldots\} \\ \text{CONSTITS} & \{\ldots \boxed{2}[\text{CONT} \quad \text{i}]\ldots\} \\ \text{CONTENT} & \text{p} \end{array}\right]$$

$$\Rightarrow$$

$$\left[\begin{array}{ll} \text{CONTENT|MSG-ARG} & \text{?A.p} \\ \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & \text{?i.p} \end{array}\right]$$

More specifically, the input of the rule singles out a contextual parameter i, which is the content of an element of the daughter set of the utterance $\boxed{2}$. Intuitively, i is a parameter whose value is problematic or lacking. The sub-utterance $\boxed{2}$ is specified to constitute the value of the feature SAL-UTT associated with the context of the clarification utterance cu_0. The descriptive content of cu_0 is a question, any question whose open proposition p is identical to the (uninstantiated) content of the clarified utterance.[50] MAX-QUD associated with the clarification is fully specified as a question whose open proposition is p and whose PARAMS set consists of the 'problematic' parameter i.

We can exemplify the effect of parameter focussing with respect to clarifying an utterance of (32). The output this yields, when applied to *Bo*'s index, b, is the partial specification in (54). Such an utterance will have as its MAX-QUD a question cq_0 paraphrasable as *who$_b$, named Bo,*

[49] Recall that all signs of type *root-cl* have as their content a proposition (whose SOA value is) of type *illoc-rel*, one of whose subtypes is *ask-rel*. The descriptive content of a root utterance is given in terms of the feature MSG-ARG.

[50] The main relation of this proposition is the illocutionary force of the antecedent utterance, *ask* for a query, *assert* for a proposition etc.

are you asking if b left, whereas its SAL-UTT is the sub-utterance of *Bo*. The content is underspecified:

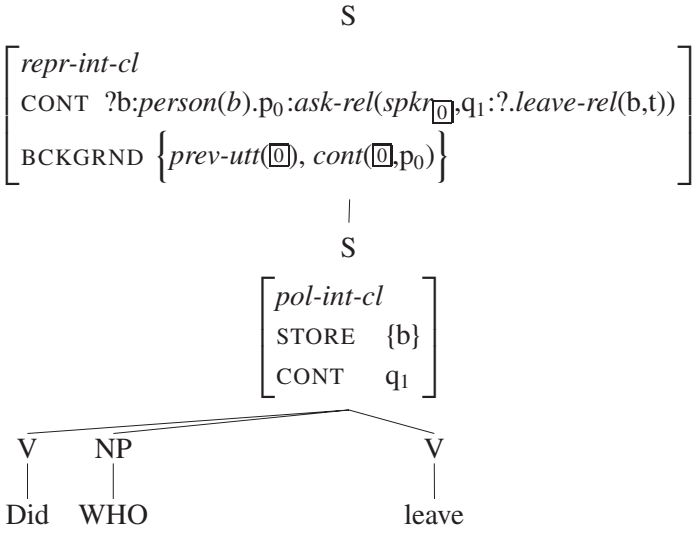
(54) $\begin{bmatrix} \text{CONT|MSG-ARG} & \text{?A.p: ask-rel(i,j,}\textit{leave-rel}\text{(b,t))} \\ \text{SAL-UTT} & \boxed{5} \\ \text{MAX-QUD} & \text{?b:}\textit{named}\text{(Bo)(b).p: ask-rel(i,j,}\textit{leave-rel}\text{(b,t))} \end{bmatrix}$

This (partial) specification allows for clarification questions such as the following:

(55) a. Did WHO leave?
b. WHO?
c. BO? (= Are you asking if BO left?)

Perhaps the simplest example is a reprise sentence such as (55a). The grammar of (G&S-00) associates with such a sentence the content given in (56b), which unifies with the specification provided above in (54):

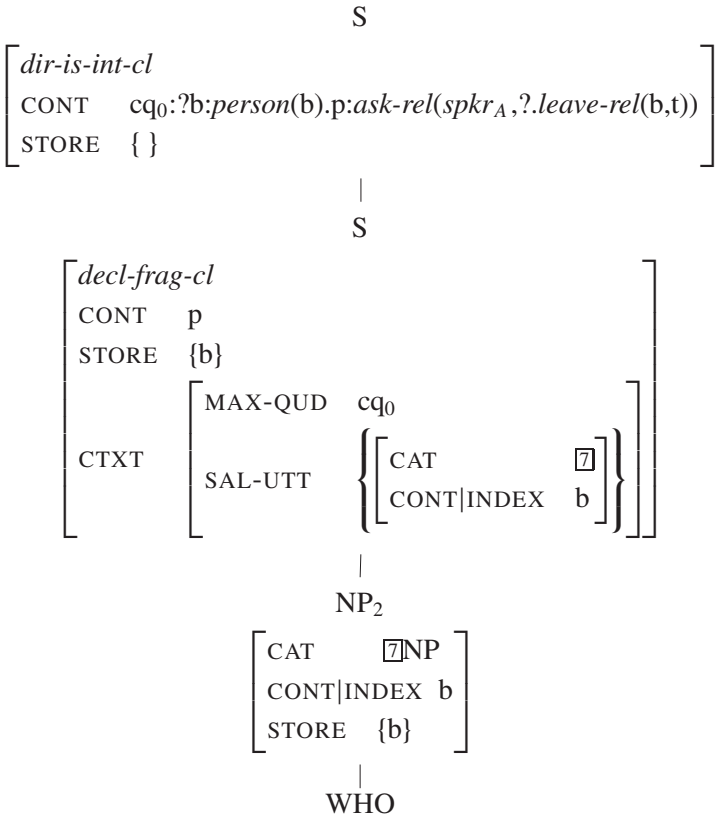
(56) Did WHO leave?



Let us consider (55b): we will analyze this as expressing an identical content to the one expressed by (55a). However, the way this content arises is somewhat different, given the elliptical nature of (55b). The QUD-maximality of cq_0 allows us to analyze the fragment as a 'short answer' to cq_0, using the type *decl-frag-cl*, introduced in Section 4.3 above. More precisely, *decl-frag-cl* enables us to build the proposition of the question,

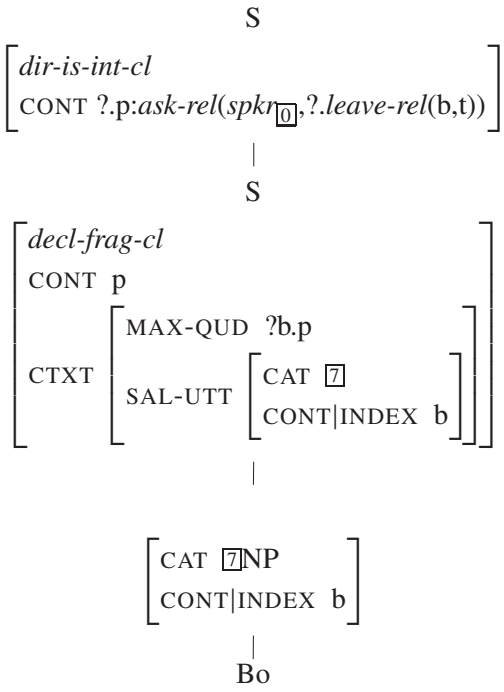
which requires the parameter which constitutes the content of *WHO* to remain in storage. The retrieval of the latter parameter is effected using the type *dir-is-int-cl*, which as we mentioned in Section 4.2 allows a question to be constructed by retrieving from storage zero or more parameters from a proposition–denoting head daughter:

(57)



Let us finally turn to (55c). The analysis we offer for this case is very similar to that offered to (55b). Once again the QUD-maximality of cq_0 allows us to analyze the fragment as a 'short answer' to cq_0, using the type *decl-frag-cl*. And out of the proposition which emerges courtesy of *decl-frag-cl* a question is constructed using *dir-is-int-cl*. The crucial difference is that in the case of (55c) there is no parameter to retrieve from storage – the only 'retrieval' that takes place is vacuous, leading to a question with an empty PARAMS set, in other words a polar question. However, given that this polar question is posed in a context where cq_0 is QUD-maximal, the reading this yields includes a focal component – essentially, *Are you asking if BO, of all people, left*?

(58)



5.2. *Parameter Identification*

The second coercion operation we discussed previously is parameter identification: for a given problematic contextual parameter its output is a partial specification for a sign whose content and MAX-QUD involve a question querying what the speaker intended to convey with the utterance requiring clarification:[51]

(59) parameter identification$_i$:

$$\left[\begin{array}{l} \textit{root-cl} \\ \text{C-PARAMS} \ \{\ldots \text{i} \ldots\} \\ \text{CONSTITS} \ \left\{\ldots \boxed{2}\left[\text{CONT} \quad \text{i}\right]\ldots\right\} \\ \ldots \end{array}\right] \Rightarrow$$

[51] The relation which types MAX-QUD|PROP|SOA in (59) is dubbed *spkr-meaning-rel* to suggest Grice's notion of speaker meaning. This suggestion follows our assumption that the requested clarification targets the entire utterance content, not solely its conventional content.

$$\begin{bmatrix} \text{CONTENT|MSG-ARG} & ?A.p \\ \text{C-PARAMS} & \{\ldots \text{k:}\mathit{addr}(\text{k}) \ldots\} \\ \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & ?c.p:\mathit{spkr\text{-}meaning\text{-}rel}(\text{k},\boxed{2},\text{c}) \end{bmatrix}$$

To exemplify: when this operation is applied to (32), it will yield as output the partial specification in (60):

(60)
$$\begin{bmatrix} \text{CONT| MSG-ARG} & ?A.p \\ \text{C-PARAMS} & \{\text{k:}\mathit{addr}(\text{k})\} \\ \text{SAL-UTT} & \boxed{5}\begin{bmatrix} \text{PHON} & \text{bo} \\ \text{CAT} & \text{NP} \\ \text{CONT|INDEX} & \text{b} \end{bmatrix} \\ \text{MAX-QUD} & ?c.p:\mathit{spkr\text{-}meaning\text{-}rel}(\text{k},\boxed{5},\text{c}) \end{bmatrix}$$

We now show how this specification allows for clarification questions such as the following:

(61) a. Who do you mean Bo?
b. WHO? (= who is Bo)
c. Bo? (= who is Bo)

Utterances such as (61a,c) are tricky in that they contain a constituent *Bo* which is being used in a somewhat 'non-standard' way. By this we mean that in these examples *Bo* is not being used to refer to an individual – the point of these utterances is after all to highlight the inability of the speaker to do so. In some sense these uses of *Bo* involve mention and not use, to use a familiar distinction. However, dubbing these uses as *mention* is not quite right either, at least if by *mention* one means a use in which reference is made simply to the form itself, as in (62):

(62) a. 'Bo' has two letters.

b. 'Bo' is a noun.

The point is that uses such as (61a,c), as in (63), involve an intrinsic anaphoricity:

(63) a. A: Did Bo leave? B: Who is Bo?

b. A: We're fed up. B: Who is we?

In these examples the issue is not *who is Bo/we* **in general**, but *who is Bo/we* **in the sub-utterance needing clarification**. In other words, all these uses involve anaphora to the phonologically identical SAL-UTT, the to–be–clarified sub-utterance. We can accommodate such uses by positing a non-branching phrase type *utt-anaph-ph* that allows NPs to denote the SAL-UTT, assuming phonological parallelism:[52]

(64)
$$\begin{bmatrix} \textit{utt-anaph-ph} & \\ \text{CONT} & \boxed{1}:\boxed{2} = \boxed{1} \\ \text{HD-DTR} & \begin{bmatrix} \text{CAT} & \text{NP} \\ \text{PHON} & \boxed{3} \\ \text{SAL-UTT} & \boxed{2}\begin{bmatrix}\text{PHON} & \boxed{3}\end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Given this, (61a) is quite straightforward. The content arises as a regular extracted *wh*-interrogative, akin to (40):[53]

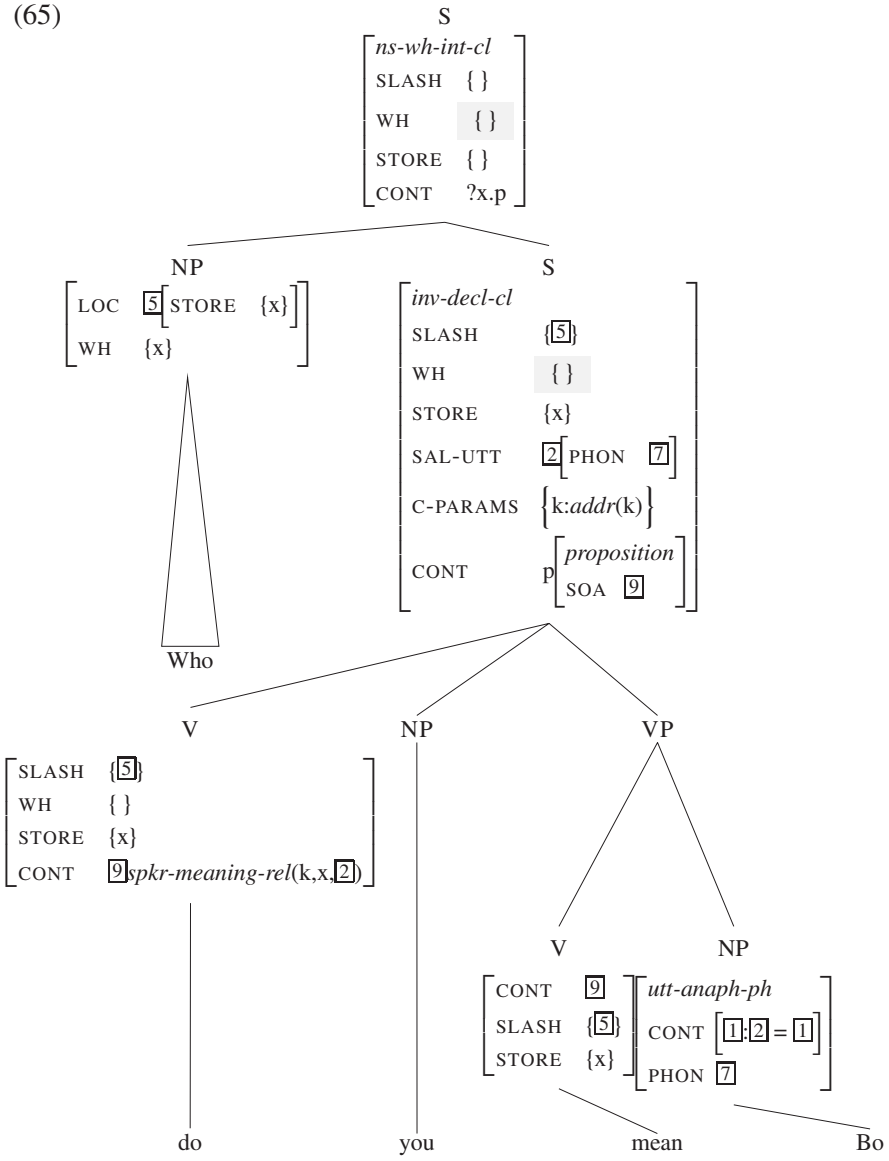
[52] In line with earlier comments, we formulate this rule for NPs only. Note also that for the success of utterance anaphora significant deixis seems to be required, by means of gesture and/or enunciation that somehow mimics the original speaker. We abstract away from this here.

[53] The diagram in (65) illustrates one pitfall of the head–driven approach to semantic composition which we employ here, using as we are HPSG. The content associated here with the auxiliary 'do' is the SOA *spkr-meaning-rel*(k,x,2). This is because in subject-auxiliary constructions, the auxiliary is assumed to be the semantic head, whose SOA value is stipulated to be shared with the mother. Associating this SOA with the auxiliary is somewhat counterintuitive precisely because in an exchange such as

(i) B: Did Bo leave? A: Who do you mean Bo? B: do?

B's CR does not seem to be querying the content *spkr-meaning-rel*(k,x,2). The approach we develop in this paper is entirely compatible with non–head-driven approaches to semantic composition.

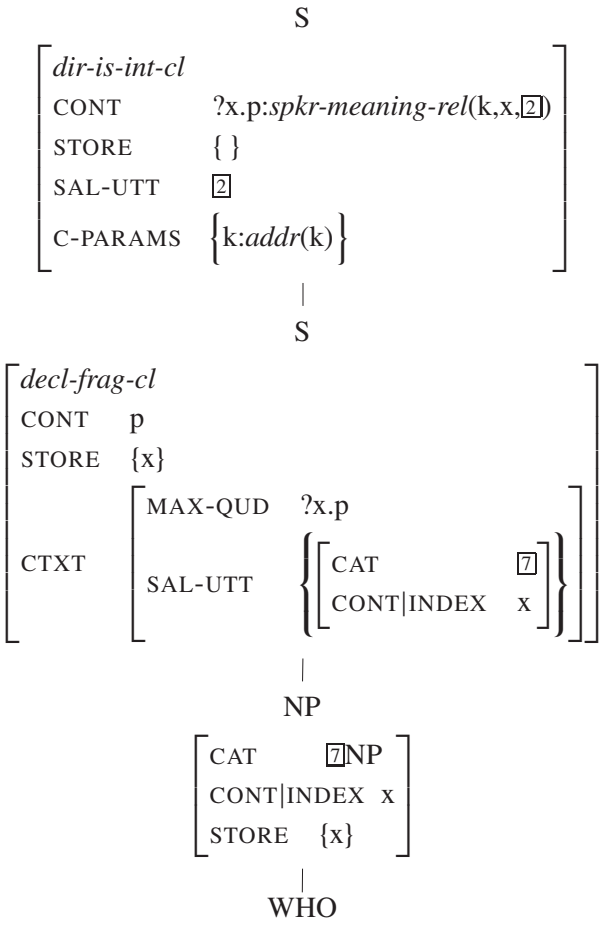
(65)



(61b) arises in a way entirely analogous to (55b):[54]

[54] Note that the analysis we provide can be paraphrased *which person did you mean as the content of the utterance pronounced Bo.*

(66)



Finally we turn to (61c), which is the most interesting but also intricate example. The content we wish to assign to such utterances is in fact identical to the question which constitutes the value of MAX-QUD. Such a content cannot arise using *decl-frag-cl*, the short-answer/reprise sluice phrasal type we have been appealing to extensively, regardless of whether we analyze the NP fragment as denoting its standard conventional content or alternatively as denoting an anaphoric element to the phonologically identical to–be–clarified sub-utterance. Given this, we need to posit a new phrasal type, *constit(uent)-clar(ification)-int-cl*. This will encapsulate the two idiosyncratic facets of such utterances, namely the MAX-QUD/CONTENT identity and the HD-DTR being an *utt-anaph-ph*:[55]

[55] One could reformulate this constraint without assuming the existence of *utt-anaph-ph*, i.e. while employing the fragment's standard conventional content. But this would then require two additional stipulations: one concerning phonological parallelism between the

(67) *constit-clar-int-cl*:

$$\begin{bmatrix} \text{CONT} & \boxed{1} \\ \text{CTXT|\ MAX-QUD} & \boxed{1} \end{bmatrix} \rightarrow \mathbf{H}\left[\textit{utt-anaph-ph}\right]$$

Given this, we can offer the following analysis of (61c):

(68)

$$\begin{array}{c} \text{S} \\ \left[\begin{array}{l} \textit{constit-clar-int-cl} \\ \text{CONT}\ \ \text{cq}_0\text{: ?x.}\textit{spkr-meaning-rel}(\text{k,x,}\boxed{2}) \\ \text{MAX-QUD}\ \ \text{cq}_0 \\ \text{SAL-UTT}\ \ \boxed{2}\left[\text{PHON}\ \ \boxed{7}\right] \\ \text{C-PARAMS}\ \ \left\{\text{k:}\textit{addr}(\text{k})\right\} \end{array}\right] \\ | \\ \text{NP} \\ \left[\begin{array}{ll} \textit{utt-anaph-ph} & \\ \text{PHON} & \boxed{7}\,\texttt{bu} \end{array}\right] \\ | \\ \text{BO} \end{array}$$

5.3. *Extensions*

We now consider some possible extensions and modifications to the coercion rules introduced in this section. The first class of cases is illustrated by B's response in (69a):

(69) Ariadne: Did Bo kowtow?

a. Bora: Bo, hmm. (I'm not sure).

b. Bora: You mean Bro.

The content of an acknowledgement such as (69a) is, in certain respects, even harder to pin down than that of a clarification. This is because such

fragment and the SAL-UTT; the other is a means of ensuring that the C-PARAMS value of the fragment is not inherited by the mother (since the reprise does not involve reference to the conventionally associated referent.). In contrast, positing *utt-anaph-ph* simplifies the constraint as well as providing us with the means of analyzing non-elliptical reprises such as (61a).

an utterance is intended primarily to acknowledge the entire previous utterance, not merely its antecedent sub-utterance, and sometimes is uttered almost as a reflex follow up. Thus, any of the three following contents could be argued to be the content of (69a):

(70) a. I understand your utterance, the one that contained the word pronounced *Bo*.

b. (I notice) You're asking if BO (of all people) kowtowed.

c. (I notice) You're referring to Bo.

We think that, in fact, (69a) does not have a univocal paraphrase – a speaker could be argued to intend any of these three contents in this context; (70a) is entailed by (70b) – the difference between them lying in that the latter commits the acknowledger to a particular analysis of the previous utterance's content. This construal seems convincing in contexts where the acknowledger is, as it were, being reflective about the utterance:

(71) A: Did Bo kowtow?

B: Bo, hmm, good question. (I'm not sure)

There are, however, contexts in which an acknowledgement arises via repetition of a constituent but this need not be understood as involving an intention to highlight that constituent:

(72) (Context: B is a waitress in an Edinburgh diner) A: I'll be having chips and beans and a cappuccino. B: and a cappuccino, OK. (attested example)

We will not offer an explicit account of how (the arguably existing) reading/understanding (70a) emerges. The framework we have introduced here could accommodate (70b,c) fairly easily. The reading in (70b) can be derived if we postulated a variant of parameter focussing. This variant would have the same SAL-UTT and MAX-QUD components as the original parameter focussing. It would differ solely in the content it would associate with (acknowledgement) utterances. The descriptive content would be a

fact rather than a question, embedded under an illocutionary force of, say, *exclaiming* rather than *asking*:[56,57]

(73) parameter focussing$_i$ acknowledgement:

$$\left[\begin{array}{ll} \textit{root-cl} & \\ \text{C-PARAMS} & \boxed{1}\{\ldots \boxed{i} \ldots\} \\ \text{CONSTITS} & \{\ldots \boxed{2}[\text{CONT} \quad \text{i}]\ldots\} \\ \text{CONTENT} & \text{p} \end{array}\right]$$

$$\Rightarrow$$

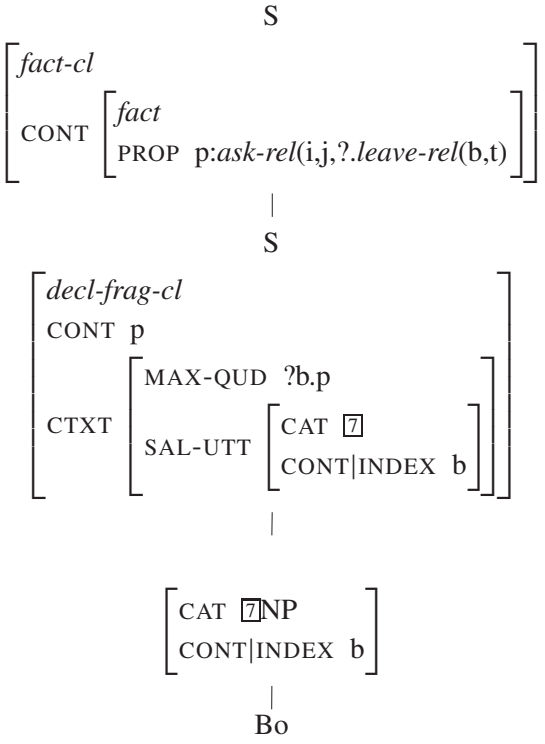
$$\left[\begin{array}{ll} \text{CONTENT|MSG-ARG} & \left[\begin{array}{ll} \textit{fact} & \\ \text{PROP} & \boxed{p} \end{array}\right] \\ \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & ?\text{i.p} \end{array}\right]$$

Given this, (70b) could then be analyzed by means of the short answer type *decl-frag-cl* (cf. (57)) from which a fact would be built using the type *fact-cl*:

[56] In the framework of (G&S-00) a *root-cl* whose MSG-ARG is of type *fact* has its NUCL value resolved to be of type *exclaim-rel* which is a subtype of *illoc-rel*. The type *fact* is appropriate for the feature PROP whose value is of type *proposition*. This reflects the ontological assumption that facts and more generally possibilities are structured objects constructed from propositions. (G&S-00) posit a phrasal type *fact-cl* which allows fact–denoting clauses to be constructed from finite proposition–denoting clauses.

[57] In a more detailed discussion, one would try to collapse the coercion rules posited here in a way which would allow via type inference derivation of the variants which lead to clarification queries and acknowledgements respectively.

(74)



The reading (70c) can be explicated in entirely analogous fashion – the sole change being the postulation of a variant on the rule parameter identification in which descriptive content would be a fact rather than a question, embedded once again under an illocutionary force of *exclaiming* rather than *asking*:

(75) parameter identification$_i$ acknowledgement:

$$\begin{bmatrix} \textit{root-cl} \\ \text{C-PARAMS} \left\{ \ldots \text{i} \ldots \right\} \\ \text{CONSTITS} \left\{ \ldots \boxed{2} \begin{bmatrix} \text{CONT} & \text{i} \end{bmatrix} \ldots \right\} \\ \ldots \end{bmatrix}$$

$\Rightarrow$

$$\begin{bmatrix} \text{CONTENT}|\text{MSG-ARG} & \begin{bmatrix} \textit{fact} \\ \text{PROP} \quad \text{p} \end{bmatrix} \\ \text{C-PARAMS} & \{\ldots \text{k}{:}\textit{addr}(\text{k})\} \\ \text{SAL-UTT} & \boxed{2} \\ \text{MAX-QUD} & ?\text{c.p}{:}\textit{spkr-meaning-rel}(\text{k},\boxed{2},\text{c}) \end{bmatrix}$$

Parameter identification acknowledgement could also be used to explicate responses such as (76b). Here B's response involves the exclamation that the person intended by A as the referent of the sub-utterance A pronounced as *Bo* is actually (the person named) Bro:

(76) a. A: Did Bo kowtow?

b. B: You mean Bro.

Finally, we return to an issue raised in Section 2: the fact that addressees often when encountering a problematic contextual parameter do not generate a clarification but instead simply existentially quantify away the problematic parameter.[58] We need to formulate a means of doing this which can either be applied in case the addressee decides to ground the utterance or alternatively for cases when an addressee decides to existentially quantify away some parameters and seek clarification about others.[59] We achieve this effect by formulating a coercion operation which takes signs of types *root-cl* into type *root-cl*, while altering solely the content of the sign:

(77) contextual existential generalization$_i$:

$$\begin{bmatrix} \textit{root-cl} & & \\ \text{CTXT-PARAMS} & \boxed{1}\{\ldots \text{i} \ldots\} & \\ \text{CONTENT} & \begin{bmatrix} \textit{proposition} & \\ \text{SIT} & \text{s} \\ \text{SOA} & \begin{bmatrix} \text{QUANTS} & \langle\rangle \\ \text{NUCL} & \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

[58] The existence of this possibility has been emphasized particularly in the work of David Israel and John Perry, e.g. (Israel and Perry 1991). They refer to the content in which all contextual parameters are existentially quantified away as the *pure content*, whereas any content stronger than this, where a contextual parameter is instantiated using contextual information is referred to as an *incremental content*.

[59] The need to take account of the latter option was pointed out to us by David Milward.

$$\Rightarrow \begin{bmatrix} \text{CTXT-PARAMS} & \boxed{1} \setminus \{\text{i}\} \\ \text{CONTENT} & \begin{bmatrix} \textit{proposition} & \\ \text{SIT} & \text{s} \\ \text{SOA} & \begin{bmatrix} \text{QUANTS} & \left\langle \begin{bmatrix} \exists\textit{-rel} \\ \text{i} \end{bmatrix} \right\rangle \\ \text{NUCL} & \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

This rule, then, means that for every sign of type *root-cl* one can construct a corresponding sign where a contextual parameter is existentially quantified away with widest possible scope. An addressee who finds a given meaning problematic can utilize this contextually less dependent meaning, either by taking it as the content of the utterance or by using this meaning as input to one of the previously discussed coercion rules (including a "reapplication" of contextual existential generalization.).

Let us exemplify this with reference to our running example (32). Applying contextual existential generalization to the parameter *b* will yield as output the following sign:

(78)
$$\begin{bmatrix} \textit{root-cl} & \\ \text{PHON} & \text{did bo leave} \\ \text{CAT} & \text{V[+fin]} \\ \text{C-PARAMS} & \left\{ \begin{array}{l} \text{t:}\textit{precedes}\text{(t,k),i:}\textit{spkr}\text{(i),} \\ \text{j:}\textit{addr}\text{(j),k:}\textit{utt-time}\text{(k)} \end{array} \right\} \\ \text{CONT} & \exists\text{b:}\textit{named}\text{(Bo)(b). ASK-REL(i,j,}\boxed{q_0}\text{?.}\textit{leave-rel}\text{(b,t))} \\ \text{CONSTITS} & \left\{ \begin{array}{ll} \boxed{4}\begin{bmatrix}\text{PHON} & \text{Did}\end{bmatrix}, & \boxed{5}\begin{bmatrix}\text{PHON} & \text{Bo}\end{bmatrix}, \\ \boxed{6}\begin{bmatrix}\text{PHON} & \text{leave}\end{bmatrix}, & \boxed{7}\begin{bmatrix}\text{PHON} & \text{Did Bo leave}\end{bmatrix} \end{array} \right\} \end{bmatrix}$$

6. Integrating Utterances in Information States

The final component we need is a formal version of the discussion in Section 2 of how utterances get integrated in a CP's information state, leading either to grounding or clarification. We restrict attention essentially to clarification and grounding. Moreover, for simplicity we ignore the

contextual existential generalization coercion operation introduced above. See (Ginzburg forthcoming) for additional motivation of various aspects of the protocol, as well as for a more encompassing protocol that covers acknowledgements and corrections such as those discussed in the previous section.

We formulate this as a protocol – a sequence of instructions to a CP about to update her DGB, from the point where she believes an utterance (hers possibly) has taken place, say a Transition Relevance Point (TRP),[60] and through to the point where she has offered a response. In this way, then, both perspectives, that of the speaker and that of the addressee are covered in a way that accommodates the transient nature of being speaker and addressee. An initial version of the protocol is the following:

(79) **Utterance processing protocol** (initial version)

For an agent A with IS I: if an utterance u is Maximal in PENDING:

(a) Try to:
(1) Find an assignment f in I for σ, where σ is the (maximal description available for) the sign associated with u
(2) Update LATEST-MOVE with u.
(3) React to content(u) according to querying/assertion protocols.
(4) If successful, u is removed from PENDING

(b) Else: make an utterance appropriate for a context such that MAX-QUD and SAL-UTT get values according to the specification in $\mathrm{coe}^i(u, \sigma)$, where coe^i is one of the available coercion operations.

A number of points concerning (79) require further elucidation. The first is that we posit a stack PENDING whose elements are utterances. This is used as a repository for utterances that have yet to be integrated in the DGB, either because they have just occurred or because they need to be set aside while clarification takes place. A second point to note is a speaker/addressee asymmetry, emphasized by the word *try*: we are assuming in line with discussion in Section 2, that it is a fundamental requirement in conversations for a speaker, the author of a given utterance, A, to be aware of the content she intends to convey. Her perspective on the utterance then arises as a special case by assuming (79a) to be vacuously satisfied. Moreover, we explicitly assume that a speaker incorporates her utterance into the DGB as soon as she makes it. Thus, having made her utterance

[60] That is, a point at which, in Conversation Analysis terms, the turn could change.

A needs to update LATEST-MOVE with the content of her utterance, as it is this update that triggers the update moves associated with querying and assertion (e.g. QUD gets updated). These updates must take place immediately following (or simultaneously with) the utterance, if only to allow also for the case where the speaker does not immediately surrender her turn. On the other hand, nothing ensures that for an addressee the precondition associated with (79a) gets satisfied. Hence, it is only addressees for whom the integration of an utterance can give rise to clarification.

A third point which requires elaboration concerns step (79a(2)) – what updating LATEST-MOVE with an utterance amounts to. There are two essential components to such an update, the actual incrementation of LATEST-MOVE and the fate of 'old' values of LATEST-MOVE. As for the former: most speech acts inspired approaches (e.g. the various information state approaches discussed in Traum et al. (1999)) would have the value of LATEST-MOVE be a fact/proposition specifying a speech act. For reasons we have made clear above, the value of LATEST-MOVE actually needs to be an utterance, from which a content of the latter type can be read off. In other words, the value is an utterance skeleton combined with an assignment. In the current formalization this amounts to a pair $\langle \sigma, f \rangle$, σ a sign and f an assignment. The second component of LATEST-MOVE update is backwards looking: should the existing value of LATEST-MOVE become presupposed information? In other words, is there evidence that all CPs believe that the utterance which constitutes the existing value of LATEST-MOVE is *grounded* in the sense of Clark (1996)? If there is, then the existing value of LATEST-MOVE should be added to FACTS, the common ground component of the DGB. With Clark (1996), we take the primary evidence for the grounding of a complete utterance u to be the provision of an adjacency pair response in the sense discussed above:[61]

(80) a. Given a dialogue participant A, an utterance u in PENDING, and A's information state I, I | LATEST-MOVE is grounded if either:
(a) spkr(LATEST-MOVE) $\neq$ A,[62] or
(b) spkr(LATEST-MOVE) = A and u grounds LATEST-MOVE.

b. u grounds LATEST-MOVE if either: (a) content(u) is 'Spkr(u) understands L-M', or:
(b) content(u) is content(L-M)–specific.

[61] In contrast, grounding of parts of an utterance is typically signalled by an affirmative act such as an utterance *yeah, mmh, right* or a corresponding gesture.

[62] Since A was not the speaker of the utterance she integrated into LATEST-MOVE, the fact that she integrated it into her DGB means she could ground it.

Our decision on the type of value of LATEST-MOVE impacts on how 'old' values of LATEST-MOVE increment FACTS: were we to assume the value to be fact/proposition, we could simply assume that if LATEST-MOVE is grounded, it gets added to FACTS in the same way asserted propositions get added to FACTS. However, assuming the value to be the more complex structured object we have argued for means that formulating this incrementation operation is also somewhat more complex. There are two essential choices: either (a) project away the non-semantic information from the utterance information, in which case incrementation reduces to the standard FACTS update or (b) maintain the utterance information intact, while ensuring that the semantic objects that constitute the elements of FACTS are all of a similar ontological nature. For instance, one could assign signs an interpretation as restricted abstracts (not just their semantic components, as we proposed earlier) with the restrictions including as 'presuppositions' not only conventional background information such as naming, but also non-semantic information that characterizes the utterance. One would then assume that all elements of FACTS are abstract/assignment pairs and one would define an appropriate 'merge' operation, many examples of which can be found in the literature on dynamic semantics.[63] The choice between these strategies is partly an empirical issue and partly dependent on the area of intended application. The cognitive psychology literature on memory for discourse provides some evidence for the rapid decay of purely structural information in many but by no means all circumstances.[64] Adding to the mixed picture is the existence of a growing body of work that demonstrates the existence of non-semantically-based syntactic priming (see e.g. (Branigan et al. 2000)). Purver et al. (2002) report that approximately 80% of the clarification requests found in a random sample of the BNC concern the most recent utterance, whereas 96% concern one of the 4 most recent utterances. The latter two facts suggest that non-semantic information associated with an utterance u has decreasing utility for conversationalists the further time passes from u. Nonetheless, *decreasing* does not mean *vanishing* and preservation of a highly structured utterance representation does have utility particularly in circumstances where the potential for misunderstanding is high, with the concomitant need for belief revision, for instance. We believe, therefore, that it is likely to be fruitful to pursue strategy (b) (or some variant thereon). Given that this raises a variety of issues we cannot go into in the current paper, we choose here the more simplistic (a) strategy. This can be formulated simply as follows:

63 See e.g. (Vermeulen 1993; van Eijck and Kamp 1997).

64 See (Fletcher 1994) for a review.

(81) LATEST-MOVE update:

If f is an assignment in information state I for σ, where σ is (the maximal description available for) the sign associated with u

1. If LATEST-MOVE is grounded, then FACTS:= FACTS + content(LATEST-MOVE);
2. LATEST-MOVE := $\langle \sigma, f \rangle$

Let us exemplify this with reference to our running example. A utters (82), which B cannot ground because, say, she cannot resolve the reference of *Bo*. We assume she applies parameter focussing to generate a partial specification for her clarification request. Given this, our protocol would lead to the following respective DGBs:

(82) a. A: Did Bo leave?

b. A's information state:

$$\begin{bmatrix}
\text{MAX-QUD} & q_0 \\
\text{SAL-UTT} & \{\} \\
\text{LATEST-MOVE} & \begin{bmatrix}
\text{SIGN} & \begin{bmatrix}
\textit{root-cl} \\
\text{PHON} \ \text{did bo leave} \\
\text{CAT} \ \text{V[+fin]} \\
\text{C-PARAMS} \ \left\{ \begin{array}{l} \text{b:}\textit{named}\text{(Bo)(b),t:}\textit{precedes}\text{(t,k),i:}\textit{spkr}\text{(i),} \\ \text{j:}\textit{addr}\text{(j),k:}\textit{utt-time}\text{(k)} \end{array} \right\} \\
\text{CONT} \ \text{ASK-REL(i,j,}q_0\text{:?.}\textit{leave-rel}\text{(b,t))} \\
\text{CONSTITS} \ \left\{ \begin{array}{l} \boxed{4}\left[\text{PHON Did}\right], \boxed{5}\left[\text{PHON Bo}\right], \\ \boxed{6}\left[\text{PHON leave}\right], \boxed{7}\left[\text{PHON Did Bo leave}\right] \end{array} \right\}
\end{bmatrix} \\
\text{CTXT-ASSGN} & \begin{bmatrix} \text{b B} \\ \text{t T0} \\ \text{i A} \\ \text{j B} \\ \text{k T1} \\ \text{s S0} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}$$

c. B's information state:

$$\begin{bmatrix}
\text{PENDING} & \begin{bmatrix}
\text{SIGN} & \begin{bmatrix}
\textit{root-cl} & \\
\text{PHON} & \text{did bo leave} \\
\text{CAT} & \text{V[+fin]} \\
\text{C-PARAMS} & \left\{\begin{array}{l}\text{b:}\textit{named}\text{(Bo)(b),t:}\textit{precedes}\text{(t,k),i:}\textit{spkr}\text{(i),} \\ \text{j:}\textit{addr}\text{(j),k:}\textit{utt-time}\text{(k)}\end{array}\right\} \\
\text{CONT} & \text{ASK-REL(i,j,?.}\textit{leave-rel}\text{(b,t))} \\
\text{CONSTITS} & \left\{\begin{array}{l}\boxed{4}\begin{bmatrix}\text{PHON} & \text{Did}\end{bmatrix}, \boxed{5}\begin{bmatrix}\text{PHON} & \text{Bo}\end{bmatrix}, \\ \boxed{6}\begin{bmatrix}\text{PHON} & \text{leave}\end{bmatrix}, \boxed{7}\begin{bmatrix}\text{PHON} & \text{Did Bo leave}\end{bmatrix}\end{array}\right\}
\end{bmatrix} \\
\text{CTXT-ASSGN} & \begin{bmatrix}
\text{t} & \text{T0} \\
\text{i} & \text{A} \\
\text{j} & \text{B} \\
\text{k} & \text{T1} \\
\text{s} & \text{S0}
\end{bmatrix}
\end{bmatrix} \\
\text{MAX-QUD} & \text{?x:}\textit{person}\text{(x).ASK-REL(i,j,?.}\textit{leave-rel}\text{(x,t))} \\
\text{SAL-UTT} & \boxed{5}
\end{bmatrix}$$

(82) has one perplexing and seemingly problematic consequence – A and B have processed the same utterance and yet find themselves in distinct DGB configurations: whereas A has the question of *whether Bo left* as the value of MAX-QUD and the empty set as the value of SAL-UTT, B has the question *who$_t$, named Bo, are you asking if t left* as the value of MAX-QUD and the utterance *Bo* as SAL-UTT. In other words, what has emerged is **a mismatch between the CPs at the level of public context**. This mismatch can be used to explicate the Turn Taking Puzzle (see examples (33)). However, as Ginzburg (1998) notes, this mismatch is exhibited intrinsically on the level of production, but need not arise at the level of comprehension. That is, whereas A cannot intend a fragment she utters if she keeps the turn to be understood as a clarification, she can certainly resolve the ellipsis satisfactorily if B utters the same form and intends it as a clarification:

(83) A(1): Who left Bill? B(2): Bill?

However, as we have set things up so far, A will have only one way of construing (83[2]), namely as an answer to the question she posed; she can also pose a clarification question concerning B's clarification question. But she cannot, in this set up, comprehend B's utterance as a clarification question about her original utterance. In order to enable this possibility, we offer such an additional, backtracking alternative in the protocol, given as

option (b) below. The final version of the protocol can then be formulated as follows:

(84) **Utterance processing protocol** (final version)

For an agent B with IS I: if an utterance u is Maximal in PENDING:

(a) Try to:
(1) find an assignment f in I for σ, where σ is (the maximal description available for) the sign associated with u
(2) update LATEST-MOVE with u:

1. If LATEST-MOVE is grounded, then FACTS:= FACTS + LATEST-MOVE;
2. LATEST-MOVE := $\langle \sigma, f \rangle$

(3) react to content(u) according to querying/assertion protocols.
(4) if successful, u is removed from PENDING

(b) Else: repeat from stage (a) with MAX-QUD and SAL-UTT obtaining the various values of $\text{coe}^i(\tau)|\text{MAX-QUD/SAL-UTT}$, where τ is the sign associated with LATEST-MOVE and coe^i is one of the available coercion operations;

(c) Else: make an utterance appropriate for a context such that MAX-QUD and SAL-UTT get values according to the specification in $\text{coe}^i(u, \sigma)$, where coe^i is one of the available coercion operations.

7. Conclusions and Further Work

This paper has provided an analysis of how requests for clarifications emerge in dialogue, with particular focus on the elliptical construction we have dubbed Clarification Ellipsis (CE). Our analysis takes as a starting point a by now classical view of meanings as functions/abstracts over contextual parameters. It proceeds from this to offer a proposal in which contextual updates involve operations on utterance representations that combine phonological, syntactic, semantic, and contextual information. Our account suggests that the potential for clarification spawned by an utterance depends in part on that utterance's phonological and syntactic structure. Thus, two utterances whose meaning is identical can have distinct clarification potentials if their phonological and syntactic structures

are distinct.[65] In so far as this proposal is correct, it provides evidence for a strongly representationalist perspective on dynamic semantics.

Moreover, the cost of integrating CE into the grammar – at least the grammar as envisioned in a theory like HPSG – has been rather low: it involves positing (a) several phrasal types with independently motivated grammatical features, and (b) the existence of anaphora to utterances, which fall under the rubric of event anaphora, assuming utterances to be a kind of event. This grammar interfaces into a context on which we assume conversationalists can compute coercion operations on signs, the entities we assume to effect contextual updates.

Thus, our account calls into question the servile role Montague reserved for form (i.e. phonology and syntax) by comparison with meaning. This view we have suggested survives in a weakened form in all existing approaches to dynamic semantics. On the other hand, our account stresses the importance of a fractal approach to utterance representations and relies on a constructionist approach to phrasal grammar. The former, a generalization of the 'rule to rule' approach initiated by Montague (1974a), distinguishes recent work in Constraint Based Grammar (e.g. HPSG, Categorial Grammar, and LFG) from transformational approaches. In order to state our coercion rules, we depend on the grammar satisfying **Fractal Heterogeneity**: for each sub-utterance the representation must encode phonological, syntactic, semantic, and contextual information. Constructionism is a feature of recent work in Construction Grammar, Word

[65] Both an anonymous reviewer for *Linguistics and Philosophy* and Manfred Krifka have suggested to us that attempting to characterize clarification potential goes beyond the purview of linguistics proper. The anonymous reviewer writes:

So, yes, to model real dialogue understanding we need to keep track of utterances and their associated contents – to the extent these were resolvable – and partial contents where those exist, but the theory of meaning, even in a dynamic context, is surely about what a fully interpretable meaning does to a context.

We are happy to concede some terminological territory here – albeit the fact that on the standard Montague/Kaplan view *meanings* are not interpreted or interpretable. We have given indications throughout this paper that the potential for CRs is as rule governed as the potential for 'successful, full interpretation'. In both cases it is difficult sometimes to decide exactly how to formulate the rule, but there are clear intuitions about central cases. If someone prefers not to call the task of characterizing CR potential part of a theory of meaning, so be it. This task is, nonetheless, a fundamental part of a theory of linguistic competence that pertains to language understanding, however one wishes to call it. Indeed, this task is closely related to what is by now a sizeable and important literature on underspecification. This latter raises an important theoretical question – equally outside the purview of 'what a fully interpretable meaning does to a context' – what kind of updates can be based on underspecified meanings?

Grammar, and HPSG. We exploit this to internalize ellipsis resolution in a grammar where dialogue context plays a significant role.

Much work remains to be done to extend the work on clarification outlined here. For a start, the account of CE we presented in this paper is applicable directly only to referential NPs. Extending the account to verbs, adjectives, common nouns and other similar words is not straightforward within current versions of HPSG. This is because in those versions each such word directly denotes a type which introduces argument roles as the features appropriate for that type. Allowing for clarification of uses of such words requires them to introduce in some way an instantiable contextual parameter, with concomitant 'underspecification' of the associated argument roles. Such an account has been developed in Purver (2002), who applies this to a computational proposal of how the acquisition of new words can result from dialogue interaction.

In contrast to this, explicating how clarification applies to quantificational NPs, as in (85), seems a more genuinely framework independent problem:

(85) a. A: Is everybody happy?
B: Everybody? (= who counts as everybody)

b. A: Are you going to hire someone for the summer?
B: Someone?
A: A programmer, say, or a corpus hacker.

c. A: Is everyone in the College associated with somebody?
B: Somebody?
(= what function f are you asking if everyone$_x$ in the College is associated with f(x)?)
A: An adviser.

We will not make a proposal here as to how this problem should be tackled.[66] It does, nonetheless, seem clear that an account will involve adopting an approach where QNPs have 'stand alone' meanings (as in Montague (1974b)), given the need to be able to seek clarification of the QNP sub-utterance as such. Data such as (85b,c), however, offer initial indications that the requisite meanings will not turn out to be higher order operators similar to those proposed in Montague (1974b). These clarifications do not seem to concern in some way the denotation of a property of properties; moreover, whether the existential is construed as wide or narrow seems to affect its clarification possibilities. This is not easily

[66] But see (Purver and Ginzburg 2003) for such an account.

consonant with a Montogovian approach, where scope differences simply correspond to order of application of functors.

Finally, we wish to point to the broader perspective which we hope the current work suggests. In Section 6, we discussed two possible approaches concerning how grounded utterances get integrated in the common ground. One approach, which we adopted here for reasons of simplicity, involved projecting the content of an utterance into the FACTS component of the DGB. An alternative we mentioned would be to project a more structured entity, which corresponds to a sign/assignment pair. We believe this latter to be a fruitful strategy for a number of reasons. For a start consider a theory of nominal anaphora in dialogue. Hitherto in this paper, we have considered two types of context dependence: indexicality, which gets analyzed via the meaning/content distinction, and non-sentential ellipsis, which gets analyzed via the dynamics of QUD. We have not said anything about pronominal anaphora and the contextual evolution that underwrites it. This has been the focus of a considerable amount of research for text/monologue, though relatively little for dialogue (see e.g. (Dekker 1997; Asher and Lascarides 1998)).

We believe this is not an accidental lacuna: dialogue brings with it features such as disagreement, distinct illocutionary forces of successive turns, and misunderstanding, which require significant modifications to frameworks such as DRT (Kamp and Reyle 1993) or DPL (Groenendijk and Stokhof 1991b). These frameworks were designed for texts, where such features are not present. To take one example, we are not aware of a treatment within such frameworks of a dialogue such as the following, where B cannot resolve the reference of *Jill*:

(86) A: Did any of my students phone?
B: Possibly.
A: Did Jill phone?
B: Which one is she?

(86) shows that anaphora can arise even in the context of a clarification request, where the speaker lacks a referent for the anaphor, whose antecedent is 'referential' (for the other conversationalist). (Ginzburg 2001a), building on (Milward 1995), sketches an analysis of pronouns, which can cover examples such as (86). This treats pronouns as definite descriptions, whose uniqueness domain is provided by the antecedent utterance situation. Such a situation is introduced by each NP as a contextual parameter, and hence enters into the context when a meaning gets (partially) instantiated in grounding or clarification. An extended discussion of the semantics of pronouns is obviously beyond the scope of the current paper. What we

wish to suggest, nonetheless, is that if there are independent grounds for maintaining structured utterance representations in context, as we have suggested for CE, such representations can serve as underpin of an analysis of pronouns in dialogue. Tying such an analysis to utterance representations (as opposed to purely semantic representations such as DRSs or sets of assignments) has the advantage that it enables one to deal with grammatical gender phenomena, as illustrated in (87). Hebrew, which lacks a neuter gender and correspondingly has no neuter pronouns, has two words which correspond to the English *car*, one is masculine, the other feminine. Subsequent pronominal reference must agree with the gender that occurs in the antecedent utterance:

(87) a. A: ledani yesh óto yafe. B: eyfo hu kana otó/#ota?
A: Dani has car-masc nice-masc. B: Where he bought him/#her?
A: Dani has a nice car. B: Where did he buy it?

b. A: ledani yesh mexonit yafa. B: eyfo hu kana #otó/ota?
A: Dani has car-fem nice-fem. B: Where he bought #him/her?
A: Dani has a nice car. B: Where did he buy it?

Basing anaphoric resolution on utterance situations allows both agreement and referent identity information to be simultaneously in the context for as long as the anaphoric potential exists. Agreement constitutes a problem for dynamic systems where the only information maintained concerns referents, as e.g. in DPL and in DRT.

The final phenomenon we mention is one we brought up in the introduction, namely attitude reports. One of the most discussed puzzles in the Philosophy of Language since the early 1980s is Kripke's Pierre puzzle (Kripke 1979). This concerns the Frenchman Pierre who lives in London but is not aware that *Londres* names the same place. Consequently, without being irrational he can be described as believing that London ("Londres") is pretty and also that London ("London") is not pretty. The puzzle arises from the fact that there are good reasons to think that, semantically, uses of 'London' and 'Londres' have the same content. A conclusion various researchers have drawn from this and related puzzles is the need to individuate agent information states in a finer grained way than solely by their (semantic) content (Crimmins and Perry 1989; Crimmins 1993; Asher 1993). The various accounts of information state structure proposed by philosophical logicians have not offered an explicit dynamic theory of how following an utterance an agent who processes it modifies her existing information state and enters into a new state, characterized in part by the syntactic/phonological aspects of the utterance.

The account of utterance processing developed in previous sections goes some ways towards providing such a theory. Consider the information state of Anais given in (88):

(88) a. Anais: Is London pretty?

b. Anais' information state:

$$\begin{bmatrix}
\text{MAX-QUD} & q_0 \\
\text{SAL-UTT} & \{\} \\
\text{LATEST-MOVE} & \begin{bmatrix}
\text{SIGN } \sigma & \begin{bmatrix}
\textit{root-cl} \\
\text{PHON Iz landon prItI} \\
\text{CAT V[+fin]} \\
\text{C-PARAMS } \left\{\begin{array}{l}\text{l:}\textit{named}\text{(London)(l),i:}\textit{spkr}\text{(i),}\\ \text{j:}\textit{addr}\text{(j),}\\ \text{k:}\textit{utt-time}\text{(k)}\end{array}\right\} \\
\text{CONT ASK-REL(i,j, } q_0\text{:?.}\textit{pretty-rel}\text{(l)} \\
\text{CONSTITS } \left\{\begin{array}{l}\boxed{4}\begin{bmatrix}\text{PHON Iz}\end{bmatrix}, \boxed{5}\begin{bmatrix}\text{PHON landon}\end{bmatrix},\\ \boxed{6}\begin{bmatrix}\text{PHON prItI}\end{bmatrix},\\ \boxed{7}\begin{bmatrix}\text{PHON Iz landon prItI}\end{bmatrix}\end{array}\right\}
\end{bmatrix} \\
\text{CTXT-ASSGN f} & \begin{bmatrix}
\text{b L} \\ \text{i A} \\ \text{j B} \\ \text{k T1} \\ \text{s S0}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}$$

Given this information state, we would be justified in making the statement in (89a) to colleagues in London, but equally the statement in (89b) to colleagues in Paris:

(89) a. Anais has asked whether London is pretty.

b. Anais a demandé si Londres est jolie.

What justifies this is that these two statements have the same content, that semantic object that arises by applying the abstract $\mu(\sigma)$ we can associate with the sign σ (see (28) above) to the contextual assignment f. We can generalize this as follows:

(90) A has asked q in situation s iff there exists an information state I such that I |LATEST-MOVE = $\langle \sigma, f \rangle$ and $\mu(\sigma)[f]$ = $\mathsf{prop}(s, \langle Ask; A, q \rangle)$

Here σ the sign associated with A's utterance plays a role similar to the 'frame of mind' in (Barwise and Perry, 1983) or to a DRS in Asher (1993); in other words, it classifies the internal state of the agent. f represents the external anchoring of the information state.

For many cases, a theory of attitude reports based solely on the content of the information state is workable. One way of understanding pathological cases such as the Pierre puzzle is that, in the limit, such a theory is not quite sufficient. In its stead, one needs attitude reports to involve reference to cognitive particulars. In other words, one needs to relativize attitude relations to information states. This can be done straightforwardly as follows:

(91) A has asked q in s relative to an information state I iff I |LATEST-MOVE = $\langle \sigma, f \rangle$ and $\sigma[f]|CONT = \mathsf{prop}(s, \langle Ask; A, q \rangle)$

Such information states are sufficiently fine grained to offer a straightforward explanation of puzzles like Pierre's, but crucially their structured nature is independently motivated in terms of dialogue processing. Of course, what we have offered here is merely a sketch of how a theory of utterance processing can be applied in the domain of attitude reports. Fleshing it out – e.g. to discuss attitudes like belief as opposed to directly illocutionary ones like asking – would involve in particular development of the structured utterance update strategy we discussed above. However, we hope this all suggests a much wider application of the strategy we employed to analyzing the phenomenon of CE.

ACKNOWLEDGEMENTS

For very useful discussion and comments we would like to thank Pat Healey, Howard Gregory, Shalom Lappin, Dimitra Kolliakou, Manfred Krifka, David Milward, Matt Purver, Ivan Sag, three anonymous reviewers for the ACL 2001 conference, and an anonymous reviewer for *Linguistics and Philosophy*. We would also like to thank Matt Purver for help in using SCoRE and with the abbreviatory notations. Earlier versions of this paper were presented at colloquia at ITRI, Brighton, Queen Mary, London, at the Computer Lab, Cambridge, at the Dept of Computational Linguistics, Saarbrücken, and at the 39th meeting of the ACL in Toulouse. The research described here is funded by grant number R00022269 from the Economic and Social Research Council of the United Kingdom, by INDI (Information Exchange in Dialogue), Riksbankens Jubileumsfond 1997-0134, and by grant number GR/R04942/01 from the Engineering and Physical Sciences Research Council of the United Kingdom.

REFERENCES

Asher, N.: 1993, *Reference to Abstract Objects in English: A Philosophical Semantics for Natural Language Metaphysics*, *Studies in Linguistics and Philosophy*. Kluwer, Dordrecht.

Asher, N. and A. Lascarides: 1998, 'Questions in Dialogue', *Linguistics and Philosophy* **21**.

Barwise, J. and J. Perry: 1983, *Situations and Attitudes*, Bradford Books, MIT Press, Cambridge.

Bohlin, P., R. Cooper, E. Engdahl, and S. Larsson: 1999, 'Information States and Dialogue Move Engines', *Gothenburg Papers in Computational Linguistics*.

Branigan, H., M. Pickering, and A. Cleland: 2000, 'Syntactic Coordination in Dialogue', *Cognition* **75**, 13–25.

Bresnan, J.: 2000, *Lexical Functional Syntax*, Blackwell, Oxford.

Carroll, L.: 1865, *Alice in Wonderland*, Mad Hatter Publications, London.

Chierchia, G.: 1995, *Dynamics of Meaning*, University of Chicago Press, Chicago.

Chomsky, N.: 1955, 'The Logical Structure of Linguistic Theory', published in 1975 by University of Chicago Press.

Chomsky, N.: 1986a, *Barriers*, MIT Press, Cambridge.

Chomsky, N.: 1986b, *Knowledge of Language*, Praeger, New York.

Chomsky, N.: 1995, *The Minimalist Program*, MIT Press, Cambridge.

Clark, H.: 1993, *Arenas of Language Use*, CSLI Publications, Stanford.

Clark, H.: 1996, *Using Language*, Cambridge University Press, Cambridge.

Cooper, R.: 1998, 'Mixing Situation Theory and Type Theory to Formalize Information States in Dialogue Exchanges', in J. Hulstijn and A. Nijholt (eds.), *Proceedings of TwenDial 98, 13th Twente workshop on Language Technology*, Twente University, Twente.

Cooper, R. and J. Ginzburg: 2002, 'Clarification Ellipsis in Dependent Type Theory', in J. Bos and C. Matheson (eds.), *Proceedings of Edilog, the 6th Workshop on the Semantics and Pragmatics of Dialogue*, Edinburgh.

Cooper, R. and M. Poesio: 1994, 'Situation Theory', in *Fracas Deliverable D8*, Centre for Cognitive Science, The Fracas Consortium, Edinburgh.

Copestake, A., D. Flickinger, C. Pollard, and I. A. Sag: n.d., 'Minimal Recursion Semantics: An Introduction', Stanford University and Ohio State University.

Crimmins, M.: 1993, *Talk about Beliefs*, Bradford Books, MIT Press, Cambridge.

Crimmins, M. and J. Perry: 1989, 'The Prince and the Phone Booth: Reporting Puzzling Beliefs', *Journal of Philosophy* pp. 685–711.

Dalrymple, M., F. Pereira, and S. Shieber: 1991, 'Ellipsis and Higher Order Unification', *Linguistics and Philosophy* **14**, 399–452.

Dekker, P.: 1997, 'First Order Information Exchange', in G. Jaeger and A. Benz (eds.), *Proceedings of MunDial 97 (Technical Report 97-106)*. Universitaet Muenchen Centrum fuer Informations- und Sprachverarbeitung, Muenchen.

Fletcher, C.: 1994, 'Levels of Representation in Memory for Discourse', in M. A. Gernsbacher (ed.), *Handbook of Psycholinguistics*, Academic Press.

Garvey, C.: 1979, 'Contingent Queries and their Relations in Discourse', in E. Ochs and B. Schieffelin (eds.), *Developmental Pragmatics*, pp. 68–97, Academic Press, New York.

Gawron, M. and S. Peters: 1990, *Anaphora and Quantification in Situation Semantics*, CSLI Lecture Notes, CSLI, Stanford CA.

Gazdar, G. and C. Mellish: 1988, *Natural Language Processing in Prolog*, Addison Wesley.

Ginzburg, J.: 1996, 'Interrogatives: Questions, Facts, and Dialogue', in S. Lappin (ed.), *Handbook of Contemporary Semantic Theory*, Blackwell, Oxford.

Ginzburg, J.: 1997a, 'On Some Semantic Consequences of Turn Taking', in P. Dekker, M. Stokhof, and Y. Venema (eds.), *Proceedings of the 11th Amsterdam Colloquium on Formal Semantics and Logic*, pp. 145–150, ILLC, Amsterdam.

Ginzburg, J.: 1997b, 'Structural Mismatch in Dialogue', in G. Jaeger and A. Benz (eds.), *Proceedings of MunDial 97 (Technical Report 97-106)*, pp. 59–80, Universitaet Muenchen Centrum fuer Informations- und Sprachverarbeitung, Muenchen.

Ginzburg, J.: 1998, 'Clarifying Utterances', in J. Hulstijn and A. Nijholt (eds.), *Proceedings of TwenDial 98, 13th Twente Workshop on Language Technology*, pp. 11–30, Twente University, Twente.

Ginzburg, J.: 2001a, 'Clarification Ellipsis and Nominal Anaphora', in H. Bunt, R. Muskens, and E. Thijsse (eds.), *Computing Meaning: Volume 2*, No. 77 in Studies in Linguistics and Philosophy, Kluwer, Dordrecht.

Ginzburg, J.: 2001b, 'Turn Taking Puzzles and the Semantics of Adjuncts', King's College, London. Unpublished MS.

Ginzburg, J.: (forthcoming), *Semantics and Interaction in Dialogue*, CSLI Publications and University of Chicago Press, Stanford CA. Draft chapters available from http://www.dcs.kcl.ac.uk/staff/ginzburg.

Ginzburg, J., H. Gregory, and S. Lappin: 2001, 'SHARDS: Fragment Resolution in Dialogue', in H. Bunt (ed.), *Proceedings of the 4th International Workshop on Computational Semantics*, ITK, Tilburg University, Tilburg.

Ginzburg, J. and I. A. Sag: 2000, *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*, No. 123 in CSLI Lecture Notes, CSLI Publications, Stanford CA.

Gregory, H. and S. Lappin: 1999, 'Antecedent Contained Ellipsis in HPSG', in G. Webelhuth, J. P. Koenig, and A. Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, pp. 331–356, CSLI Publications, Stanford.

Groenendijk, J. and M. Stokhof: 1991a, 'Dynamic Montague Grammar', in M. Stokhof, J. Groenendijk, and D. Beaver (eds.), *Quantification and Anaphora I, DYANA Report R2.2.A*, Centre for Cognitive Science, University of Edinburgh.

Groenendijk, J. and M. Stokhof: 1991b, 'Dynamic Predicate Logic', *Linguistics and Philosophy* **14**(1), 39–100.

Hamblin, C. L.: 1970, *Fallacies*, Methuen, London.

Hardt, D.: 1993, 'Verb Phrase Ellipsis: Form, Meaning, and Processing', Ph.D. thesis, University of Pennsylvania.

Heim, I.: 1982, 'The Semantics of Definite and Indefinite Noun Phrases', Ph.D. thesis, University of Massachusetts, Amherst.

Israel, D. and J. Perry: 1991, 'What is Information', *CSLI Report* **91-145**.

Johnson, D. and S. Lappin: 1999, *Local Constraints versus Economy*, CSLI Publications, Stanford CA.

Kamp, H.: 1981, 'A Theory of Truth and Semantic Representation', in J. Groenendijk (ed.), *Formal Methods in Semantics*, Amsterdam Centre for Mathematics.

Kamp, H. and U. Reyle: 1993, *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, No. 42 in Studies in Linguistics and Philosophy, Kluwer, Dordrecht.

Kaplan, D.: 1989, 'Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals', in J. A. et al. (ed.), *Themes*

from Kaplan, pp. 481–614, Oxford University Press, New York. An earlier unpublished version exists as a UCLA Ms from ca. 1977.

Kehler, A.: 1993, 'Common Topics and Coherent Situations: Interpreting Ellipsis in the Context of Discourse Inference', *Proceedings of the 32nd Meeting of the Association for Computational Linguistics*.

Kripke, S.: 1979, 'A Puzzle about Belief', in A. Margalit (ed.), *Meaning and Use*, Synthese Language Library, Reidel, Dordrecht.

Lappin, S. and H. Gregory: 1997, 'A Computational Model of Ellipsis Resolution', in R. Oehrle (ed.), *Proceedings of the Conference on Formal Grammar*, ESSLLI, Aix en Provence.

Larsson, S.: 2002, *Issue based Dialogue Management*, Ph.D. thesis, Gothenburg University.

McTear, M.: 1987, *Children's Conversation*, Blackwell, Oxford.

Milward, D.: 1995, 'Integrating Situations into a Theory of Discourse Anaphora', in P. Dekker and M. Stokhof (eds.), *Proceedings of the 10th Amsterdam Colloquium*, ILLC, Amsterdam.

Milward, D.: 2000, 'Distributing Representation for Robust Interpretation of Dialogue Utterances', *Proceedings of the 38th Meeting of the Association for Computational Linguistics*.

Montague, R.: 1974a, 'Pragmatics', in R. Thomason (ed.), *Formal Philosophy*, Yale University Press, New Haven.

Montague, R.: 1974b, 'The Proper Treatment of Quantification in Ordinary English', in R. Thomason (ed.), *Formal Philosophy*, Yale University Press, New Haven.

Moortgat, M.: 1997, 'Categorial Grammar', in J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Linguistics*, North Holland, Amsterdam.

Ninio, A. and C. Snow: 1996, *Pragmatic Development*, Westview.

Poesio, M.: 1998, 'Underspecified Interpretations and a Theory of Language Processing', *Centre for Cognitive Science*, University of Edinburgh Ms.

Poesio, M. and R. Muskens: 1997, 'The Dynamics of Discourse Situations', in P. Dekker, M. Stokhof, and Y. Venema (eds.), *Proceedings of the 11th Amsterdam Colloquium*, ILLC, Amsterdam.

Poesio, M. and D. Traum: 1997, 'Conversational Actions and Discourse Situations', *Computational Intelligence* **13**, 309–347.

Pollard, C. and I. A. Sag: 1994, *Head Driven Phrase Structure Grammar*, University of Chicago Press and CSLI, Chicago.

Pulman, S.: 1997, 'Focus and Higher Order Unification', *Linguistics and Philosophy* **20**.

Purver, M.: 2001, 'SCoRE: A Tool for Searching the BNC', Technical report, King's College, London.

Purver, M.: 2002, 'Processing of Unknown Words in a Dialogue System', in K. Jokinnen (ed.), *Proceedings of the 3rd SigDial Workshop on Discourse and Dialogue*.

Purver, M. and J. Ginzburg: 2003, 'Clarifying Noun Phrase Semantics', unpublished ms., King's College, London.

Purver, M., J. Ginzburg, and P. Healey: 2002, 'On the Means for Clarification in Dialogue', in J. van Kuppevelt and R. Smith (eds.), *Advances in Discourse and Dialogue*, Kluwer, Dordrecht.

Ranta, A.: 1994, *Type Theoretical Grammar*, Oxford University Press, Oxford.

Reinhart, T.: 1991, 'Bare NP Ellipsis', in A. Kasher (ed.), *The Chomskyan Turn*, Blackwell, Oxford.

Sag, I. A. and T. Wasow (eds.), 1999, *Syntactic Theory: A Formal Introduction*, CSLI, Stanford.

Seligman, J. and L. Moss: 1997, 'Situation Theory', in J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Linguistics*, North Holland, Amsterdam.

Stalnaker, R. C.: 1978, 'Assertion', in P. Cole (ed.), *Syntax and Semantics, Volume 9*, pp. 315–332, Academic Press, New York.

Traum, D.: 1994, 'A Computational Theory of Grounding in Natural Language Conversations', Ph.D. thesis, University of Rochester.

Traum, D., J. Bos, R. Cooper, S. Larsson, I. Lewin, C. Matheson, and M. Poesio: 1999, 'A Model of Dialogue Moves and Information State Revision', in *TRINDI Deliverable 2.1*, University of Gothenburg, Gothenburg. Available from http://www.ling.gu.se/research/projects/trindi.

van der Sandt, R.: 1992, 'Presupposition Projection as Anaphora Resolution', *Journal of Semantics* **9**(4).

van Eijck, J. and H. Kamp: 1997, 'Representing Discourse in Context', in J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Linguistics*, North Holland, Amsterdam.

Vermeulen, C. F. M.: 1993, 'Sequence Semantics for Dynamic Predicate Logic', *Journal of Logic, Language, and Information* **2**(3), 217–254.

Jonathan Ginzburg
Department of Computer Science
King's College, London
The Strand, London WC2R 2LS
UK
ginzburg@dcs.kcl.ac.uk

Robin Cooper
Department of Linguistics
Göteborg University
Box 200, 405 30 Göteborg
Sweden
cooper@ling.gu.se