This is a contribution from *Alignment in Communication. Towards a new theory of communication*.
Edited by Ipke Wachsmuth, Jan de Ruiter, Petra Jaecks and Stefan Kopp.

On making syntax dynamic

The challenge of compound utterances and the architecture of the grammar

Eleni Gregoromichelaki, Ruth Kempson, Christine Howes and Arash Eshghi

The Pickering and Garrod model (Pickering & Garrod, 2013) represents a significant advance within the language-as-action paradigm in providing a mechanistic non-inferential account of dialogue. However, we suggest that, in maintaining several aspects of the language-as-product tradition, it does not go far enough in addressing the dynamic nature of the mechanisms involved. We argue for a radical extension of the language-as-action account, showing how compound-utterance phenomena necessitate a grammar-internal characterization which can only be met with a shift of perspective into one in which linguistic knowledge is seen as procedural. This shift provides a more psychologically plausible model of language-in-use, a basis for allowing intentions and speech-acts to be co-constructed, as well as a computationally tractable basis for dialogue models.

1. Introduction

Pickering and Garrod (2013; this volume, *P&G* henceforth) propose that production and comprehension in dialogue are as tightly interwoven as argued in current computational neuroscience models linking action, action perception and joint action. A key mechanism is *prediction*. The tight integration of perception and action is achieved via "forward models" which predict the outcomes of action commands before an action is executed and support covert imitation of an agent's actions during their perception. Transferring such ideas from the domain of action to the domain of linguistic processing, P&G argue that people predict their own utterances (now conceived as actions) at different levels of representation (semantics, syntax, and phonology) via forward modelling. They also covertly imitate and predict their interlocutors' utterances via the same mechanism.

In this paper, we argue for a radical extension of this language-as-action perspective adopted by P&G based on the phenomenon of *compound utterances*, the type of utterances illustrated in (1), a phenomenon whose modelling, in our view, crucially depends on predictive mechanisms.

(1) Context: Friends of the Earth club meeting
A: So what is that? Is that er... booklet or something?
B: It's a book
C: Book
B: Just ... talking about al you know alternative
D: On erm... renewable yeah
B: energy really I think.....
A: Yeah [BNC:D97][1]

We show that the full array of compound-utterance data demands a grammar-internal characterization, in that the licensing of the complete structure, and ultimately the discourse effects of such moves, depends on syntactic/semantic constraints. As a result, in our view, a uniform account of such data within the grammar itself (rather than a separate processing component, see e.g. Peldszus & Schlangen, 2012; Poesio & Rieser, 2010) can only be given with a shift of perspective into one in which linguistic knowledge is seen as action-based (procedural), i.e., a set of unencapsulated processing mechanisms. For this reason, we suggest that the predictive mechanisms supporting compound utterances in dialogue are not necessarily at the level of the forward model as P&G suggest. In our view, given the tight linking with syntactic constraints, this type of predictivity is at a more basic level: contrary to the view of the grammar P&G and others assume, under the view we present, syntactic constraints are not constructs reflecting abstract modular knowledge of licensed representations; rather, "syntactic" licensing, the combinatorial mechanisms, can be reduced to the unencapsulated mapping mechanism from message to utterance, eliminating a separate level of representation for syntax; this mechanism crucially incorporates an incremental predictive element, as an integral component of the grammar itself, which can be utilized both for the imposition of classic "syntactic" combinatorial constraints and explain the seamless generation and processing of compound utterances. In consequence, we indicate that the dubious "impoverished" nature of the linguistic efferent copies postulated by P&G (as noted also in Jaeger & Ferreira, 2013) is an artefact of the representational architecture they assume. Instead, we suggest that the architecture of the grammar itself provides a more plausible alternative

1. BNC refers to data found in the British National Corpus, see Burnard (2000); for further data, see Purver et al. (2009).

explanation for some of the coordination phenomena P&G discuss so that forward modelling is reserved for higher-level predictions, for example, cases amenable to explicit conceptualization of the communication process by the participants, cases of break-down in the communication, monologue planning, deception, irony etc. For all other cases, a domain-general action-oriented model that accounts for both the subsentential, the supra-sentential and cross-modal structure of an interaction (a *grammar*) provides an adequate base of explanation. In our view, this shift provides a more psychologically plausible model of language-in-use, a basis for allowing intentions and speech acts to be seen as co-constructed during interaction instead of having to be taken as predetermined causal factors, as well as a computationally tractable basis for dialogue models.

1.1 Language-as-action and the nature of linguistic knowledge

According to H. Clark (1992, 1996), psycholinguistic models fall into one of two traditions. The *language-as-product* paradigm involves standard information-processing analyses springing from early theories of transformational grammar emphasizing linguistic representations, the "product" of language processing. This is accompanied by a parallel view in theoretical linguistics where core language processes have been conceptualized as idiosyncratic and encapsulated (Hauser, Chomsky, & Fitch, 2002) and linguistic capacity is conceived as relying on a body of declarative knowledge of rules and representations ("competence"). The processing theories of this tradition focus on the individual cognitive processes during which language users employ context-independent propositional representations, as outputs of an encapsulated system that can be computed quickly and efficiently. Such representations are taken as general enough in order to serve as the input for more computationally complex context-specific representations. However, a substantial amount of evidence indicates that language users combine linguistic information and context-dependent content very early during processing (see e.g. Altmann & Steedman, 1988; Marslen-Wilson & Tyler, 1980), which led to a methodological separation between competence/performance theories to justify the postulates of the conflicting models.

In contrast, the *language-as-action* tradition, stemming from work in the Ordinary Language philosophy (e.g. Austin, 1962; Grice, 1975; Searle, 1969), and work on conversational analysis (e.g. Schegloff, 2007), emphasizes how people use language to perform (speech) acts and involves investigations of interactive dialogue considered as the basic form of language use. With language conceived as an idiosyncratic cognitive module, theories of action would seem to have little to say about it. In accordance with this, work in theoretical linguistics, despite evidence that even basic aspects of interpretation involve pragmatic processing

(e.g. Levinson 2000; Sperber & Wilson 1995) linked to interactive participant coordination (Clark 1996), has sought to reduce such mechanisms to internal reasoning processes compatible with individualistic processing. Computationally intractable inferential mechanisms, propositional attitude mindreading, strategic planning or game-theoretic deliberation are then postulated to account for joint activity mediated through language generating puzzles like the *mutual knowledge paradox* (Clark & Marshall, 1981), according to which, interlocutors have to compute an infinite series of beliefs in finite time. In this respect, even dialogue-oriented psycholinguistic models make heavy use of concepts like Gricean intention-recognition and mind-reading which contrast[2] with the automaticity, fastness and efficiency that characterizes online linguistic interaction. In addition, the separation between competence and performance has led even dialogue-oriented psycholinguists, e.g. Clark (1996), to distinguish language$_S$ (language structure), which involves systems of rules and representations, from language$_U$ (language-in-use), which preserves the roots of the language-as-action hypothesis. The latter has been recently enhanced by the neurophysiological discoveries of strong parallels between language and action (see, e.g., Rizzolatti & Craighero, 2004) and direct mappings and common coding for production/perception processes. However, even under this view, the dilemma remains that there is plenty of evidence for apparent means-ends understanding and *audience design* in conversation, both types of coordinative behaviour, posing the problem of how to model the interlocutors' abilities and the context that allow them to achieve this rapidly and efficiently during online processing. There have been attempts to reconceptualize the classical (neo-) Gricean accounts of communication in terms of implicit subpersonal and interpersonal processes, sometimes even rejecting the BDI model of explanation while attempting to maintain that inferential mental state ascription is the primary basis for communication (see e.g., Davies & Stone, 1995; de Ruiter et al., 2007, 2010; Sperber & Wilson, 1995, a.o.). However, in our view, such attempts risk to introduce unnecessary conceptual confusion in two respects. Firstly, the view that attribution of mental states is the sine-qua-non for communication is taken as axiomatic, rather than a position to be defended (see also de Bruin, Strijbos & Slors, 2011) thus ignoring a range of alternatives to be explored (see e.g. Ginzburg, 2012; ch. 7; Mills, 2011; Mills & Gregoromichelaki, 2010; Piwek, 2011). Secondly, as a consequence of this stance, even when behaviours or situations are encountered that cannot be properly explained through the necessary attribution of folk-psychological abilities (e.g. "theory of mind" evidence in

2. For indicative literature on the complexity involved in propositional attitude mindreading see Bermúdez, 2003; Apperly, Back, Samson & France, 2008; McKinnon & Moscovitch, 2007; Apperly, 2011 and references there, esp. Ch. 5.

animals/infants/autistic patients, vagueness of speech act content, collaborative emergence of structures and intentions in dialogue), researchers still seek to postulate something weaker as a substitute elevating such constructs as the mechanisms enabling "intention recognition". What is missed here is that attribution of propositional attitude mindreading is only justified under the assumption that the agents understand, employ and engage with the complex causal structures that the logic of such states requires (see e.g., Davidson, 1980; for further explication see Apperly, 2011, Ch. 5; Bermúdez, 2003). Especially for Gricean intentions, this should involve multiple levels of metarepresentation. More pertinently for our purposes here, from the point of view of standard psychological and computational models where communication is conceptualized as crucially involving Gricean propositional attitude mindreading, interspersed within low-level processing steps, conversation appears to be very complex (see e.g. Poesio & Rieser, 2010) for an admirably thorough illustration of this complexity in accounting for a single type of compound utterances). This is because, in conversation, interlocutors must be modelled as having to switch between production and comprehension, perform both acts at once, and develop their plans/intentions on the fly (Pickering & Garrod, 2004). In contrast, the work of Pickering and Garrod presents alternatives that can support more realistic models of real-time language processing (see also Gann & Barr, 2012).

1.2 Dialogue within an action-based framework: Pickering and Garrod (2013)

The model presented by P&G develops the basis of a psychological account of human communication that promises to provide a compromise between the language-as-product and language-as-action paradigms in a way that reconciles realistic fast processing in dialogue with the interpersonal and subpersonal mechanisms that support fluent intersubjectivity. Standard modular accounts of language separate production and comprehension by postulating an intermediate cognitive level of integration, a perspective that is incompatible both with the demands of communication and with extensive data P&G present indicating that production and comprehension are tightly interwoven at a very fine-grained level. As regards our concerns here, for example, as shown in (1) earlier, interlocutors clarify, repair and extend each other's utterances, even in the middle of an emergent clause (*compound utterances*) switching fluently among planning, comprehension, production and integration of contextual cross-modal inputs.

In order to solve the puzzle of rapid and fluent language-based interaction, P&G propose to conceptualize language processing in terms analogous to recent accounts of instrumental action perception and motor action. In the light of

current evidence regarding interactions between perception and action (e.g., Bargh & Chartrand, 1999; Sebanz, Bekkering, & Knoblich, 2006), mechanistic frameworks have been developed in computational neuroscience that make use of the notion of *internal models* (e.g., Grush, 2004; Wolpert, Doya & Kawato, 2003; see also Hurley 2008). On these views, during execution of goal-directed actions, it is more efficient to derive and use a predictive (forward) model of the expected sensory consequences rather than simply waiting to react on the basis of actual reafferent feedback. Accordingly, during execution, an *efference copy* of the motor command is created causing the forward action model to generate the predicted act and its consequences, which are then compared with the actual feedback for adjustment and learning purposes. Similarly, during perception, an inverse model (plus the context) can be used to covertly imitate the actor and predict their subsequent movements thus either leading to overt imitation or achieving goal-understanding as well as coordination in joint action cases. In these accounts of goal-directed action, a central role is assigned to prediction in both action execution and action understanding, with subpersonal low-level online perception-action links being utilized to achieve the intersubjective understanding/coordination for which offline inferential models had previously been presumed to be needed. P&G apply these mechanisms to language production and comprehension for which there is a lot of evidence that they crucially involve predictive processes (e.g., comprehension: Levy, 2008; production: Pickering & Garrod, 2007; Jaeger, 2010). According to P&G, speakers use forward models to predict their upcoming utterances thus adjusting their output accordingly (audience design could be argued to be based on such a mechanism, but see also Gann & Barr, 2012; Horton & Gerrig, 2005). Listeners covertly imitate speakers through use of inverse models which through learned associations and the shared current context provide the background for understanding the speaker's "intention" in uttering the current input. They then use forward models based on their own potential next motor command to predict what speakers are likely to say next (this constitutes the "simulation route" to comprehension).

1.2.1 *Incrementality and predictivity in dialogue and the role of grammar*

Despite the radical nature of their model, in our view, P&G maintain a conservative stance as regards the online progress of interaction, rehearsing standard assumptions about how linguistic processing is executed. They assume that linguistic information is organized hierarchically and represented at different levels between message and articulation: (at least) semantics, syntax, and phonology. These levels are ordered "higher" to "lower," so that a message causes a semantic representation, semantics evokes a syntactic representation, this in turn maps to phonology, and from phonology to speech sounds. Thus, a *production process* goes from message to sound via each of these levels (message → semantics → syntax →

phonology → sound) whereas a *comprehension process* goes from sound to message in the opposite direction. Given the forward model that speakers and hearers use to predict what is likely to come next, this means that producing utterances involves not only production processes but also comprehension processes; similarly, comprehending utterances involves comprehension processes but also incorporates production processes. Furthermore, reflecting the relationship between the linguistic levels, the production command is taken to constitute the message that the speaker wishes to convey, including information about communicative force, pragmatic context, and a nonlinguistic situation model, which is then mapped to the representational levels assumed at the action execution phase.

This model might seem to be confirmed by the type of compound-utterances termed *collaborative completions* as in (2) and (3):

(2) Conversation from A and B, to C:
A: We're going to ...
B: Bristol, where Jo lives.

(3) A: Are you left or
B: Right-handed.

However, it is very much less compatible with the many other types of continuations in conversation. As (4)–(5) show, such completions by no means need to be what the original speaker actually had in mind, so they don't need to involve prediction at the message or semantic levels:

(4) Morse: in any case the question was
Suspect: a VERY good question inspector [Morse, BBC radio 7]

(5) Daughter: Oh here dad, a good way to get those corners out
Dad: is to stick yer finger inside.
Daughter: well, that's one way. [from Lerner, 2004]

In fact, such continuations can be completely the opposite of what the original speaker might have intended, as in what we will call *hostile continuations* or *devious suggestions* which are nevertheless collaboratively constructed from a structural point of view:

(6) (A and B arguing:)
A: In fact what this shows is
B: that you are an idiot

(7) (A mother, B son)
A: This afternoon first you'll do your homework, then wash the dishes and then
B: you'll give me £10?

In (4)–(7), the string of words ("sentence") that the completion yields is not at all what either participant takes themselves to have had in mind as message (or semantic representation) and there is no reason to suggest that the hearer first predicted the original speaker's expected continuation, then rejected it, and then constructed a novel one fitting more appropriately their own purposes. This would predict substantial complexity in the use of such utterances, unlike the aims of a realistic dialogue model.

Such data also cast doubt on the long-held assumption that in all successful acts of communication, the speaker must have in mind some definitive propositional content which they intend to convey to their hearer, whose task, conversely, is to succeed in grasping that particular content. Some variant of this assumption underpins many current pragmatic theories (see e.g. Bach & Harnish, 1982; Sperber & Wilson, 1995). Taking this view, one possible analysis of the data in (4)–(7) can be given through imposing distinctions between "private" and "public" intentions and a contrast between the apparent and missing full propositional contents (this was suggested to us by a reviewer). However, in our view, this approach places such data under the same category as failures and non-standard uses (see Airenti, Bara, & Colombetti, 1993 for an analysis of such phenomena) in that there is a clash between which proposition was privately "intended" by the original speaker and what instead was imposed on him/her. Besides the fact that we don't think that speakers have to have fully-formed propositional intentions in order to start to speak, it seems to us that such a categorization is misleading in that such data are not qualitatively different as regards what happens in "ordinary" cases of communication (see e.g. (1)). When people engage in conversational dialogue they do not do so in order to divert or support some other speaker's plan, they genuinely expertly engage in a joint task in a way that does not necessitate to consider the other person's mental states (i.e. meanings and goals are transparent to the participants and not mediated through propositional attitude ascriptions see e.g. Millikan, 1984, Ch. 3). The sequential nature of the conversational structure (see e.g. Schegloff, 2007) and, in general, the very nature of direct perception of "meaning" in language (McDowell, 1998) provide an adequate background for the execution of a joint project. Accordingly, a public/private contrast obscures the significance of the fact that these utterances are constructed as a *joint* speech-act with a single propositional content rather than as two clashing independent ones. What is then missed is that the employment of such joint structures has interactional effects: for example, in some contexts, invited completions of another's utterance have been argued to exploit the vagueness/covertness of the speech act involved to avoid overt/intrusive elicitation of information (*grammar-induced speech acts*, see below in (8) as well as (12)–(15), and Gregoromichelaki, Cann, & Kempson, 2013):

(8) (Lana = client; Ralph = therapist)
Ralph: Your sponsor before ...
Lana: was a woman [from Ferrara, 1992]

Here the therapist uses an invited completion in a way that gives the patient the opportunity to reveal or not as much information as she is willing to reveal. Should we attribute this technique to a fully-propositional private intention or premeditated plan that occurs inside the mind of the therapist or rather to the training and practice of an expert professional employing (sub-consciously) well-rehearsed patterns of appropriate interaction? Along with other researchers, we suggest that intentions are always "public" in that they should not be seen as causal factors driving communication but, instead, as discursive constructs that are employed by participants, as part of a (meta-) language regarding the coordination process itself, when they need to conceptualize their own and others' performance for purposes of explicit deliberation or accountability when trouble arises. Empirical evidence for this stance come from studies showing that in task-oriented dialogue experiments explicit negotiation is neither a preferential nor an effective means of coordination (Garrod & Anderson, 1987). If it occurs at all, it usually happens after participants have already developed some familiarity with the task. Further more specific evidence have been provided by experiments probing participants' awareness of even their own intentions in early and late stages of task-oriented dialogue leading to expert performance (see e.g. Mills, in press; Mills & Gregoromichelaki, 2010). It has been shown that as participants become more and more expert in the task, awareness of plans/intentions emerges and can then be utilized as a means of coordination when trouble ensues (see also Suchman, 2007).

Hence, in our view, the production/comprehension of compound contributions cannot be taken to causally rely on the determination of a pre-planned speaker-intended speech-act. Indeed, in our view, fixed joint intentionality is decidedly non-normal in dialogue: to the contrary, joint intentionality has to develop through engagement in the task, hence it is emergent rather than constitutive of dialogue acts. On the other hand, there is something that the participants share *ab initio*, i.e., a set of processing mechanisms and practices, in our view, the "grammar", that can ground further coordination. From this point of view, the important observation that comes from compound utterance data is that their licensing crucially employs this grammar. For example, the dependency that licenses the reflexive anaphor *myself* in the third turn of (9) relies on its antecedent *you* in B's previous turn. And in (10), the dependency holds across turns between a Negative Polarity Item and its triggering environment, the question:

(9) with smoke coming from the kitchen:
A: I'm afraid I burnt the kitchen ceiling
B: But have *you*
A: burned *myself*? Fortunately not.

(10) A: *Have* you mended
B: *any* of your chairs? Not yet.

Such split-participant realizations range over the entire range of syntactic and semantic dependencies, and are observable in all languages (Howes et al., 2011; Kempson, Gregoromichelaki, & Chatzikyriakidis, 2012; Purver et al., 2009). Given that such dependencies are defined grammar-internally, a grammar formalism has to be able to license such compound utterances if it is to meet minimal conditions of adequacy. However, these data are highly problematic for all standard frameworks, given the commitment to models of linguistic knowledge (competence grammars) licensing such dependencies over sentence-strings independent of any performance realization. Even for frameworks that employ incrementality at the level of the processing components (Peldszus & Schlangen, 2012; Poesio & Rieser, 2010) data like (9)–(10), as well as (11) below, counter-intuitively, will have to be treated as syntactically/semantically deviant. This is because such models maintain an independent level of syntactic representation over strings of words with the result that, at this level, such data have to either be classed as "ungrammatical" (**But have you burned myself*) or assigned inappropriate string/semantic pairings (e.g. in (11), the string *Is this yours or yours* should be assigned the interpretation "Is this Yo's or Eleni's"). The only way out for such models would then be to consider such utterances as "performance data", not relevant for the definition of the grammar, in that they become acceptable only through the operation of the performance modules which tolerate "ungrammatical" input.

There is, however, an alternative where the intuitive acceptability of such collaboratively constructed utterances is not delegated to performance but rather naturally emerges from the architecture of the grammar formalism itself. *Dynamic Syntax* (*DS*, Cann, Kempson, & Marten, 2005; Kempson, Meyer-Viol, & Gabbay, 2001; Purver et al., 2011), is an action-based grammatical framework eschewing representations of sentence strings and modelling directly both language comprehension and production as the incremental steps leading bidirectionally from message to articulation.[3] On this view, compound contributions are not merely

3. At the content end of the spectrum, Guhe (2007), Guhe et al. (2000) have argued for the incremental conceptualization of observed events resulting in the generation of preverbal messages in an incremental manner guiding semantic/syntactic formulation. At the other extreme, phonology can be conceived in processing terms (see, e.g., Kaye, 1989; Lahiri & Plank, 2010).

characterizable, but rather predictable consequences of the grammar architecture itself. By assuming an action-based formalism for the characterization of the combinatorial properties of language, in effect, on this view, the grammar involves the crystallization of motor mechanisms originally evolved to control/represent the hierarchical structure of instrumental action (for a similar view of how "syntax" emerged, see also Gallese, 2007, Section 8; Hurley, 2008; Pulvermüller & Fadiga, 2010). Thus, in parallel to assumptions in the P&G model, but more radically transferred within the grammar itself, the DS combinatorial mechanisms employ an architecture similar to those assumed in the control of the hierarchies that emerge in the analysis of goal-directed actions. But since these mechanisms constitute a relatively fixed and stable architecture that can be employed rapidly, reliably and automatically there is no need to assume the necessary employment of forward/inverse models whose usual function is in the service of learning and adjustment. Instead, predictivity/goal-directedness is built right inside the operation of the grammar for efficiency and control purposes. That is, the grammar design includes a top-down element that provides the source for the generation of predictions (which can further be simulated in a forward model but need not necessarily be so); and the coupling of parser/generator is intrinsically modelled as a form of covert imitation and prediction through the employment of identical mechanisms in a shared context. As we will demonstrate in more detail below, crucially, such predictions guide lexical access at a subpropositional level, for both speaker and listener in parallel, irrespective of what role they realize currently. It is this more basic mechanism (at a similarly low-level as the "association route" in the P&G model) that participants exploit in the generation of compound utterances in order to steer the conversation towards their own goals without necessarily having to consider the current speakers' intended messages. Under this view, participants can progress via an associative route, guided by the goals generated by the grammar and, on this basis, negotiate derivative constructs like intentions and strategies overtly at the social level ("externalized inference", see also Pickering & Garrod, 2004).

This view also allows the possibility that the components of compound utterances can play multiple roles at the same time (e.g. the second-turn fragments in (3) earlier and (11) below can be simultaneously taken as question/clarification/completion/acknowledgment/answer), a phenomenon not commensurate with P&G's assumptions of a predefined communicative force included in the motor command:

(11) Eleni: Is this *yours* or
Yo: *Yours*. [natural data]

As expected, notice also that co-construction at the sub-propositional level can be employed for the performance of speech acts without fully expressed propositional contents. We have argued that this is accomplished by establishing "(syntactic) conditional relevances",[4] i.e., exploiting the grammatical dependencies themselves to induce a response by the listener (*grammar-induced speech acts*, see Gregoromichelaki et al., 2013). For example, completions might be explicitly invited by the speaker thus forming a question-answer pair without the speech act content involving a full proposition (see also (8) earlier):

(12) A: And you're leaving at ...
B: 3.00 o'clock

(13) A: And they ignored the conspirators who were ...
B: Geoff Hoon and Patricia Hewitt
[radio 4, Today programme, 06/01/10]

(14) Jim: The Holy Spirit is one who <pause>gives us?
Unknown: Strength.
Jim: Strength. Yes, indeed. <pause>The Holy Spirit is one who gives us? <pause>
Unknown: Comfort. [BNC HDD: 277–282]

(15) George: Cos they <unclear>they used to come in here for water and bunkers you see.
Anon 1: Water and?
George: Bunkers, coal, they all coal furnace you see, ...
[BNC, H5H: 59–61]

Seen from this perspective, the P&G model represents a significant advance within the language-as-action paradigm in providing a mechanistic non-inferential account for action understanding and production in dialogue. However, we suggest that in maintaining several aspects of the language-as-product tradition, it does not go far enough in extending the action-based architecture. The multiple representational levels assumed between the motor command and articulation is one such element causing trouble. As a result, we suspect that the noted dubious "impoverished" nature of the postulated efferent copies is an artefact of the abstract representational architecture P&G assume. Instead, under the DS proposal, along with mechanisms for incremental construction of messages (e.g., Guhe, 2007) and a view of phonology as a guide to parsing, it would be possible to implement a view where the efference copy is directly mapped to the predicted meaning

4. For the concept of *conditional relevance* in conversation see, e.g., Schegloff (2007).

(rather than multiple intermediate linguistic units) as has been suggested by Jaeger and Ferreira (2013).

Reconceptualizing the grammar along the lines suggested by DS promises to solve another problem having to do with the relevance of neuroscience evidence for models of linguistic competence. Linguists have long disputed the compatibility of current theories of brain function with (competence) theories of syntactic structure (see e.g. Jackendoff, 2002). Because no alternative to standard competence models has been conceived, it has been taken for granted that the alleged abstract nature of syntactic structure conflicts with the requisite direct matching between perceptual linguistic information and corresponding motor plans that recent neuroscience models advocate:

> In spite of great progress in the neurosciences in understanding the mechanisms of language and conceptual thought, there is still one domain that appears to be largely immune to brain evidence. This domain, syntax, is far removed from concrete events and is built on highly sophisticated abstraction. Even its most basic phenomena are best described in terms of abstract formula and it therefore may appear doubtful – if not impossible – that its core principles can be translated into the profane language of nerve cells and circuits." (Pulvermüller, 2010, p. 167).

Especially for the kind of evidence that P&G cite, regarding the close affinity between action and linguistic processing, as well as, current neuroscience results pointing in the same direction (Gallese, 2007, Section 8; Hurley, 2008; Pulvermüller & Fadiga, 2010), the view of syntax as an abstract domain of knowledge, as assumed by standard grammars, constitutes the biggest stumbling block for further progress (as also noted by Patel 2008, Section 5.4.3). This standard view of syntax as an abstract intermediary has led to specific claims that this immunity to brain evidence is due to the very nature of syntactic phenomena that are, it is claimed, not amenable to time-linear sequential explanations (Tettamanti & Moro, 2012; cf. Pulvermüller, 2010). According to this standard view, syntactic explanations rely on complex hierarchical structures that become hidden to the bodily senses due to their linearization into strings of words. Hence, it is claimed, this inaccessibility to perceptual systems implies that syntactic processing must rely on different capacities than those involved in matching perceptual linguistic information onto corresponding motor plans as assumed in the P&G model. However, from the DS perspective, there is an alternative action-based view of "syntax" which makes it directly compatible with architectures like the P&G model as well as with currently proposed neurobiological mechanisms mediating action understanding/execution. We turn to a more detailed presentation of DS next.

2. Dynamic syntax

2.1 Grammar and coordination in joint activities

Along with P&G, the DS account is concerned with "embodiment of form" (the "vehicle level" Gallese, 2007) in that action and procedural knowledge underpin both comprehension and production. However, even more radically, according to the DS perspective, such procedural knowledge is constitutive of the grammar itself. DS proposes a distinct reconciliation between the "language-as-action" and "language-as-product" traditions while at the same time shifting the boundaries between grammar and pragmatics. All traditional syntactic puzzles (including those declared as impervious to such explanations by Tettamanti & Moro, 2012) have been shown to be amenable to time-linear accounts (see e.g. Cann et al., 2005; Eshghi et al. 2010, 2011; Kempson et al., 2001; Kempson et al., 2012; Kempson, Gregoromichelaki, & Howes, 2011)[5] within a system which crucially involves:

- an action-based architecture that employs unitary representations integrating multiple sources of contextual information
- word-by-word incrementality and predictivity within the grammar formalism

As a consequence of this stance, what have been identified as inherent features of the DS grammar architecture employed to solve traditional grammatical puzzles have also been shown to underlie many features of language use in dialogue. Firstly, the function of items like inserts, repairs, hesitation markers etc., interact with the grammar at a sub-sentential level (Clark & Fox Tree, 2002). Hence the grammar must be equipped to deal with those in a timely and integrated manner. In addition, the turn-taking system (see, e.g., Sacks, Schegloff, & Jefferson, 1974) seems to rely on the grammar, based on the predictability of (potential) turn endings; in this respect, recent experimental evidence has shown that this predictability is grounded on syntactic recognition rather than prosodic cues etc. (De Ruiter, Mitterer & Enfield, 2006); and further evidence shows that people seem to exploit such predictions to manage the timing of their contributions (Henetz & Clark, 2011). More importantly for our concerns here, incremental planning in production allows the grammar to account for how the interlocutors interact sub-sententially in dialogue to derive joint meanings, actions and syntactic constructions taking in multi-modal aspects of communication and feedback, a fact claimed to be a basic characteristic of interaction (Goodwin, 1981). Such mechanisms can in fact serve as the means for *discovering* one's own and others' intentions (see Gregoromichelaki, Cann,

5. For an initial neural model of the sequential nature of syntactic constraints see also (Pulvermüller, 2010).

& Kempson, 2013; Mills & Gregoromichelaki, 2010). In addition, such a view extends in the domain of pragmatics in so far as the claim is that automatic sensorimotor couplings provide the basis for semantic and pragmatic parity: "actions done by other individuals become messages that are understood by an observer without any cognitive mediation" (Rizzolatti & Craighero, 2004).

2.2 Dynamic action-based grammars and dialogue coordination

We have argued that the view emerging from dialogue data is that an appropriately defined model should be able to provide the basis for direct modelling of dialogue coordination as an immediate consequence of the grammar architecture. Uncharacteristically for grammars, being an action-based model, the core notion in DS is goal-directed incremental information growth/linearization following the time-linear flow of parsing/production. Utterance contents, represented as binary tree-structures of predicate-argument form,[6] are built up relative to a context which evolves in parallel keeping a record of extra-linguistic information, the sequence of unfolding partial tree-structures and the actions used to build them. The process of building up such representations in context is what is taken to constitute NL "syntax": syntactic constraints, as well as word-entries, are modelled as sets of procedures that define how parts of representations of content can be incrementally introduced and updated.

The general process is taken to involve building as output a tree whose nodes reflect the content of some utterance – in the simplified case of a sentence uttered in isolation, a complete propositional formula. This is expressed from the beginning as an imposed top-down goal (*?Ty(t)* in Figure 1) to be achieved eventually after interaction with the context and steps of processing.

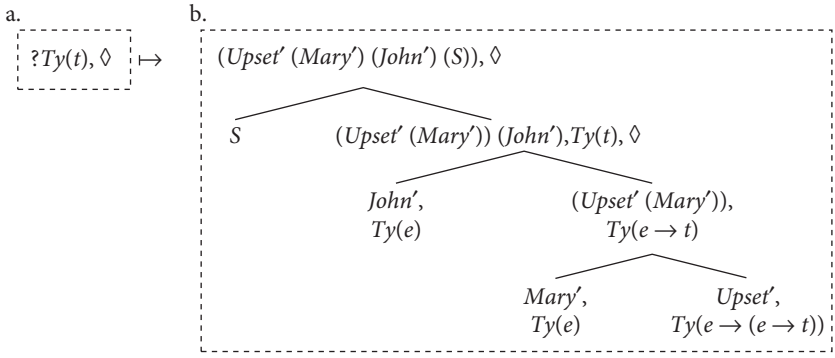

Figure 1. Processing of *John upset Mary*

6. These representations do not have to be conceptualized necessarily as traditional symbolic representations; they can also be seen as *embodied representations* (Pezzulo, 2011), thus removing another layer of the "cognitive sandwich".

In DS terms, in such a simple case, the input to the processing task is represented as a minimal tree that does nothing more than state at the root node the "goal" to be achieved, namely, to establish some propositional formula (?*Ty*(*t*) in Figure 1; goals are represented with *?* in front of annotations). For example, in the parse of the string *John upset Mary*, the output tree in Figure 1 to the right of the → constitutes some final end result: it is a tree in which the propositional formula itself annotates the root node, and its various subterms appear on the dominated nodes rather like a proof tree in which all the nodes are labelled with a formula and a semantic type. These DS trees are invariably binary, and, by convention, the argument always appears on the left branch, and the functor on the right branch (a pointer, ◊, identifies the node under development). Hence they reflect conceptual structure, not structures over strings. Each node in a complete tree is annotated not with words but contents, i.e. terms of a logical language (e.g. *Mary'*, *λx.Upset'x*), these being subterms of the resulting propositional representation ('*Upset'(Mary') (John')* holds at index *S'* in Figure 1 above). The processing task is to use both lexical input, computational actions (e.g. *Introduction and Prediction* in step 1 in Figure 2) and information from context to progressively enrich the input tree satisfying all the sub-goals imposed (in Figure 2, the satisfaction of goals introduced with *?* initially is indicated by removal of *?* and cross-out when satisfied).

These interpretation trees are the only representations constructed during processing, hence no distinct syntactic level of representation is assumed. Production follows exactly the same procedures, but with the added requirement of a subsumption relation to some richer "goal" tree. For example, the tree T_g, the *goal tree*, shown in Figure 2, step 4, will be present from the beginning as the target of processing in case the speaker has planned a full proposition in advance. However, more partial trees can be assumed as targets in production, the only requirement is that the goal-tree is always at least one processing step ahead of the currently processed tree.

As in DRT and related frameworks (see also Jaszczolt, 2005), semantic, truth-conditional evaluation applies solely to these contextually-enriched representations, hence no semantic content is ever assigned to structures inhabited by elements of strings of words (sentences). The distinguishing feature of DS as compared to DRT is that this process of progressive building of semantically transparent structures is taken as core syntax: there is no other level of structure interfacing the phonological sequence and some ascribable content. Furthermore, all syntactic dependencies, including those mentioned by Tettamanti and Moro (2012), are seen in procedural terms, including, in particular, the classical evidence for denying the direct correspondence between NL-structure and semantic content that led to accounts via transformations (Cann et al., 2005; Kempson et al., 2001;

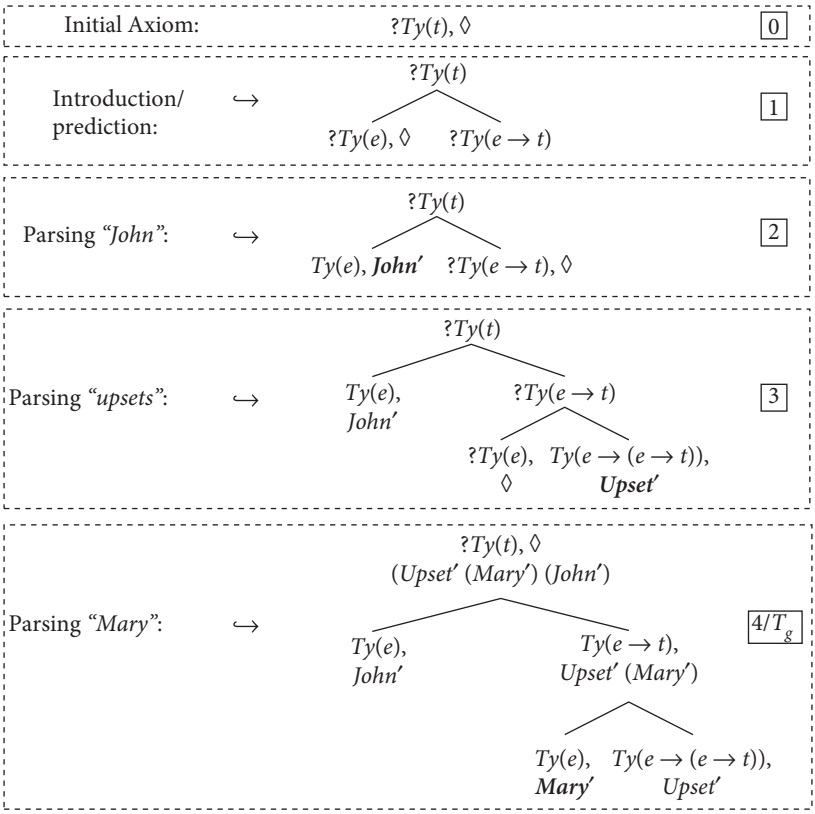

Figure 2. The processing steps for *John upsets Mary*

see Kempson et al. (2011) for various analyses of distinct phenomena crosslinguistically). For example, "movement" cases (*Who did you see; The man who you saw*) are analyzed in terms of the initial projection of an underspecified dominance relation between the input provided by the wh-element and the predictively induced predicate-argument structure. Later update to a fixed dominance relation occurs at the point at which, in movement accounts, an associated "empty category" is posited. Cases of anaphora or ellipsis may occur when the linguistic input includes anaphoric elements that have to be obligatorily enriched from the surrounding context. Anaphoric elements introduce *metavariables*, symbolized as *U, V, W*, along with goals triggering context search for their replacement with contextually-available semantic terms.

The gradual unfolding of the emergent DS semantic trees is crucial for accounts of dialogue phenomena. For example, it has been shown both by corpus research (Fox & Jasperson, 1995) and experimental results (Eshghi et al., 2010)

that repair processes in dialogue target primarily, what in other frameworks are characterized as "constituents". Additionally, as we saw, use of fragments during interaction follows syntactic constraints indicating their appropriate integration in some structured representation (see e.g. (9)–(10)). This is more evident in languages with rich morphology. For example, languages like German and Greek require that the fragment bears appropriate case specifications, otherwise it is perceived as ungrammatical (for similar data in other constructions see also Ginzburg, 2012):

(16) Context: A and B enter a room and see a woman lying on the floor:
A to B: Schnell, den Arzt/*der Arzt
"Quick, the doctor.ACC /*the doctor.NOM" [German]

(17) A is contemplating the space under the mirror while re-arranging the furniture and B brings her a chair:
tin karekla tis mamas?/*I karekla tis mamas? Ise treli?
the chair.ACC of mum's/*the chair.NOM of mum's. Are you crazy?
[Greek] [clarification]

On the DS account, such morpho-syntactic particularities do not justify distinct levels of representation, for the morphological information is defined as introducing constraints for appropriate integration in the unfolding semantic tree: in particular, case information such as "accusative" in Greek is taken to project a constraint that the content of an expression bearing this feature must occupy the first argument position of a predicate; depending on the case system of a specific language a combination of lexical information and general computational rules ensure the appropriate contribution of morphological information to the semantics without a separate level of syntactic representation having to be assumed (for various crosslinguistic analyses see Chatzikyriakidis & Kempson, 2011; Gregoromichelaki 2013; Kempson et al., 2011).

Modelling NLs as encoding constraints on growth of interpretation relative to context is exactly the assumption that allows the handling of dialogue phenomena, such as compound contributions, in a straightforward manner, i.e., as continuations involving genuinely jointly constructed contents. If, instead, a separate level of syntactic representation is insisted upon, such data can only be treated as fragments requiring propositional reconstruction or mechanisms overriding the morphosyntactic information they bear. This is because, as shown below in (18) and earlier in (9), splicing together the two partial strings gives incorrect interpretations since elements like indexicals have to switch form in order to be interpretable as intended or for grammaticality:

(18) G: when you say it happens for a reason, it's like, it happened to get *you* off
D: off *my* ass [from Clancy et al., 1996]

A characterization of such structures as ungrammatical/performance data is a potential avenue that frameworks assuming a separate syntactic level might take (e.g., Peldszus & Schlangen, 2012; Poesio & Rieser, 2010) but, in our view, this is just an artefact of the preoccupation of standard grammars with sentential/propositional structures.

2.3 Incrementality and predictivity within the grammar architecture

Instead of data such as those in (1)–(8) and (18) being indicative of language-particular levels of syntax/morphology, use of the licensing mechanisms both by a single speaker and coupled across interlocutors, as in a DS-style dynamic account, is what enables handling of dialogue phenomena. The two architectural features of DS that underlie this dynamicity and its direct licensing of partial subsentential constructs are *incrementality* and *predictivity*, features conventionally associated only with parsers (Sturt & Lombardo, 2005). Incrementality, i.e. the licensing of subsentential elements as they become available in a time-linear manner, is an essential characteristic for the modelling of dialogue coordination. Dialogue phenomena like interruptions (as in (1)–(8) earlier), self-repair (as in (19) below), corrections (as in (20) below) etc. rely on the multi-modal incrementality of both understanding and production, in order to be modelled as making a timely contribution:

(19) "Sure enough ten minutes later the bell r-the doorbell rang"
(Schegloff et al., 1977)

(20) B to A who is pointing at Harry:
(Oxi,) tin.ACC aderfi.ACC tu [Greek] [correction]
(No,) his sister.ACC

But since, as we saw, the grammar must license such constructions, the elements it needs to manipulate must be partial/non-fully-sentential constructs. Because the syntactic licensing defined by DS is procedural and word-by-word incremental, fragments can be taken as just that, and not themselves sentential in nature. Accordingly, they may provide regular update to emerging partial structures irrespectively of who has initiated these structures, as in the fragment interruptions in (1), or when the fragment is interpreted as an extension of a non-propositional structure given in context, as in (16)–(17) and (20).

There is a lot of evidence that both comprehension and production involve predictive processes (e.g., Kutas, DeLong, & Smith, 2011; Pickering & Garrod, 2007; comprehension: Marslen-Wilson, 1973; Levy, 2008; production: Jaeger, 2010). Hence incremental integration of contents is coupled in DS with

generalized *predictivity/goal-directedness* in that the parser/generator is always predicting top-down structural goals to be achieved in the next steps (see Figure 2 earlier). Although generally the motivation for this type of architecture is efficiency considerations in parsing, in fact, coordination phenomena in dialogue can be seen to be exploiting this processing characteristic. Because DS is *bidirectional*, i.e., a model of both parsing and production mechanisms that operate concurrently in a synchronized manner, its goal-directedness/predictivity applies symmetrically *both* in parsing and generation (for modelling predictivity in production see also Demberg-Winterfors, 2010). This means that the tight coordination of turn-taking, and switches between speakers at transition relevance places (Sacks et al., 1974) can be unproblematically accounted for.

A DS-style predictive architecture for the grammar models licensing mechanisms for how interlocutors construct joint meanings, actions and syntactic constructs through interaction by means of the generation of goals to be achieved symmetrically by both the parser and the producer, the listener/parser usually awaiting input from the speaker for fulfilling these goals. Such goals are also what activates the search of the lexicon ('lexical access') in production in order to recover a suitable NL word for the concept to be conveyed. As a result, an initial listener/parser who achieves a successful lexical retrieval before processing the anticipated linguistic input provided by the original speaker can spontaneously become the producer and take over. As seen in all cases (1)–(15) and (18) above, the original listener is, indeed, using such a structural anticipation to take over and offer a completion that, even though licensed as a grammatical continuation of the initial fragment, might not necessarily be identical to the one the original speaker would have accessed had they been allowed to continue their utterance (as in (4)–(7)). And since the original speaker is licensed to operate with partial structures without having a fully-formed intention/plan as to how it will develop (as the psycholinguistic models in any case suggest), they can integrate immediately such offerings without having to be modelled as necessarily revising their original intended message. By way of illustration, we take a simplified variant of (9) (for detailed analyses see Eshghi et al. 2010, 2011; Gargett et al., 2008, 2009; Gregoromichelaki et al., 2013; Kempson et al., 2011; Purver et al., 2006, 2009, 2011):

(21) Ann: Did *you* burn
Bob: *myself*?

Here, the reconstruction of the string as **Did you burn myself*? is unacceptable (at least with a reflexive reading of *myself*)[7], illustrating the problem for purely

7. This is not a matter of perspective-taking as P&G suggest: use of *yourself* conveys a distinct meaning.

syntactic accounts of split utterances. But under DS assumptions, with representations only of structured content, not of putative structure over strings of words, the switch of person is entirely straightforward. Consider, in Figure 3, the partial tree induced by parsing Ann's utterance *Did you burn* which involves a substitution of the metavariable (**U**) contributed by *you* by the constant standing for the listener/parser (*Bob'*) and imposed predictions/goals for all the other nodes of the tree except the predicate node annotated with the concept *Burn'*.

At the point illustrated in Figure 3, Bob can complete the utterance with the reflexive as what such an expression does in general is to induce a set of actions that copy a formula from a local co-argument node onto the current node, just in case that formula satisfies the conditions set by the person, number and, here, participant role of the uttered reflexive. So, in this case, the restriction is that the metavariable stands for a local co-argument that is *currently* the speaker (in contrast to *yourself* which would require a local co-argument that is currently the addressee). This indeed will satisfy the extant prediction/goal on this node *(?Ty(e))* and provide an appropriate value for the metavariable U as can be seen in Figure 4.

Hence, the absence of a "syntactic" level of representation distinct from that of semantic representations allows the direct successful integration of such fragments through the grammatical mechanisms themselves, rather than necessitating their analysis as sentential ellipsis. In addition, this predictive feature of DS is fully compatible with observations in interactional accounts of conversation where it is noted that 'anticipatory planning'/audience design takes place (Arundale & Good, 2002), this "planning" here supported by low-level architectural features of the grammar. Furthermore, given the format of the semantic representations employed by DS (trees annotated with conceptual content in functor-argument format), a second stage of composition of what has been built incrementally also occurs at constituent boundaries thus giving the opportunity for 'retroactive assessment' of the derived content in a new context (as noted again by Arundale & Good, 2002). The output tree resulting from the parse/production is shown in Figure 5.

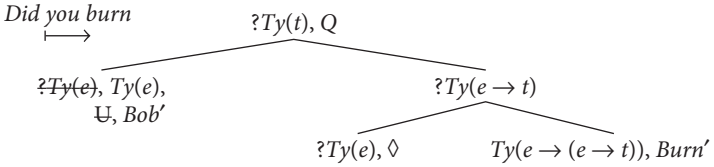

Figure 3. Parsing/producing Ann's utterance *Did you burn*

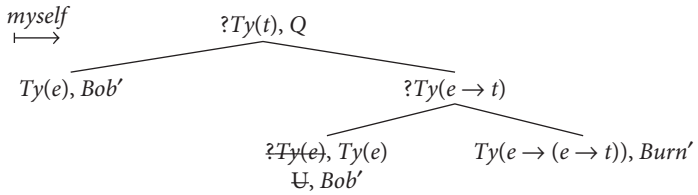

Figure 4. Parsing/producing Bob's utterance *myself*

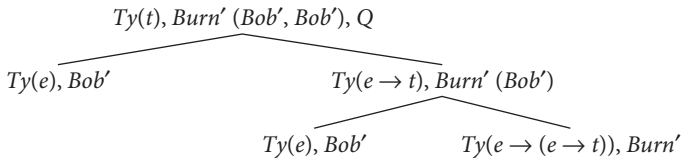

Figure 5. Interpretation constructed by speaker and hearer for joint utterance *Did you burn myself?*

Modular approaches to the grammar/pragmatics interface propose that the grammar delivers underspecified propositional representations as input to pragmatic processes that achieve full interpretations and discourse integration (see e.g. Schlangen, 2003, following an SDRT model). However, an essential feature of language use in dialogue is the observation that on-going interaction and feedback shapes utterances and their contents (Goodwin, 1981), hence it is essential that the grammar does not have to license whole propositional units before semantic and pragmatic evaluation can take place. And this is the strategy DS adopts, operating directly with partial constructs whether induced by speaker or listener: in either case such constructs are fully licensed by the antecedently constructed context and thereupon integrated into the emergent tree by updating it, without having to consider such fragments as sentences which happen to be elliptical, or as sentences which are in some sense not well-formed despite their success as utterances.

Thus DS reflects directly and explicitly, from within the grammar itself, how the possibility arises for joint-construction of utterances, meanings and structures in dialogue and how this is achieved. And these explanations are fundamentally based on the same (subpersonal) mechanisms underlying language structure: since the grammar licenses partial, incrementally constructed objects, speakers can start an utterance without a fully-formed intention/plan as to how it will develop relying on feedback from the hearer to shape its structure and its construal. Moreover, the syntactic constraints themselves can be exploited ad hoc as a source of "conditional relevance" (see e.g. Schegloff, 2007) by setting up sequences (joint speech acts or 'adjacency pairs') sub-sententially without involving speech acts

with complete propositional contents (see (12)–(14) above). Thus, syntactic devices and their goal-directed, projectible nature can be manipulated by interlocutors to manage conversational organization. Given these results, in our view, the dichotomy between *language*$_S$ (language structure) and *language*$_U$ (language use) postulated in standard linguistic models does not withstand the test of application in dialogue, the primary site of language use. Instead, the grammar has to be seen as underpinning communication with, as DS suggests, the syntactic architecture viewed in dynamic terms as the crystallization of action patterns derived from language use and wider cognitive/social considerations.

3. Conclusion: low-level mechanisms for linguistic coordination and emergent intentions

Despite the widespread assumption in pragmatic theorizing that successful retrieval of intended propositional contents is the sine-qua-non of communicative success, the fundamental role of intention recognition and the primary significance of speaker meaning in dialogue has been disputed in interactional accounts of communication. In these, intentions, instead of assuming causal/explanatory force can be characterized as "emergent" in that the participants can be taken to jointly construct the content of the interaction (Gibbs, 2001; Haugh, 2008; Mills, in press; Mills & Gregoromichelaki, 2010). This aspect of joint action has been explicated via the assumption of the "non-summativity of dyadic cognition" (Arundale, 2008; Arundale & Good, 2002; Haugh, 2012; Haugh & Jaszczolt, 2012) or in terms of "interactive emergence" (A. Clark, 1997; Gibbs, 2001). This view gains experimental backing through the observation of the differential performance of participants vs. over-hearers in conversation (Clark & Schaefer, 1987; Schober & Clark, 1989) and the gradual emergence of intentional explanations in task-oriented dialogue (Mills & Gregoromichelaki, 2010). Such views have serious consequences for dialogue models. Typically, these are serial, modular and operate on complete utterances underpinned by a speaker plan and its recognition, all in contra-distinction to the evidence discussed above. Moreover, the output of each module is the input for another with speaking and listening seen as autonomous processes. This directly conflicts with the observation that, in ordinary conversation, utterances are shaped genuinely incrementally and "opportunistically" according to feedback by the interlocutor (as already pointed out by Clark, 1996) thus genuinely engendering co-constructions of utterances, structures and meanings (see e.g. Lerner, 2004).

In our view, the main reason for this inadequacy in dialogue modelling are methodological assumptions justified by the competence/performance distinction,

separating the grammar from the parser/generator and the pragmatic modules, with the result that the grammatical models employed lack the capability to fully manipulate and integrate partial structures in an incremental manner (for recent incremental systems see Peldszus & Schlangen, 2012; Poesio and Rieser, 2011). In sharp contrast, a model which defines "syntax" as mechanisms reflecting real-time processing provides a wholly natural basis for the incremental co-construction of compound utterances without need of mind-reading capability by either participant. In a similar vein, the P&G model also presents a welcome new addition to socio-cultural and psychological accounts of emergent intentions in that understanding and control of goal-directed action does not have to be seen as mediated by costly inferential mechanisms modelled on the basis of offline reasoning processes. However, the "simulation route" of the P&G model has to be seen, in our view, as one of several means available to interlocutors for achieving coordination. Additional means such as priming, alignment, entrainment, mirroring (Böckler et al., 2010; Pickering & Garrod, 2004), and, importantly, the grammar as modelled in a dynamic framework like DS are also, perhaps more readily, available mechanisms.

Acknowledgements

We gratefully acknowledge various helpful discussions with Gregory J. Mills, Ronnie Cann, Matt Purver, Pat Healey, Jonathan Ginzburg, Yo Sato, and Graham White. We thank two anonymous reviewers and the editor J. P. de Ruiter who provided valuable comments and observations. This work was supported by grant ESRC RES-062-23-0962.

References

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238.

Apperly, I. (2011). *Mindreaders: The cognitive basis of "Theory of Mind"*. Hove and New York: Psychology Press.

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *106*, 1093–1108.

Airenti, G., Bara, B., & Colombetti, M. (1993). Failures, exploitations and deceits in communication. *Journal of Pragmatics, 20*, 303–326.

Arundale, R. B. (2008). Against (Gricean) intentions at the heart of human interaction. *Intercultural Pragmatics*, *5*(2), 229–258.

Arundale, R. B., & Good, D. (2002). Boundaries and sequences in studying conversation. In A. Fetzer, & C. Meierkord (Eds.), *Rethinking sequentiality: Linguistics meets conversational interaction* (pp. 121–150). Amsterdam: John Benjamins.

Austin, J. L. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462–479.

Bach, K., & Harnish, R. M. (1982). *Linguistic communication and speech acts*. Cambridge, MA: MIT Press.

Bermúdez, J. L. (2003). The domain of folk psychology. In A. O' Hear (Ed.), *Minds and Persons* (pp. 25–48). Cambridge: Cambridge University Press.

Böckler, A., Knoblich, G., & Sebanz, N. (2010). Socializing Cognition. In B.M. Glatzeder, V. Goel, & A. Müller (Eds.), *Towards a theory of thinking: Building blocks for a conceptual framework* (pp. 233–250). Berlin: Springer.

Burnard, L. (2000). *Reference guide for the British National Corpus (World Edition)*. Oxford: Oxford University Computing Services.

Cann, R., Kempson, R., & Marten, L. (2005). *The dynamics of language*. Oxford: Elsevier.

Chatzikyriakidis, S., & Kempson, R. (2011). Standard modern and pontic Greek person restrictions: A feature-free dynamic account. *Journal of Greek Linguistics, 11*(2): 127–166.

Clancy, P. M., Thompson, S. A., Suzukia, R., & Tao, H. (1996). The conversational use of reactive tokens in English, Japanese, and Mandarin. *Journal of Pragmatics*, *26*(3), 355–387.

Clark, A. (1997). *Being There*. Cambridge, MA: MIT Press.

Clark, H. H. (1992). *Arenas of Language Use*. Chicago: University of Chicago Press & CSLI.

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

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In: A. K. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding*. Cambridge: Cambridge University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speech. *Cognition*, *84*, 73–111.

Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language*, 26, 209–225.

Davies, M., & Stone, T. (1995). *Folk Psychology*. Oxford: Basil Blackwell.

Davidson, D. (1980). Towards a unified theory of meaning and action. *Grazer Philosophische Studien*, *11*, 1–12.

De Bruin, L. C., Strijbos, D. W., & Slors, M. (2011). Early social cognition: Alternatives to implicit mindreading. *Review of Philosophy & Psychology*, *2*(3), 499–517.

De Ruiter, J., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speakers turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515–535.

De Ruiter, J. P., Noordzij, M. L., Newman-Norlund, S. E., Hagoort, P., & Toni, I. (2007). On the origin of intentions. In P. Haggart, Y. Rosetti, & M. Kawato (Eds.), *Attention & Performance XXII. Sensorimotor foundation of higher cognition* (pp. 593–610). Oxford: Oxford University Press.

De Ruiter, J. P., Noordzij, M. L., Newman-Norlund, S., Newman-Norlund, R., Hagoort, P. & Levinson, S. C. (2010). Exploring the cognitive infrastructure of communication. *Interaction Studies*, *11*, 51–77.

Demberg-Winterfors, V. (2010). *A broad-coverage model of prediction in human sentence processing*. PhD thesis. University of Edinburgh.

Eshghi, A., Healey, P. G. T., Purver, M., Howes, C., Gregoromichelaki, E., & Kempson, R. (2010). Incremental turn processing in dialogue. In *Proceedings of the 16th Annual Conference on Architectures and Mechanisms for Language Processing (AmLAP)*, York, UK, September 2010, p. 170.

Eshghi, A., Purver, M. & Hough, J. (2011). *Dylan: Parser for dynamic syntax*. Technical Report EECSRR-11-05, School of Electronic Engineering, Computer Science, Queen Mary University of London. London: Queen Mary University of London.

Ferrara, K. (1992). The interactive achievement of a sentence: Joint productions in therapeutic discourse. *Discourse Processes*, *15*(2), 207–228.

Jaeger, T. F., & Ferreira, V. (2013). Seeking predictions from a predictive framework. Commentary in Pickering and Garrod (2013). *Behavioral and Brain Sciences*, 36, 359–360.

Fox, B., A. & Jasperson, R. (1995). A syntactic exploration of repair in English conversation. In P. W. Davis (Ed.), *Descriptive and theoretical modes in the alternative linguistics* (pp. 77–134). Amsterdam: John Benjamins.

Gann, T. M., & Barr, D. J. (2012). Speaking from experience: Audience design as expert performance. *Language and Cognitive Processes*. DOI:10.1080/01690965.2011.641388

Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 659–669.

Gargett, A., Gregoromichelaki, E., Howes, C., & Sato Y. (2008). Dialogue-grammar correspondence in dynamic syntax. In *Proceedings of the 12th SemDial (LonDial)*, London, pp. 43–50.

Gargett, A., Gregoromichelaki, E., Kempson, R., Purver, M., & Sato Y. (2009). Grammar Resources for Modelling Dialogue Dynamically. In *Journal of Cognitive Neurodynamics*, *3*(4), 347–363.

Garrod S., & Anderson., A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition*, *27*, 181–218.

Gibbs, R.W., Jr. (2001). Intentions as emergent products of social interactions. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and Intentionality* (pp. 105–122). Cambridge, MA: MIT Press.

Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Grice, P. (1975). Logic and Conversation. In P. Cole, & J.Morgan (Eds.), *Syntax and Semantics, Volume 3: Speech acts* (pp. 41–58). New York: Academic Press.

Gregoromichelaki, E. (2013). Clitic Left Dislocation and Clitic Doubling: A Dynamic Perspective on Left-Right Asymmetries in Greek. In G. Webelhuth, M. Sailer, H. Walker (Eds.), *Rightward movement in a comparative perspective* (pp. 321–368). Amsterdam: John Benjamins.

Gregoromichelaki, E., Cann, R., & Kempson R. (2013). On coordination in dialogue: Subsentential speech and its implications. In L. Goldstein (Ed.), *On Brevity* (pp. 53–73). Oxford: Oxford University Press.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*(3), 377–96.

Guhe, M. (2007). *Incremental Conceptualization for Language Production*. Hillsdale NJ: Lawrence Erlbaum Associates.

Guhe, M., Habel, C., & Tappe, H. (2000). Incremental event conceptualization and natural language generation in monitoring environments. In *Proceedings of the first international*

conference on Natural language generation, Volume 14 (pp. 85–92). Association for Computational Linguistics.
Haugh, M. (2008). Intention in pragmatics. *Intercultural Pragmatics*, 5(2). 99–110.
Haugh, M. (2012). Conversational interaction. In K. Allan, & K. M. Jaszczolt (Eds.), *Cambridge handbook of pragmatics* (pp. 251–274). Cambridge: Cambridge University Press.
Haugh, M., & Jaszczolt, K. M. (2012). Speaker intentions and intentionality. In K. Allan, & K. M. Jaszczolt (Eds.), *Cambridge handbook of pragmatics* (pp. 87–112). Cambridge: Cambridge University Press.
Hauser et al. 2002Hauser, M., Chomsky, N., & Fitch, W. (2002), 'The faculty of language: What is it, who has it, and how did it evolve?', *Science, 298*(5598), 1569–1579.
Henetz, T., & Clark, H. H. (2011). Managing delays in speaking. Paper presented at the IPrA meeting, Manchester July 2011.
Howes, C., Purver, M., Healey, P. G. T., Mills, G. J. (2011). On incrementality in dialogue: Evidence from compound contributions. *Dialogue and Discourse*, *2*(1), 279–311.
Hurley, S. L. (2008) The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral and Brain Sciences, 31*(1), 1–58.
Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology, 61*, 23–62.
Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.
Jaszczolt, K. M. (2005). *Default Semantics: Foundations of a compositional theory of acts of communication.* Oxford: Oxford University Press.
Kaye, J. (1989). *Phonology: A cognitive view*. New York: Lawrence Erlbaum.
Kempson, R., Meyer-Viol, W., & Gabbay, D. (2001). *Dynamic syntax*. Oxford: Blackwell.
Kempson, R., Gregoromichelaki, E., & Howes, C. (Eds.) (2011). *The dynamics of lexical interfaces*. CSLI – Studies in Constraint Based Lexicalism.
Kempson, R., Gregoromichelaki, E., & Chatzikyriakidis, S. (2012). Joint utterances in Greek: their implications for linguistic modelling. In *Proceedings of 33*rd *Annual Linguistics Meeting "Syntactic Theories and the Syntax of Greek"*. Thessaloniki, 26–27 April 2012, 107–129.
Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press.
Lahiri, A., & Plank, G. (2010). Phonological phrasing: the judgement of history confirmed through experimentation. *Proceedings of the Philological Society 108*(3), 370–398.
Lerner, G. H. (2004). Collaborative turn sequences. In G. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 225–256). Amsterdam: John Benjamins.
Levinson, S. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177.
Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*, 522–523.
McDowell, J. (1998). *Mind, value and reality*. Cambridge, Mass.: Harvard University Press.
McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179–218.
Millikan, R. G. (1984). *Language thought and other biological categories*. Cambridge, MA: MIT Press.

Mills, G. J. (2011). The emergence of procedural conventions in dialogue. In L. Carlson, C. Hoelscher, & T.F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 471–476). Austin, TX: Cognitive Science Society.

Mills, G. J. (in press). Dialogue in joint activity: Complementarity, convergence and conventionalization. In: *New Ideas in Psychology*, Available online 24 April 2013. http://dx.doi.org/10.1016/i.newideapsych.2013.03.006.

Mills, G. J., & Gregoromichelaki, E. (2010). Establishing coherence in dialogue: Sequentiality, intentions and negotiation. In P. Łupkowski & M. Purver (Eds.), *Aspects of semantics and pragmatics of dialogue (SemDial)* (pp. 17–24). Poznań: Polish Society for Cognitive Science.

Patel, A. D. (2008). *Music, language, and the brain.* Oxford: Oxford University Press.

Peldszus, A., & Schlangen D. (2012). Incremental construction of robust but deep semantic representations for use in responsive dialogue systems. In *Proceedings of the COLING 2012 Workshop on Advances in discourse analysis and its computational aspects (ADACA)*, Mumbai, India, December 2012, pp. 59–75.

Pezzulo G. (2011). Grounding procedural and declarative knowledge in sensorimotor anticipation. *Mind and Language*, *26*(1), 78–114.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–226.

Pickering, M.J., & Garrod, S., (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Science, 11*, 105–110.

Pickering, M., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36, 329–347.

Piwek, P. (2011). Dialogue structure and logical expressivism. *Synthese, 183*, 33–58.

Poesio, M., & Rieser H. (2010). Completions, coordination, and alignment in dialogue. *Dialogue and Discourse 1*(1), 1–89.

Purver, M., Cann, R., & Kempson, R. (2006). Grammars as parsers: Meeting the dialogue challenge. *Research on Language and Computation*, 4(2/3), 289–326.

Purver, M., Howes, C., Gregoromichelaki, E., & Healey, P. G. T. (2009). Split utterances in dialogue: A corpus study. In *Proceedings of SIGDIAL 2009* (pp. 262–271). London, UK: Association for Computational Linguistics.

Purver, M., Eshghi, A., & Hough, J. (2011). Incremental semantic construction in a dialogue system. In J. Bos, & S. Pulman (Eds.), *Proceedings of the 9th International Conference on Computational Semantics* (pp. 365–369). Oxford, UK.

Pulvermüller, F. (2010). Brain embodiment of syntax and grammar: discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain and Language*, *112*(3), 167–179.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.

Schegloff 2007Schegloff, E. (2007). *Sequence organization in interaction: A primer in conversation analysis I*. Cambridge: Cambridge University Press.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382.

Schlangen, D. (2003). *A Coherence-Based Approach to the Interpretation of Non-Sentential Utterances in Dialogue*. PhD thesis, University of Edinburgh.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhears. *Cognitive Psychology, 21*(2), 211–232.
Searle 1969Searle, J. R. (1969). *Speech Acts*. Cambridge: Cambridge University Press.
Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Science*, *10*(2), 70–76.
Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell.
Sturt, P., & Lombardo, V. (2005). Processing coordinate structures: Incrementality and connectedness. *Cognitive Science*, *29*, 291–305.
Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated action*. Cambridge: Cambridge University Press.
Tettamanti, M., & Moro, A. (2012). Can syntax appear in a mirror (system)? *Cortex*, *48*, 923–945.
Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, l–71.
Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 593–602.