Pragmatics 24:2.275-296 (2014)
International Pragmatics Association
DOI: 10.1075/prag.24.2.05gio

GRAMMAR, CONTEXT AND THE HEARER: A PROPOSAL FOR AN ADDRESSEE-ORIENTED MODEL OF FUNCTIONAL DISCOURSE GRAMMAR[1]

Riccardo Giomi

Abstract

This paper addresses two issues related to the overarching question of how to integrate Functional Discourse Grammar (FDG) into a wider theory of verbal interaction (Hengeveld and Mackenzie 2008: 1). First, it proposes an addressee-oriented version of the Grammatical Component, presenting a first attempt to develop an FDG account of language comprehension; second, it aims to shed light on the interaction between the Grammatical and Contextual Components of FDG by exploring this crucial aspect of verbal communication from the perspective of the addressee.

Keywords: Functional Discourse Grammar; Language parsing; Context; Discourse; Inference.

1. Introduction

One basic assumption in FDG is that "a model of grammar will be more effective the more its organization resembles language processing in the individual" (Hengeveld and Mackenzie 2008: 1-2). Although *processing* does not necessarily mean *producing*, since Hengeveld and Mackenzie's presentation of the theory virtually all contributions to the FDG framework have exclusively focussed on the individual *speaker*'s use of language. Yet, FDG is not a model of the speaker but the Grammatical Component of a "wider theory of verbal interaction". It thus provides, by its very nature, means of developing a general theory in which both speech participants are represented:

> [...] although the presentation of the FDG model will focus on the generation of utterances, the model could in principle be turned on its head to account for the parsing of utterances. It is clear that listeners analyse phonetic input into phonological representations, which are subsequently grouped into morphosyntactic constituents, from which meaningful representations are then constructed. (Hengeveld and Mackenzie 2008: 2)

[1] I am grateful to all the participants in the workshop *The Interaction between the Grammatical Component and the Contextual Component* for their insightful comments and suggestions. Special thanks are due to Lachlan Mackenzie for commenting on previous versions of this paper. All errors are mine.

The general purpose of this paper is to contribute to the development of such a general FDG theory of language, which of course requires the "standard" speaker-oriented model to be integrated with an explicit account of language comprehension. The architecture of this addressee-oriented model of FDG will be outlined in section 2. Since linguistic interaction is always supported by a specific context, which is cooperatively created by the speech participants in the course of the exchange, another crucial element of a comprehensive functional-discursive theory is a good understanding of the role played by the Contextual Component (henceforth CtxC) and of its relation to the Grammatical Component (GC). This subject is treated in section 3, again from the perspective of language understanding. In section 4 I will discuss the implications of the proposed approach for the nature and theoretical status of the CtxC (4.1) and bring together the speaker-oriened and the hearer-oriented versions of FDG to describe the whole process of transfer of information from the speaker to the addressee (4.2). The conclusions of the paper are summarized in section 5.

2. Grammar and the hearer

Looking at grammar from the vantage point of the addressee means reversing the perspective from which FDG has been conceived. However, it seems clear that an addressee-oriented FDG model should mirror the speaker-oriented one in its organization: Assuming the contrary would imply that language users have two different grammars in their minds, one for production and one for comprehension, while it seems much less counter-intuitive for language users to have only one grammar at their disposal, which functions in two different ways depending on the use it is put to.[2] This is suggested by the fact that (in normal conditions) any competent speaker of a language only produces structures she[3] would be able to process as a hearer, which is only possible if the same grammatical levels are relevant in production and understanding and if the same sets of primitives (i.e. of linguistic units and rules) are available in both processes.

On the other hand, it should be noted that language users are able to understand syntactic constructions or lexemes they would never use as speakers, and even constructions or lexemes they hear for the first time. Therefore, the inventories of primitives available for comprehension must necessarily be thought of as open classes: Were it not so, language acquisition would not be possible at all. In other words, an individual's grammar is more "flexible" when used for language parsing than it is in production, at least as concerns the primitives it can process and the way it interacts with the CtxC – since it is normally the context that allows listeners to "get the sense" of words and structures they have never heard before. Another important factor in interpreting unknown primitives is the recognition of formal and functional similarities with the primitives the GC already has at its disposal, which suggests that, at least for pragmatic and semantc frames and morphosyntactic and phonolgical templates, a precise set of constraints on the theoretically possible primitives may exist for each language.

---

[2] It is thus only for the sake of simplicity that, in this article, I will refer to the two versions of the model as "the hearer's/the speaker's grammar" or "the hearer's/the speaker's GC".

[3] To avoid confusion, I use feminine pronouns to refer to the speaker and masculine pronouns for the hearer.

Where we may expect to find crucial differences is between the kind of operations performed in production and in comprehension, as speakers and listeners have exactly opposite goals in language processing: The speaker's GC translates prelinguistic communicative intentions into grammatical representations, then sends an instruction to Articulation to produce an output consisting of sound waves (or visually perceivable signs, in gestual and written language); conversely, in comprehension, sounds waves (or visible signs) are perceived and transmitted to the brain, in the form of electric impulses, by the ear and the acoustic nerve (or the eye and the optic nerve), so that one may wish to talk about an extra-grammatical operation of "Transmission". This operation generates the phonetic Input to be analysed by the GC, which translates it into a series of linguistic representations from which the communicative intentions of the speaker are ultimately reconstructed. Since FDG takes language production to be a top-down process, starting with a communicative intention and ending with a perceivable output, it follows that language understanding, which reconstructs the former starting from the latter, should be seen as a bottom-up process. More specifically, the GC decodes the Input to generate phonological and morphosyntactic configurations, then processes these to reconstruct their semantic and pragmatic content. That is to say, it starts with Phonological and Morphosyntactic *Decoding* and continues with the elaboration of representational and interpersonal configurations, which I will refer to as *Interpretation*.

Similarly to production, comprehension develops in time through a series of stages in such a way that it is able to handle incrementality. Evidence for this comes from neurolinguistic research, supporting the view that addressees sequentially create several levels of representation (Bornkessel-Schlesewsky and Schlesewsky 2009: 88). There also is empirical evidence for an incremental parsing in which information from the Decoding levels is sent ahead to Interpretation before these levels have been completely elaborated: Were it not so, it would be impossible for listeners to figure out the full semantic and pragmatic content of an utterance before the speaker has finished producing it, and experience tells us that this is not the case. Language parsing thus seems to obey the "depth-first" principle (Hengeveld and Mackenzie 2008: 23-25), though reversing the "direction" of the prediction. In addition, Interpretation must be fed by the CtxC, since it is mainly by computing contextual information that listeners can infer the whole meaning of an utterance before it is complete.

At the same time, if we consider the nature and function of the primitives it makes use of, it appears that Interpretation also requires an input from the Conceptual Component (henceforth CptC), which is necessary for linking the linguistic material it handles to the corresponding ideational and interactive conceptual material (see Hengeveld and Mackenzie 2008: 7). What lies behind the pragmatic and semantic distinctions relevant to a language is an extremely wide and complex set of non-linguistic mental constructs, including abstract representations of the real world and of human psychological/emotional experience, which is shared by the collectivity of speakers of a language[4] and is created by the millenary stratification of social, psychological and intellectual changes that any human community inevitably undergoes. The grammatical distinctions operated at Formulation and Interpretation are only meaningful (and their existence is only justified) insofar as they reflect such non-

[4] Without ignoring the fact that individual speakers have their own, personal understanding of those conceptual and psychological constructs and of the associated linguistic items.

linguistic conceptual representations, which implies that these must necessarily be evoked each time that the corresponding semantic or interpersonal primitives are used, either in production or in comprehension.[5] This is not to suggest that conceptual content should itself be integrated into the grammatical representation: It simply follows from the observation that one cannot use or interpret a certain lexeme or grammatical structure properly if one fails to conceive the related notion or to associate it with that linguistic item. In other words, such a thing as "purely linguistic" meaning is hardly conceivable at all, which means that the operations of Formulation and Interpretation always need to have access to the inventory of conceptual representations underlying the grammatical distinctions relevant to the language. It is a major task of the CptC to provide Formulation and Interpretation with the means of associating a given conceptual element with the linguistic sign used to represent it.

There also is another sense in which the CptC directly feeds Interpretation. Here again, a process performed in comprehension parallels an aspect of production, although in a less systematic way: If Formulation is always triggered by communicative intentions, Interpretation can be strongly influenced by the addressee's expectations about the speaker's communicative intention (which, in turn, largely depend on the information contained in the CtxC). For instance, the hearer may fail to recognize an Irony operator at the Interpersonal Level, thus dramatically misunderstanding the meaning of the Discourse Act: One language in which this may happen is English, which can express irony by means of an intonation contour which is also used as a strategy for signaling emphasis (as in "That IS interesting", cf. Hengeveld and Mackenzie 2008: 65, 441).

It thus seems necessary to revise our first hypothesis about the functioning of the addressee's GC: Rather than a wholly bottom-up process, language parsing appears to be partially a bottom-up and partially a top-down process, fed by both the Input and the Conceptual Components.

Note finally that, although language parsing has no "material" output – the linguistic representations it generates are not further translated into any perceivable sign – it still clearly has its own impact on the CtxC, since a linguistic expression (of any kind) cannot be said to have really updated a context if it has not been perceived and understood by the addressee(s). This seems to be also suggested by Hengeveld and Mackenzie (2008: 14):

> The various levels of representation within the [speaker's] grammar feed into the Contextual Component, thus enabling subsequent reference to the various kinds of entity relevant at each of these levels once they are introduced into the discourse. [...] Note that the representation of these feeding relations [...] is a simplification when looked at from the perspective of the language user. In order to create a contextual specification, the Addressee has to reconstruct all the levels of representation within the grammar on the basis of the actual output of that grammar, i.e. the phonetic utterance. Since in this book we restrict ourselves to the perspective of language production and concentrate on the Grammatical Component, we abstract away from this complication by provisionally assuming direct feeding relationships between the [speaker's] Grammatical Component and the Contextual Component.

[5] This seems to be also suggested by O'Neill (this issue) as he postulates a direct "connection between concepts and primitives".

The role of the addressee in the creation of a context relevant to the ongoing discourse is explicitly recognized by Mackenzie (2012), who, drawing on Pickering and Garrod's (2004, 2005) "mechanistic" theory of dialogue, proposes to re-interpret the Contextual Component of FDG "as an implicit common ground and as a conduit for the interactive alignment of grammatical processes", which is shared and jointly constructed by the speech participants as the exchange unfolds. As will become appearent in the next section, this is also the way the CtxC is conceived of in this paper.

The considerations developed in this section are summarized in Figure 1 below, which integrates the proposed addressee-oriented FDG model with its speaker-oriented counterpart. Note that Decoding is the only operation that does not receive any input from the CtxC, while Interpretation is not assumed to feed back directly into the CptC. These topics will be discussed in detail in sections 3 and 4 respectively.

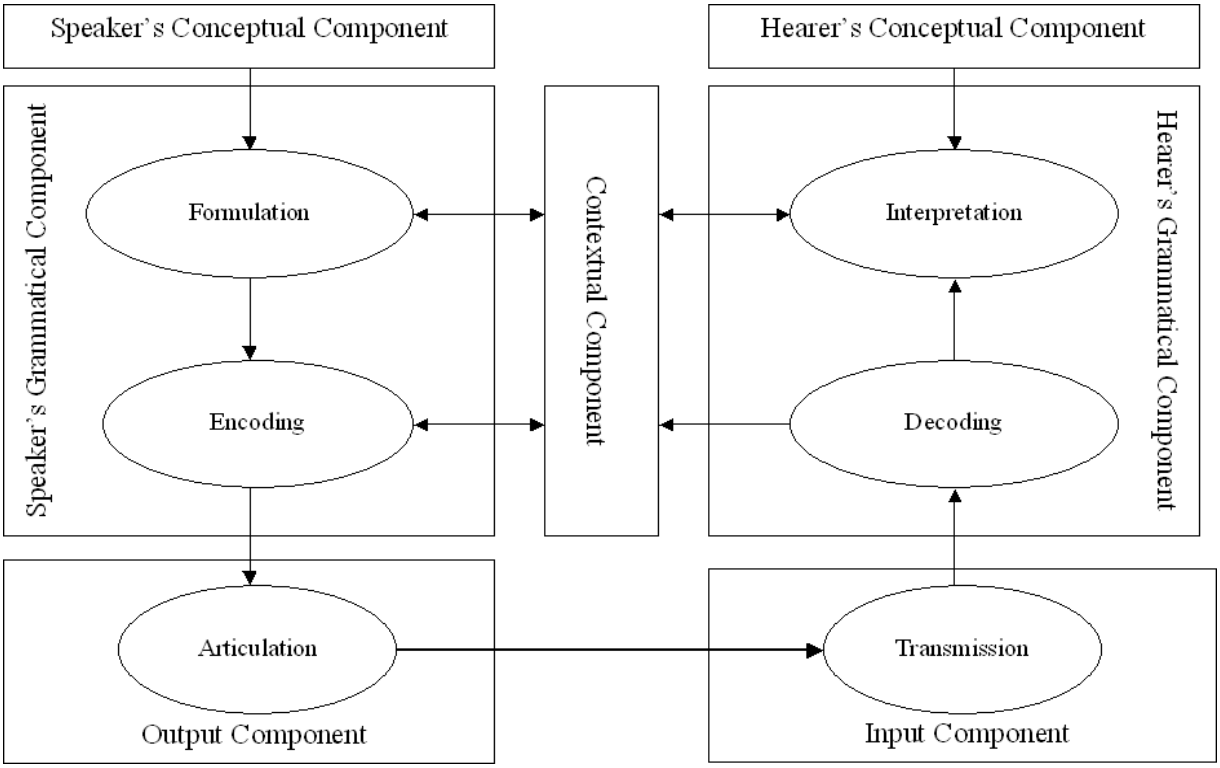

Figure 1: Preliminary proposal for an integrated model of FDG

3. The hearer's grammar and the context

This section addresses two questions: How the CtxC contributes to the elaboration of appropriate linguistic representations by Decoding and Interpretation (3.1) and how and where the discourse fragment corresponding to these representations enters the CtxC (3.2). For both these matters, it is crucial that the CtxC not be treated as a whole: On the one hand, we need to identify and categorize the elements of the CtxC that can serve as inputs for Decoding and Interpretation, in order to understand how the two grammatical

operations handle contextual information; on the other, grammatical and psycholinguistic evidence suggests that linguistic information from each level is incorporated into the CtxC independently. The analysis of contextual categories will be based on Connolly's (2007; this issue) Extended Model of Context (EMC).

3.1. *From context to grammar*

3.1.1. From context to Decoding

When comparing the processes of language production and comprehension, we first need to acknowledge that Encoding and Decoding are much less sensitive to contextual information than Formulation and Interpretation. Decoding, in particular, may not be context-sensitive at all, all issues related to the role of the CtxC becoming relevant at Interpretation, when Decoding has already played its part. In fact, the impossibility of accessing the contextual information necessary for construing meaningful representations does not deny Decoding the potential to generate the correct representations: Provided that the input-utterance is correctly perceived (by a fluent speaker of the language), Phonological Decoding always generates a representation which perfectly matches the one generated by Encoding, and this, in turn, is translated into a morphosyntactic structure which also matches the one given in the speaker's mind.

One special case in this regard is that of morphosyntactic ambiguity (as in English "old men and women"), where the phonological configuration serving as input for Morphosyntactic Decoding can yield two alternative morphosyntactic structures. Under a modular view of language parsing (i.e. one in which only one morphosyntactic representation at time is built), it is possible that in such cases the first structure generated by Morphosyntactic Decoding is not the same as the one given in the speaker's GC. But if this happens, a repair sequence will be initiated as soon as the representational and interpersonal configurations associated to that morphosyntactic structure are checked against the available contextual information and found inappropriate or irrelevant: it is at this point that a new morphosyntactic representation is created, which will now correspond to the one generated by the speaker's Encoder and whose semantic and pragmatic content will expectedly represent a more contextually appropriate output. Thus, although the ambiguity originates at the Morphosyntactic Level, disambiguation only depends on the interaction between the CtxC and Interpretation – where meaningful linguistic items become available – while the two alternative morphosyntactic configurations, being a-functional in nature, cannot be selected or discarded for their appropriateness to the context.

Let us now consider some optional morphosyntactic alternations that are invoked in other contributions to this volume in order to show how the CtxC affects the speaker's grammar through the CptC: Non-raising vs. raising constructions (García Velasco), full clauses vs. nominalizations (Van de Velde), and active vs. passive constructions (Keizer). Since these alternations are motivated by the activation status of referents (Chafe 1994; Lambrecht 1994), it may happen that the speaker treats as contextually active a referent that is not readily accessible for the addressee – causing some sense of anomaly or surprise – but this has no consequences for the representations generated by Decoding. And the same holds for systematic, rule-

governed phenomena in which contextual information directly influences Encoding, such as anaphora and "contextual agreement" at ML (Hengeveld and Mackenzie, this issue) or the choice between strong and weak pronouns at PL: With these, a mismatch in the accessiblity of contextual information for the two dialogists can actually prevent the addressee from understanding (some aspect of) the speaker's utterance, but in none of these cases does the addressee's failure to retrieve the relevant information prevent Decoding from regularly parsing the input.

To find an example of interaction between the CtxC and Decoding, we have to look at non-standard communicative situations. One case in point is when (a part of) the utterance is not correctly perceived, say, because of a noisy environment or because the addressee has hearing problems – which are variables of the "narrower physical context", in Connolly's EMC. However, in such cases the relevant aspect of the CtxC does not impact Decoding directly, but first affects Transmission and then, as a consequence, Decoding and Interpretation.[6]

A different case is where the hearer correctly perceives the input-utterance but fails to decode (some aspect of) its phonological and/or morphosyntactic structure, as may happen if the hearer is not a fluent speaker, being a child or an L2 speaker or having some language deficiency. Applying the EMC categories strictly, these socio-cultural or physical attributes of the hearer are variables of the narrower situational context, which may happen to directly affect Decoding – and more specifically, the primitives used in that operation. For instance, the hearer may fail to recognize a morphosyntactic template or the distribution of phonological operators, with consequences for the subsequent grammatical operations. Consider the following two sentences from Portuguese, a language that distinguishes yes/no Interrogatives from Declaratives only by means of a phonological operator:

(1) *O* *Jorge também* *vai.*
DEF.M Jorge also go.3.SG.PRS
'Jorge is going too.'

(2) *O Jorge também vai?*
'Is Jorge going too?'

The PL and IL representations of (1) and (2) are given in (3)-(4) and (5)-(6) respectively:

(3) PL: (U_i: (fIP_i: /uʒɔrʒətɐ̃bɐ̃jvaj/ (IP_i)) (U_i))

(4) IL: (M_I: (A_I: [(F_I: DECL (F_I)) (P_I)$_S$ (P_J)$_A$ (C_I:[(T_I) (R_I: Jorge (R_I))] (C_I))] (A_I)) (M_I))

(5) PL: (U_i: (rIP_i: /uʒɔrʒətɐ̃bɐ̃jvaj/ (IP_i)) (U_i))

(6) IL: (M_I: (A_I: [(F_I: INTER (F_I)) (P_I)$_S$ (P_J)$_A$ (C_I:[(T_I) (R_I: Jorge (R_I))] (C_I))] (A_I)) (M_I))

[6] Note that Transmission generates a phonetic Input independently from the subsequent grammatical operations, as shown by the fact that even uninterpretable babbling can be quoted back by the hearer.

Failure to recognize a falling or rising tone does not block the development of morphosyntactic and representational configurations (which are the same irrespective of intonation), but denies Interpretation the possibility to generate a full interpersonal representation, as the Illocution slot will remain unspecified. Even this, however, is not an uncontroversial counterexample to the idea that Decoding is never directly affected by contextual information, as it can reasonably be argued that the grammar of non-fluent language users is simply not the same as that of adult L1 speakers. In this perspective, the impossibility of decoding the phonological structure generated by the speaker does not depend on a feature of the situational context, but on the incompleteness of the inventory of primitives available to Phonological Decoding.

It is nonetheless interesting to consider the theoretical implications of (3)-(6), which show that the Phonological Level can send a direct output to Interpretation exactly as, in language production, information from both Formulation levels can be sent to Morphosyntactic or Phonological Encoding, circumventing the level(s) that lie(s) in between. Thus, it is not only the depth-first principle that applies to both production and comprehension (see section 2), but also the "maximal depth" principle, which states that "[o]nly those levels of representation that are relevant for the build-up of (a certain aspect of) an utterance are used in the production of that (aspect of the) utterance" (Hengeveld and Mackenzie 2008: 25).

3.1.2. From context to Interpretation

We have seen above that, in parsing an utterance, Decoding only needs to have at its disposal all the Phonological and Morphosyntactic primitives contained in that utterance. The reason for this is that ML and PL do not deal in any sense with the meaning of the stuctures they process. By contrast, the structures developed by Formulation are contentful representations, in which each linguistic unit corresponds to (a property of) a concrete or abstract entity of the extra-linguistic world or to a communicative action performed by the speaker: Consequently, in order to generate complete representations Interpretation often needs to have access to the contextual information licensing the grammatical options taken by the speaker. As shown by Connolly (2007, this issue), part of this information is situational in nature.

In Hengeveld and Mackenzie (this issue), situational information – roughly corresponding to Connolly's narrower situational context – is defined as "a language-specific selection of those details of the speech situation that have relevance for Formulation". Its organization is as follows:

(7) Situational information
  Speech situation: participants, utterance time, utterance place
  Physical world: perceivable and inferable entities, such as individuals, events, properties, etc.

The information that is relevant for Formulation and Interpretation in a give language is of course exactly the same.

In Hengeveld and Mackenzie's approach, each relevant element of situational information corresponds to a variable of IL or RL. This gives their classification the advantage of allowing a unified treatment of situational and linguistically expressed information (which is stored in the discoursal context and must be retrieved in

interpreting anaphora, ellipsis, holophrastic answers, etc.): In this way, the GC can be explicitly anchored to the CtxC through co-indexation of contextual and grammatically encoded information. It should be clear, however, that this is only done for practical reasons: Strictly speaking, any bit of situational information is a non-linguistic mental representation of an aspect of the extralinguistic world, and as such different in nature from the linguistic unit used to evoke or designate it. Nonetheless, any information (be it linguistically expressed or not) consists of semantic and/or pragmatic content, and can therefore be represented using the formalism of FDG. In what follows, I will make use of this conventional notation for situational information, treating it just as if it were linguistic in nature.

The unified representation of contextual and linguistic elements is meant to reflect the saturation of unspecified interpersonal and representational variables via the context, as in deixis, anaphora and the identification of speech participants. But this sort of "direct incorporation" of contextual information into the grammatical representation is not the only way in which the CtxtC influences Formulation and Interpretation. For instance, situational information is also essential for Topic and Focus assignment – which largely depends on the Given/New distinction – and both situational and discoursal information can be relevant for the temporal chaining of Episodes and Events (see Hengeveld and Mackenzie, this issue). Such processes are not a matter of simply "copying" one element or feature from the CtxtC into the GC, but of assigning an operator or function to a certain linguistic unit on the basis of the specifications contained in the CtxC: Therefore, the interaction between context and grammar cannot be captured by means of co-indexation. All the same, the development of appropriate grammatical representations by Formulation and Interpretation still requires that the relevant contextual information be immediately accessible to these operations, and this must be reflected in a model oriented to language comprehension.

In some cases, the specification of linguistic variables via the context can be represented quite easily. Consider the following English sentence:

(8) Will you come to the party?

The deictic pronoun *you* and the identifiability operator *the* on *party* indicate that the CtxC already contains a representation of the corresponding entities. As shown in Figure 2, the referent of *you* is stored as situational information in both the socio-cultural context – as a speech participant, (P_J) – and the physical context – as a first-order entity, (x_i). *The party* can be represented as a second-order unit of the (e) type. In this example, this entity is arbitrarily assumed to be present in the discoursal context, but it is easy to imagine a situation in which it is directly perceived, in which case it would be represented as situational information. Finally, the presence of an absolute tense operator (*will*) at the Episode layer requires that a record of the speech time also be represented within the situational information.

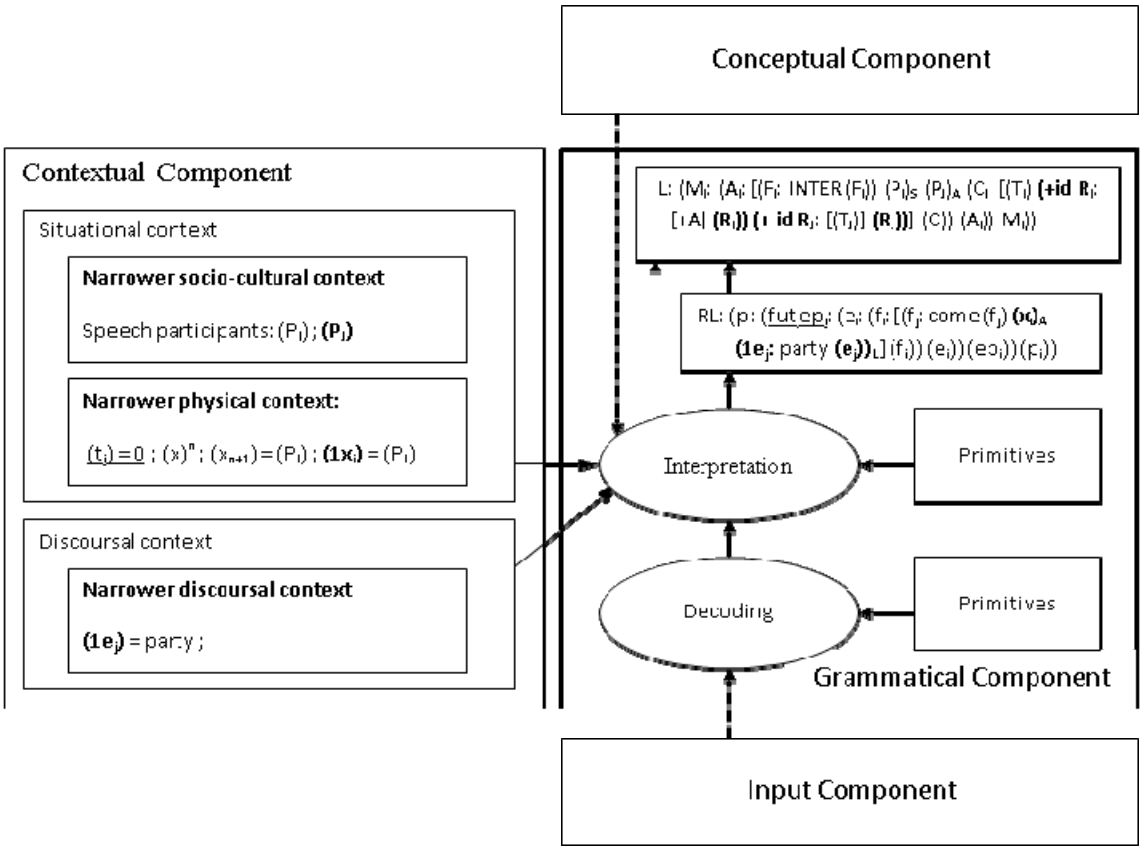

Figure 2: "Will you come to the party?"

The correlations between contextual and grammatical elements are highlighted in bold for indexicals and their referents[7] and through underlining for the contextual record of the speech time and the tense operator it licenses.

Note that in (8) *you* is potentially ambiguous, since it could designate one or more than one individual(s). For this reason, its referent is not represented as (1x) in the GC: Since no number operator is explicitly realized in the pronoun or elsewhere in the utterance, the fact that the referent is a single individual cannot be assumed to be reflected in the semantic representation developed by the grammar of either the speaker or the addressee. By contrast, the CtxC must contain a richer description of the referent intended by the speaker, including the feature with respect to which the pronoun is ambiguous: This is because, in formalizing language parsing, we need to represent the fact that some shared information is recovered in order to show that the indexical is

[7] The correlation between the contextual representation of the participant (P_J) - corresponding to the addressee – and the Referential Subact (R_I) in the GC cannot be established through co-indexation, since Subacts are not semantic units but communicative actions of evocation, and cannot be identified with the evoked entity or property. The correspondance between the two units is made explicit through the representation of (P_J) in the GC, which bears the function Addressee, and through the feature [+A] in the head of (R_I), indicating that *you* refers to (a set of entities including) the Addressee. Similarly, the relation between (R_J) and the entity *party* is established through the representational unit (e_j), co-indexed with that entity.

disambiguated – i.e., it is assigned an adequate contextual specification. The number operator in the contextual representation of the referent is inserted to render this analysis explicit. Note finally that, since the CtxC is language-specific, the inventory of abstract contextual features available for a given linguistic form only needs to be constrained by the grammatical distinctions which are generally relevant for that form at IL and RL. Thus, singular number must be specified in the CtxC in Figure 2 – and would be copied into the RL representation if it were marked by agreement with an NP – while gender or animacy (for example) never need be reflected in the analysis of English second person pronouns.

Suppose now that (8) is addressed to someone whose partner is also invited to the party, but is not taking part in the conversation. If this person has been referred to in the (immediately) preceding discourse, the pronoun will be understood to designate two individuals; conversely, if the addressee's partner is not (or no longer) an active discourse referent, *you* is much more likely to be interpreted as designating the addressee alone. In Figure 2, reference to the addressee's partner is blocked because this person is not taking part in the conversation and is not present in the surrounding environment (is not represented within the situational context), nor has s/he been activated as a referent in the preceding discourse. This kind of ambiguity is solved automatically by the CtxC, unless some contextual information the speaker relied on in using the pronoun is not immediately accessible for the addressee – for instance, if the person in question were an available referent, but the addressee did not know whether the speaker is aware that s/he is also invited to the party.

As pointed out above, if the addressee's partner has been (recently) referred to in the preceding discourse, *you* will most likely be interpreted as designating both individuals. In representing this process, a variable corresponding to the hearer's partner must be inserted in the discoursal context. But this is still not enough to allow co-indexation with the unit corresponding to the pronoun at RL: Our representation must also contain an indication that the two people in question *together* form an available referent. This can be explained along the lines proposed by Keizer (2007a-b, this issue) for kinship relations, i.e. by assuming that the relation between the two partners is semi-active in itself, being "automatically activated as part of the frame or schema" evoked by the addressee. This process is represented in Figure 3 below, where the referent of *you* is contained in a dashed box symbolizing semi-active information.[8] This is designed to suggest that the mental representation of this referent is not, strictly speaking, contained in the CtxC, but is accessible *through* contextual information (as represented by dashed arrows) and originally belongs in a separate storehouse. This storehouse could be the CptC, if we allow for a broader conception of this component than is commonly assumed in FDG, whereby, besides elaborating communicative intentions and hosting what Butler (2008) terms "conceptual proper" and "affective/interactional content", the CptC also handles long-term information such as the knowledge frames associated to linguistic and contextual elements (cf. García Velasco 2007: 183). At any rate, wherever one considers semi-active encyclopaedic information to be stored, if it can be treated as Given at Formulation and retrieved "automatically" at Interpretation

[8] Note that, as in Figure 2, the number feature of the inferred referent (where *m* stands for "plural") is not copied into the grammatical representation. In this specific case, we may also represent this feature as a dual operator rather than a plural, since this grammatical distinction is justified for English by items like both, which can take scope over a personal pronoun.

(with no particular cognitive cost for the addressee), this information can uncontroversially be said to be part of the common ground and must therefore be represented within the CtxC. At the same time, I suggest that *only* the semi-active information which is directly relevant for the production and understanding of the utterance (licensing the use of a given grammatical or lexical item and providing the addressee with the clues to solve any possible ambiguity) can be reasonably assumed to access the CtxC. I will return on this point in 4.1.

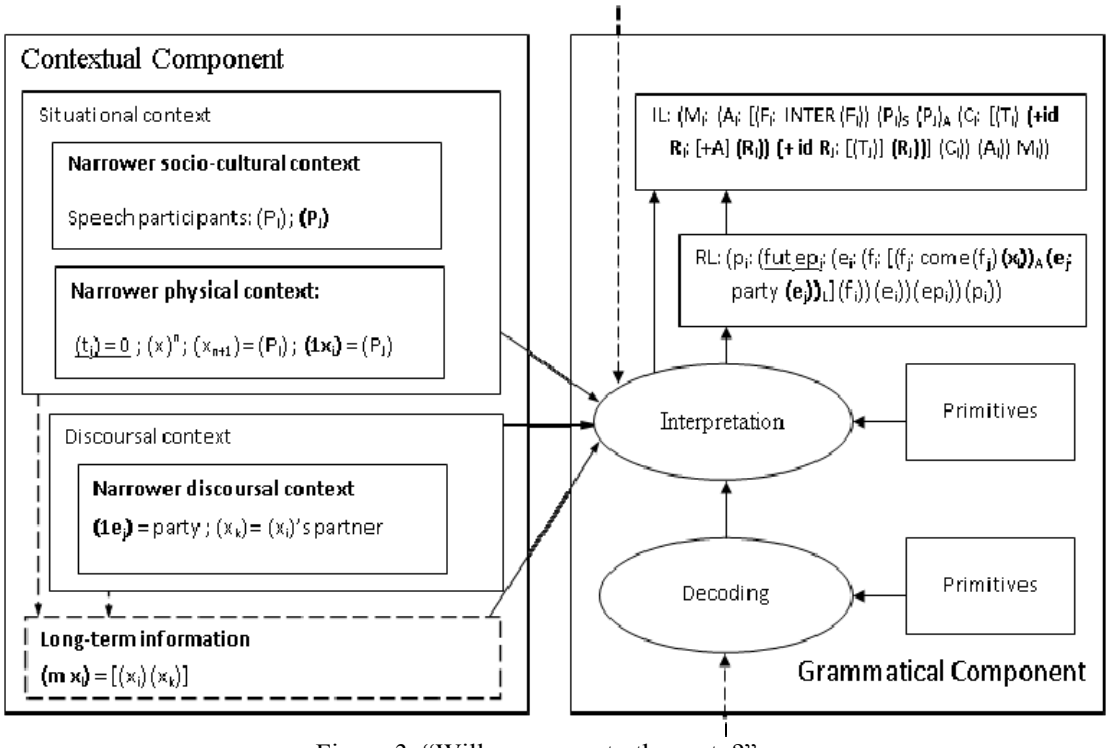

Figure 3: "Will you come to the party?"

An example of interaction between Interpretation and the CtxC involving both IL and RL is found in Italian, and concerns the identical form of the 3rd person singular feminine pronoun and a marker of high social status of the Addressee. This situation is the source for a common joke in Italian popular comedy, typically ending with the resolution of the ambiguity, as in (9b-c):

(9) a. A: *È lei l' accompagnatore?*
COP.PRS.3.SG she/you.POL DEF courier
'Is she the courier?' or 'Are you the courier?'

b. B: *Lei chi?*
'*Lei* who?'

c. A: *Lei tu!* [http://turistipercaso.it/marocco/11823/marocco.html]
'*Lei* you!'

The ambiguity arises because the dialogists have different conceptions of their social relationship with each other, i.e., they have a different access to the narrower

socio-cultural context. In (9b), it becomes clear that dialogist B has misunderstood the value of *lei* in (9a) by interpreting it as a feminine+singular marker at RL, as in (10), rather than as an high status operator on the Addressee at IL, as in (11) (see Hengeveld and Mackenzie 2008: 86).

(10) RL: lei_1: $^{f}(1x_i)$

(11) IL: lei_2: $(h\ P_I)_A$

Differently from English *you*, a single personal pronoun that is simply unspecified for number, lei_1 and lei_2 are two homonymous, functionally different forms: Therefore, the relevant features of the individual (gender and number) or participant (social status) in question must be copied from the CtxC into the GC.

3.2. *From grammar to context*

Let us now turn to another crucial aspect of the interaction between the hearer's grammar and the CtxC and see how the linguistic structures generated in language parsing are set into the CtxC.

According to Connolly, the material output of Articulation is incorporated into the extra-mental narrower discoursal context, which is a physical entity of the real world consisting of sound waves or written signs. At the same time, "[w]hen we take part in discourse activity [...], we build up a memory of the gist of what has gone before, based partly on the preceding utterances, but also partly on inferences that we have drawn in the light of the context" (Connolly 2007: 19). This continuously updated record of the preceding discourse is the narrower mental discoursal context: This is where the abstract representations elaborated by the interlocutors' GCs are incorporated. Phenomena such as morphosyntactic and phonological priming suggest that these representations are incorporated into the mental discoursal context in linguistic form. Moreover, the fact that units from any level can be anaphorically referred to in successive utterances (Hengeveld and Mackenzie 2008: 5) calls for a further categorization of the discoursal context whereby information from each level is incorporated independently. Further evidence for this is the frequency of nominalizations for designating a Propositional Content, Event or Configurational Property which had been previously expressed as a full Clause: If information from the various levels were stored altogether, it would be impossible to encode the retrieved representational unit in a new morphosyntactic template.

The most recent version of Connolly's EMC (Connolly, this issue) integrates Cornish's (2009) distinction between *text* and *discourse*. This opposition is highly relevant to the approach proposed here, as it provides a suitable basis for accommodating abstract linguistic configurations within the CtxC. Cornish (2009: 99-100) defines *discourse* as "… the hierarchically structured, mentally represented product of the sequence of utterance, propositional, illocutionary and indexical acts that the participants are jointly carrying out as the communication unfolds". Discourse is co-constructed by the dialogists "via the text in conjunction with an appropriate context", i.e. it is created on the basis of a "connected sequence of verbal signs and non-verbal signals", the *text*, "by invoking an appropriate context". As I understand it, this means

that the discourse is the participants' mental representation of *the content* of a series of communicative acts, while these acts themselves materially make up the text (Cornish 2009: 104). However, "it is arguable that **all** relevant 'context' is mentally represented, since what is crucial in communication is the users' **perception** as well as **conception** of the external world, rather than the objective 'facts' of the extra-mental universe" (Cornish 2009: 107, emphasis in the original. See also Connolly 2007: 19). That is to say, text too is only relevant as a mental representation of the preceding utterances, with all their formal properties. In my opinion, this is tantamount to saying that the text consists of a cumulative sequence of morphosyntactic and phonological representations, which is integrated into the CtxC and continually kept up to date. Similarly, it is possible to represent the discourse – the mental record of the content of utterances in a text – as a set of interpersonal and representational configurations (Cornish 2009: 99, 100) which are also integrated into the CtxC. For Cornish, the text is stored as short-term information, and "[s]hort-term memory is by definition very limited in storage capacity" (Cornish 2009: 100). Thus, the text will only contain a record of the morphosyntactic and phonological structure of the last utterances produced, from which old structures are dropped at the same time as new structures are integrated.[9] Discourse, by contrast, is stored in long-term memory and can be retrieved very long after the exchange has ended. We gain a clear view of this when we consider expressions like "I can't remember the words exactly, but the idea was… " or "the point was… ": It is exactly this "idea" or "point" that, I suggest, should be conceived of as a set of IL and RL representations (which, as time passes, are reduced to a nucleus of essential information).

It should be added that, as said in section 2, the comprehension of an utterance is of no less importance than its production in updating the context. For instance, a new referent cannot be said to have entered the discoursal context if the corresponding linguistic units of the various levels have have not been decoded and correctly interpreted. Indeed, when we regard the joint process of information transfer as a whole (as is also the case under Connolly's and Cornish's approach), the consideration that no change is brought about in the context until the utterance has been processed by the addressee suggests that text and discourse are only fed by the grammatical representations developed by Decoding and Interpretation, and not (or rather, not *directly*) by those formulated and encoded by the speaker.

Note finally that the discourse, in Cornish's sense, does not merely coincide with the semantic and pragmatic content of the preceding utterances: "[t]he text is but a sequence of 'hints' or instructions to (a) invoke a relevant context [...] and (b) create discourse as a function of it". The discourse thus comprises a significant amount of information that the interlocutors reconstruct through context-based inferences (cf. Connolly 2007: 19). Being non-linguistic in nature, inferred content cannot be incorporated into the discourse as defined here, since this, just like the text, consists of linguistically encoded information only. At the same time, inferred information differs from situational information in not being accessed "directly", through perception or conception, but always through the available discoursal and/or situational information –

[9] Note that, accordingly, priming of formal structures typically occurs in the immediately following turns. Of course, if a certain structure is found particularly important or interesting, it can be stored in long-term memory and reproduced in other situations (i.e., we can quote other speakers' words literally). However, storage of morphosyntactic and phonological configurations in long-term memory is not normally required in dialogue.

as inferences can even be drawn on the basis of both these sources at the same time (cf. Figure 3). Being neither linguistic nor situational, inferred information finds no place within the CtxC, but would be adequately captured in a broadly defined CptC, as proposed in 3.1.2 for long-term knowledge frames. Nevertheless, similarly to (semi-)active long-term information, any newly inferred content can be conventionally formalized in linguistic terms and represented within the CtxC – separately from both the discoursal and the situational information – when computation of such information is necessary for the GC to produce or parse linguistic structures.

4. Toward a general FDG theory of verbal interaction

4.1. *On the Contextual Component of FDG*

From the foregoing, the following conclusions can be drawn:

(i) different elements from both the discoursal and the situational context can interact with the GC of the addressee;
(ii) contextual information is highly relevant to Interpretation, but seems not to influence Decoding at all;
(iii) the text can be defined as the participants' mental record of the morphosyntactic and phonological structure of the utterances in the exchange, and the discourse as the record of the associated pragmatic and semantic configurations;
(iv) situational and encyclopaedic or inferred information, too, can be conventionally represented in linguistic form, when this is necessary for explaining the interaction between context and grammar;
(v) only the grammatical representations generated in language parsing are directly incorporated into the CtxC.

All these elements together now allow us to propose some considerations on the role and structure of the CtxC which should be equally valid for both the standard and the hearer-oriented FDG models.

First, it has been shown that all the contextual categories recognized by Connolly can be relevant to the grammar, so that their inclusion within the CtxC seems fully justified. However, keeping track of all the elements covered by Connolly's taxonomy in the analysis of specific utterances would make the representation extremely difficult to handle, while adding nothing to the accuracy of the analysis; rather, it seems preferable to only represent those contextual elements which directly affect the grammatical operations performed in producing and in understanding the particular utterance at stake. Another important point is that contextual information is only relevant insofar as it is present in the dialogists' mind, i.e., the CtxC must necessarily be conceived of as a purely mental construct.

Within the discoursal context, text and discourse are distinguished. In both cases, the storage and maintenance of information are ruled by what Hengeveld and Mackenzie (this issue) term "the Component's 'clock' ", which "lies at the basis of the notion of stacking and decay": I believe these useful notions should be defined in

accordance with Cornish's suggestion that text and discourse are stored in different types of memory, the latter being more long-lived than the former.[10] The text, as defined here, corresponds to Hengeveld and Mackenzie's Morphosyntactic and Phonological "Strata" of the CtxC, while discourse does not exactly coincide with the Interpersonal and Representational Strata, as it does not contain situational information. Here, as in Connolly, situational information is regarded as a separate contextual category, since it is not derived from the Discourse Acts in the exchange, but is "directly" accessed by the participants through the perception of entities that exist autonomously in the extralinguistic world or through conception of abstract notions.

Finally, I have proposed that long-term and inferred information, originally belonging in the CptC, should be incorporated into the CtxC when it is treated as shared by the speaker. Whereas Connolly and Cornish locate it within the discourse and Hengeveld and Mackenzie only mention inferable situational information (cf. (7) above), here inferred content is represented separately from both discoursal and situational information, since it differs from these both in nature and in the way it is accessed.

The inclusion of inferred information in the CtxC points to a theoretical problem concerning the very nature of this component: Since inferences are by definition subjective, representing them in the CtxC seems to imply that, rather than a single, "public" CtxC, we should distinguish two "private" CtxCs (one for each participant). However, splitting the CtxC into two sub-components does not seem an appropriate solution for both theoretical and operational reasons (cf. Mackenzie, this issue). Doing so would be necessary if one wished to account for the "personal" meaning that utterances can acquire for the two interactants, including any content the speaker does not expect the hearer to infer and, for the speaker, any inference that the hearer does not know the speaker associates with her own utterance. But such non-linguistic and non-shared mental representations – and the information these are drawn upon – have no direct relevance for the grammatical operations performed in producing and understanding linguistic structures, therefore need not be modelled in defining the interaction between context and grammar. Since we are only interested in the structures actually formulated and encoded by the speaker, no separate representation of a non-shared context is necessary: In fact, when the communication is successful, the contextual information relevant for production and comprehension is exactly the same at each stage of the exchange. So, since successful communication is the standard case for linguistic analysis, we only need to represent one, shared CtxC, in both the speaker-based and the hearer-based models. As for cases in which we need to make it explicit that the addressee has no access to certain contextual information the speaker relies on in production – which leads to misunderstanding or non-understanding – this element can still be represented within the CtxC, but will carry an indication of its non-shared

[10] Gurevich, Johnson and Goldberg have shown that "[e]xplicit and implicitly mediated verbatim memory for language exists" (2010: 73) and is not necessarily limited to a handful of seconds after the comprehension of the utterance. However, experiment-based psycholinguistic literature demonstrates that "[w]hen compared directly, it is very clear that meaning always trumps structure" (Gurevich, Johnson and Goldberg 2010: 71). Once this difference in the retention of meaningful and a-functional structures is acknowledged, it is not necessary to represent it explicitly within the model, as it is not directly relevant for grammatical analysis.

status.[11] The problem of the "public" nature of the CtxC is thus solved if we distinguish between shared and non-shared contextual information, rather than between two "private" CtxCs. Note that this is in accordance with what has been said above about the incorporation of discoursal information into the CtxC: If the context is always shared, new linguistic information can only be incorporated once the utterance has been processed by both interactants.

4.2. *Expanding the model*

One objection to the conception of language parsing underlying Figure 1 may rest on the argument that the structures developed by the addressee's GC do not only update the context, but also, undeniably, his state of information and state of mind in general, leading, if necessary, to the development of a novel communicative intention by the CptC. This should certainly be represented in an FDG model of language understanding.

An intuitively appealing way to do so would be to insert an upward arrow from the GC to the CptC. But postulating a direct relation of this kind would mean disregarding the role of the CtxC "as a companion to the Grammatical Component, collaborating with it to achieve contextually appropriate outputs" (Hengeveld and Mackenzie, this issue). This view of context is also perfectly sustainable as regards language comprehension, since it is those "contextually appropriate outputs", and not the underlying representations of the parsed utterance themselves, that actually update the state of information of the addressee, allowing him to reconstruct the original communicative intention of the speaker. In other words, if we accept that linguistic expressions only mean what they mean *in context*, it seems logical to conclude that the representations developed by the addressee's GC only exert their ultimate updating effect on his state of mind by entering (and being appropriately integrated within) the relevant context.[12] It may therefore be concluded that the CptC is not directly fed by the GC, but rather by the CtxC, into which the linguistically encoded information is incorporated.

This proposal has obvious implications for the production-oriented model, since in actual verbal interaction the development of communicative intentions in the speaker's CptC almost always involves computation of contextual information. First, the CtxC influences the very arising of a communicative intention; second, it is consulted in choosing appropriate pragmatic and semantic strategies for pursuing that intention.[13] In this sense, the dynamic implementation of the overall model begins before the GC is activated, first taking place between the CtxC and the CptC – whereby the former feeds into the latter.

---

[11] The infelicitousness of such grammatical realizations can be represented in this way regardless of whether the assumed givenness of the information is reflected at Formulation (e.g., proper names, indexicals, temporal chaining) or Encoding (e.g., weak pronominal forms, ellipsis, nominalizations).

[12] This happens systematically for the meaningful structures in the discourse, while the morphosyntactic and phonological structures in the text can probably only be said to impact the CptC when they trigger some metalinguistic consideration.

[13] Consider for instance the choice between direct and indirect speech acts or the preference for metaphors and periphrases for delicate matters: Such choices are largely determined by contextual factors, but need not be represented, in themselves, within the GC.

It is easy to see how the two processes described above parallel each other, the CptC being fed in both cases by the CtxC. In production, contextual information triggers the CptC, which in turn sends its output to the GC; in comprehension, grammatically encoded information comes to update the CptC through incorporation into the context. Schematically:

(12) **a. Production**
CtxC (discoursal and situational information) > CptC (communicative intention) > GC (Formulation > Encoding) > Output Component (Articulation)

>

b. Comprehension
Input Component (Transmission) > GC (Decoding > Interpretation) > CtxC (discoursal information) > CptC (reconstructing the communicative intention of the speaker > new communicative intention).

(12a) fits perfectly into Connolly's proposal (this issue):

> [...] The situational and discoursal context components feed into the conceptual component and influence the pre-linguistic conceptualisations formed there. The conceptualisations are passed to the grammatical component, where they are formulated and encoded into expressions. These are then passed to the output component.

But Connolly's view of the general "flow of information around the model" differs from the model proposed here as he goes on to argue that

> [t]he results impact the discoursal context component, bringing about an update to the extra-mental discoursal context component; and they may also affect the situational context component, for instance startling someone and thus causing a physical change of state. The new addition to the discoursal context component is then parsed by an addressee-oriented version of the grammatical component. The results of the parse are interpreted in the conceptual component (drawing if necessary on the discoursal and situational context components, which means that the cyclicity is not perfect). The resulting interpretation brings about an update to the mental discoursal context in the discoursal context component. The latter is now ready to feed into the conceptualisation of a new Discourse Act, if necessary.

What Connolly states here contrasts quite sharply with the idea that an utterance cannot be said to have updated the context until it has been processed by the addressee (one cannot be startled by something one has not perceived at all!). At this point, an appropriate contextual specification is created which results from the integration into the discourse of the linguistic representations developed by the addressee's GC; this new, co-constructed discourse fragment eventually feeds into the CptC, leading to "the conceptualisation of a new Discourse Act, if necessary." So, as shown in (12b), in this final phase of the process the CtxC functions as an interface between the Grammatical and the Conceptual Components of the addressee, while in Connolly's account it is always the CptC that mediates between grammar and context.

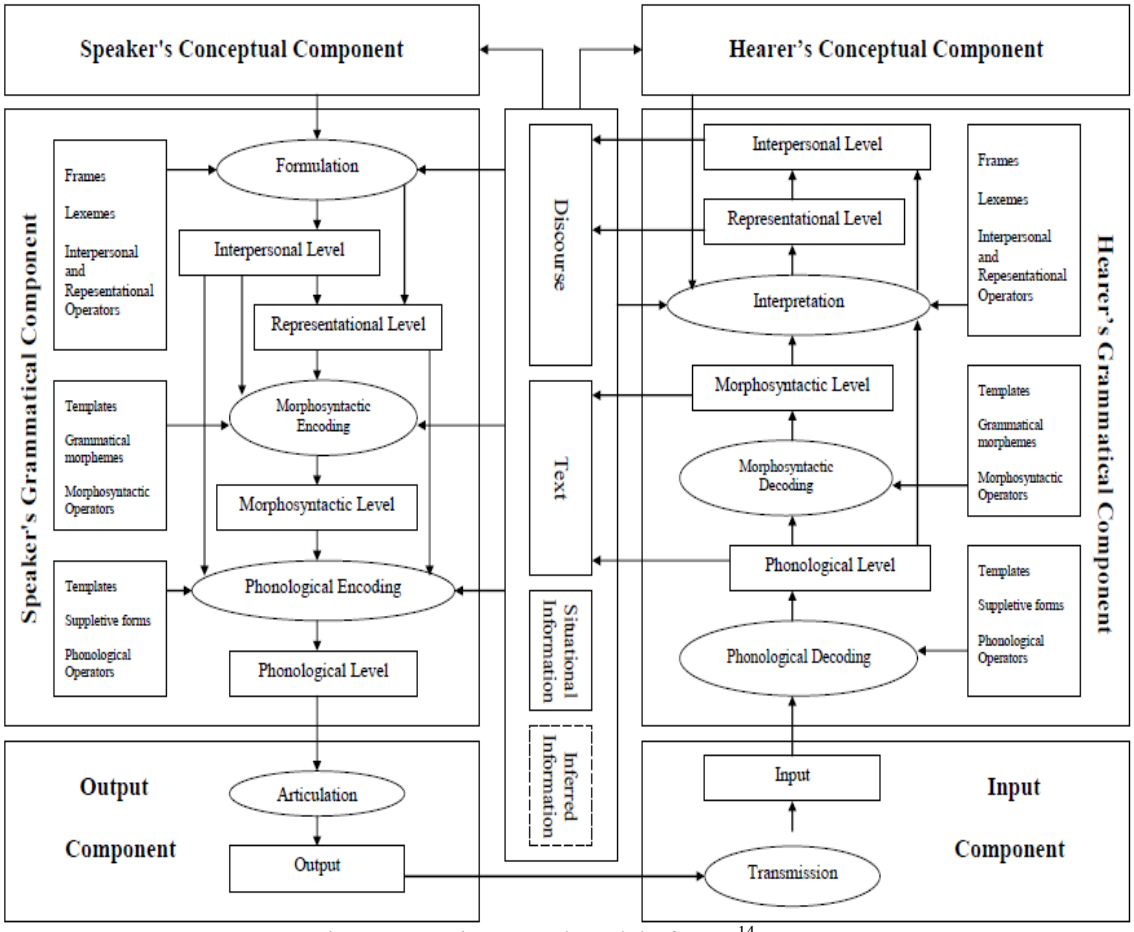

Figure 4: An integrated model of FDG[14]

Figure 4 applies what has been said so far in representing the "life cycle" of an utterance from the conception of a communicative intention by the speaker to the eventual impact of the utterance on the CptC of the hearer. As for the internal organization of the hearer's grammar, we saw in section 2 that both PL and ML can send information off to Interpretation, circumventing the level(s) and operation(s) that lie(s) in between. This is represented by upward arrows from both Decoding Levels to Interpretation. As in language production, each grammatical operation makes use of its own primitives, but only Interpretation receives a separate input from the CtxC (see 3.1.1). The representations generated by Decoding and Interpretation feed into the CtxC, where they are incorporated into the pre-existing textual and discoursal information: As shown above, it is only at this point that the context is actually updated by the utterance, and this is why no input is sent from the grammatical levels in the speaker's GC to the CtxC.[15] Finally, the arrows from the CtxC to the CptC of both

[14] A preliminary version of this figure, where I had mistakenly omitted to represent inferred knowledge within the CtxC, is cited in Cornish (2013).

[15] It should be clear that this modification of the standard model is only justified when one aims to represent the whole process of transfer of information from the speaker to the hearer. When the

interactants stand for the contribution of contextual information to the development of a communicative intention by the speaker and for the eventual updating effect of the utterance, via incorporation into the context, on the information state of the addressee (which may lead to a new communicative intention).

5. Summary

In this article, I have proposed an addressee-oriented model of FDG and tried to integrate it with the standard, speaker-oriented one to achieve a general model of verbal interaction. The overall organization of the addressee's GC is very much the same as in language production, except that the operations it performs take place in the reverse order and the implementation of grammar is partly top-down and partly bottom-up, being fed by both the Input and the Conceptual Components. The former transmits the phonetic input to Phonological Decoding, while the latter provides Interpretation with the means of associating semantic and interpersonal linguistic units with the corresponding non-linguistic mental representation. In the model proposed, the CtxC represents a purely mental construct – since contextual information is only relevant insofar as it is represented within the dialogists' minds – and is always shared and co-constructed by the speech participants. The grammatical structures developed by Decoding and Interpretation feed into the CtxC, being incorporated into the text and the discourse, respectively. As suggested by Hengeveld and Mackenzie (this issue), text and discourse may be further divided into Strata, although discourse, as conceived here, does not include situational information, since this is non-linguistic in nature and must therefore be represented separately. Furthermore, shared or new information evoked in the CptC can access the CtxC by means of an inference based on the discoursal and/or the situational information, becoming available for the interactants' GCs.

The considerations developed in this paper are a first examination of the questions that naturally arise when we look at the interaction between grammar and context from the addressee's perspective through the lens of FDG. Generally speaking, I believe that exploring this perspective can significantly increase the explanatory power of the theory. Moreover, the proposal to include and formalize long-term and inferred information in the CtxC – insofar as this is relevant for the grammatical analysis of actual linguistic structures – may open avenues to a formal treatment of many inferential, usage-based processes which have traditionally been excluded from FDG accounts and which have great relevance for a number of linguistic domains, from psycholinguistics to lexicology, from diachronic linguistics to language acquisition.

---

interaction between grammar and context is described from the speaker's perspective only, as is usual in FDG (whereby the hearer's parsing is "bypassed" in the representation), the outputs from the four grammatical levels to the CtxC must be maintained to keep track of the context-updating effect of linguistic structures.

Abbreviations

3	third person
COP	copula
DEF	definite
M	masculine
POL	polite
SG	singular
PRS	present tense

References

Bornkessel-Schlesewsky, I., and M. Schlesewsky (2009) *Processing syntax and morphology: A neurocognitive perspective*. Oxford: Oxford University Press.

Butler, C. (2008) Cognitive adequacy in structural-functional theories of language. *Language Sciences* 30.1: 1-30.

Chafe, W. (1994) *Discourse, Conciousness and Time*. Chicago: The University of Chicago Press.

Connolly, J.H. (2007) Context in Functional Discourse Grammar. *Alfa: Revista de Lingüística* 51/2: 11-33.

Connolly, J.H. (this issue) The Contextual Component within a dynamic implementation of the FDG model: Structure and interaction. *Pragmatics* 24.2: 229-248.

Cornish, F. (2009) Text and discourse as context: Discourse anaphora and the FDG contextual component. In E. Keizer, and G. Wanders (eds.), *Web Papers in Functional Discourse Grammar* 82: 97-115.

Cornish, F. (2013) On the dual nature of the Functional Discourse Grammar model: Context, the language system/language use distinction, and indexical reference in discourse. *Language Sciences* 38: 83-98.

García Velasco, D. (2007) Lexical competence and Functional Discourse Grammar. *ALFA: Revista de Lingüística* 51.2: 165-187.

García Velasco, D. (this issue) Activation and the relation between context and grammar. *Pragmatics* 24.2: 297-316.

Gurevich, O., M. Johnson, and A.E. Goldberg (2010) Incidental verbatim memory for language. *Language and Cognition* 2.1: 45-78.

Hengeveld, K., and J.L. Mackenzie (2008) *Functional Discourse Grammar*. Oxford: Oxford University Press.

Hengeveld, K., and J.L. Mackenzie (this issue) Grammar and context in Functional Discourse Grammar. *Pragmatics* 24.2: 203-227.

Keizer, M.E. (2007a) Possessive constructions in English: The proposal's supporters or the supporters of the proposal? In M. Hannay, and G. Steen (eds.), *Structural-functional studies in English grammar*. Amsterdam: John Benjamins Publishing Company, pp. 59-82.

Keizer, M.E. (2007b) *The English Noun Phrase: The Nature of Linguistic Classification*. Cambridge: Cambridge University Press.

Keizer, M.E. (this issue) Context and cognition in Functional Discourse Grammar: What, where and why? *Pragmatics* 24.2: 399-423.

Lambrecht, K. (1994) *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.

Mackenzie, J.L. (2012) Cognitive adequacy in a dialogic Functional Discourse Grammar. *Language Sciences* 34: 421-432.

Mackenzie, J.L. (this issue) The Contextual Component in a dialogic FDG. *Pragmatics* 24.2: 249-273.

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

Pickering, M.J., and S. Garrod (2005) Establishing and using routines during dialogue: Implications for psychology and linguistics. In A. Cutler (ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*. Mahwah, NJ: Erlbaum, pp. 85-102.

O'Neill, G. (this issue) Humming, whistling, singing, and yelling in Pirahã. Context and channels of communication in FDG. *Pragmatics* 24.2: 349-375.

Van de Velde, F. (this issue) The discourse motivation of argument realization in nominalisations. *Pragmatics* 24.2: 317-348.

RICCARDO GIOMI, M.A., is a Ph.D. student at the University of Lisbon and researcher at the Lisbon Institute for Theoretical and Computational Linguistics (ILTEC). He has studied linguistics at the universities of Florence and Lisbon, and participated in a research project on the acquisition and development of written language in Portugal. His main research interests include semantics, pragmatics, historical linguistics and the theory of Functional Discourse Grammar.
Address: ILTEC, Avenida Elias Garcia 147, 5º dir., 1050-099, Lisbon, Portugal.
E-mail: giombombo@gmail.com