

Information structure in formalized RRG

Kata Balogh, Laura Kallmeyer and Rainer Osswald
(Heinrich-Heine-Universität Düsseldorf)

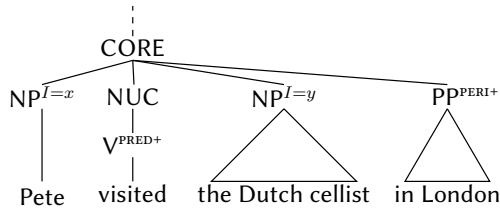
In the architecture of RRG, the linking between syntax and semantics is affected by ‘Discourse Pragmatics’. This component is captured by the ‘Focus Structure Projection’ (Van Valin, 2005). This projection, and the way of looking at information structure in RRG, is based on the theory of Lambrecht (1994). The basic building blocks of the projection are the *information units* (IUs), that are the elements of the two focus domains: the *potential focus domain* (PFD), where the focus can fall, and the *actual focus domain* (AFD). Based on these notions, Van Valin (2005) provides an explanation of various syntactic phenomena in a cross-linguistic perspective. Despite the advantages that RRG’s approach to information structure offers on various syntactic phenomena, it still asks for a further development both for a comprehensive representation of Lambrecht’s theoretical claims and for capturing core phenomena related to focusing, for example: focus sensitivity, focus marking in complex noun phrases or discontinuous focus.

A proper representation of information structure must refer both to syntactic domains, i.e., the focus domains, and to their semantic content. Both are equally important in the grammatical system. The syntactic domains with respect to information structure are crucial in the analysis of information structurally given word order variations and the structural restrictions on the location of *focus* in various languages. Reference to the semantic content of the focus domains is essential for the analysis in terms of communicative functions (Lambrecht, 1994), and for the interpretation of focus sensitive elements, e.g., the particles *only/also/even*, negation, adverbials (e.g., *always*) and so on. The crucial missing aspect in RRG’s approach to information structure is a well-defined reference to the interpretational dimension. This is essential for fully capturing Lambrecht’s insights, but also for generally capturing the contribution of focus both in free focus constructions and in association with focus, regardless whether we consider an essentially semantic (Rooth, 1992; Krifka, 2001) or pragmatic view (Lambrecht, 1994; Roberts, 2012) on focusing.

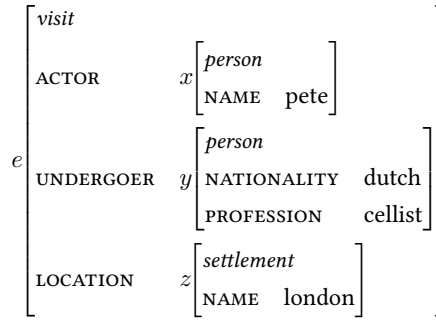
The formal analysis of any linguistic phenomenon requires a two-sided approach: theoretical claims need to be verified by formally exact models, and formal models must be built on solid theoretical grounds. In order to facilitate such an approach, our proposal towards a formalization and extension of RRG’s Focus Structure Projection is based on the formalized version of RRG (fRRG; Kallmeyer et al., 2013; Kallmeyer and Osswald, 2017; Osswald and Kallmeyer, 2018; Kallmeyer and Osswald, 2023). This formalism is based on the solid theoretical grounds of RRG, while it provides an exact, formal specification of the grammar in terms of *Tree-Wrapping Grammar* (Kallmeyer et al., 2013), strongly inspired and based on LTAG (Joshi and Schabes, 1997). The current development of fRRG provides a specification of the syntax-semantics interface, but lacks a modeling of information structure, which asks for an extension.

The basic components of the Focus Structure Projection, that make up the different focus domains, are the *information units*. Following Lambrecht (1994), Van Valin (2005: 77) argues that the minimal focus domains (i.e., the IUs) within the constituent structures are the NUCLEUS, the core arguments and the periphery PPs. These units must refer to their semantic contribution as well. We take IUs formally as pairs of a syntactic domain and its semantic representation.

(1) *Pete visited the Dutch cellist in London.*



Semantic representation:



the set of information units in (1): $\{\langle \text{NP}, f_x \rangle, \langle \text{NUC}, f_e \rangle, \langle \text{NP}, f_y \rangle, \langle \text{PP}, f_z \rangle\}$, that is

$$\left\{ \left\langle \text{NP}, \begin{matrix} \text{person} \\ \text{NAME} \text{ pete} \end{matrix} \right\rangle, \left\langle \text{NUC}, \begin{matrix} \text{visit} \\ \text{ACT} \text{ [1]} \\ \text{UNDG} \text{ [2]} \end{matrix} \right\rangle, \left\langle \text{NP}, \begin{matrix} \text{person} \\ \text{NAT} \text{ dutch} \\ \text{PROF} \text{ cellist} \end{matrix} \right\rangle, \left\langle \text{PP}, \begin{matrix} \text{[3]} \text{ LOC} \\ \text{z} \begin{matrix} \text{settlement} \\ \text{NAME} \text{ london} \end{matrix} \end{matrix} \right\rangle \right\}$$

The basic outline of the proposal is the following. The focus domains are taken as sets of information units, which essentially represent the focus structure of the given sentence. The PFD is determined as a language specific feature, while the AFD is determined by the focus marking strategies and the underlying discourse context together. Given the set of IUs and the AFD, we can straightforwardly calculate the non-focus domain (NFD) as well, which plays a role at the interpretation. The semantic content of the focus domains are derived by the unification of the semantic content of their elements, i.e., the IUs within the set. These contents correspond to the focus-background division in other approaches (see, e.g., Krifka, 1992, 2001). The sentence in (1) can have various focus structures. In English, it depends on the prosodic structure, i.e., accent placement, together with the local discourse context. The PFD is the same in all cases, while the AFD and the NFD are different in all utterances. These domains are both subsets of the set of information units (i.e., $\text{AFD}, \text{NFD} \subseteq \text{IUS}$), such that the AFD and the NFD are distinct sets ($\text{AFD} \cup \text{NFD} = \emptyset$) and their union equals the IUS ($\text{AFD} \cap \text{NFD} = \text{IUS}$).

(2) a. *Who visited Kate in London?* [PETE]^F visited Kate in London.

$$\Rightarrow \text{AFD} = \{\langle \text{NP}, f_x \rangle\}, \text{NFD} = \{\langle \text{NUC}, f_e \rangle, \langle \text{NP}, f_y \rangle, \langle \text{PP}, f_z \rangle\}$$

b. *What did Pete do in London?* Pete [visited Kate in LONDON]^F

$$\Rightarrow \text{AFD} = \{\langle \text{NUC}, f_e \rangle, \langle \text{NP}, f_y \rangle, \langle \text{PP}, f_z \rangle\}, \text{NFD} = \{\langle \text{NP}, f_x \rangle\}$$

The above example already illustrates an important advantage. The architecture of RRG and our representation of information structure is considerably different from ‘traditional’ accounts based on syntactic F-marking. The essence is that focus domains are sets of information units. Crucially, the IUs are linked to syntactic domains, but the focus domains are not determined on the nodes of the constituent structure. Therefore, when the basic IUs are defined, any combination of them can make up the actual focus domain. This predicts that neither ‘non-constituent’ focus (3a) nor discontinuous focus (3b) are problematic, that are both challenging in ‘traditional, F-marking approaches (see, e.g., Krifka, 1992; Buring, 2016).

(3) a. (*What about Kate? / What happened to Kate?*) [Pete visited]^{AFD} Kate

$$\Rightarrow \text{AFD} = \{\langle \text{NP}, f_x \rangle, \langle \text{NUC} : f_e \rangle\}, \text{NFD} = \{\langle \text{NP}, f_y \rangle\}$$

b. (*What happened between Pete and Kate?*) Pete introduced Kate to the cellist.

$$\Rightarrow \text{AFD} = \{\text{NUC} : f_e, \text{NP} : f_z\}, \text{NFD} = \{\text{NP} : f_x, \text{NP} : f_y\}$$

The above initial proposal correctly captures the basic intuitions and core insights of information structure, nevertheless it still raises several issues, both on the formal and on the theoretical

sides. The two basic questions to resolve are: (1) At which point in the derivation and how are the information units determined? and (2) What determines the sets of the focus domains, i.e., the actual focus structure of the sentence? In our talk, we address these core questions, that are crucial both for the formalization and for general theoretical considerations. For the first question, we propose that IUs are structurally determined and identified during the derivation of the sentence, where the derivation tree plays a crucial role. As for the second question, we argue that the focus structure is determined by language specific focus marking strategies and the local discourse structure together. The former aspect should be captured via the relation of syntax, semantics, prosody and morphology, while that latter aspect requires a representation of the context and the embedding of the sentence-level representation into a dynamic framework.

References

- Daniel Büring. Discontinuous foci and unalternative semantics. *Linguistica*, 56(1):67–82, 2016.
- Aravind K. Joshi and Yves Schabes. Tree-Adjoining Grammars. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, pages 69–123. Springer, Berlin, 1997.
- Laura Kallmeyer and Rainer Osswald. Combining predicate-argument structure and operator projection: Clause structure in Role and Reference Grammar. In Marco Kuhlmann and Tatjana Scheffler, editors, *Proceedings of the 13th International Workshop on Tree Adjoining Grammars and Related Formalisms*, pages 61–70. ACL Anthology, 2017. URL <https://aclanthology.org/W17-6200/>.
- Laura Kallmeyer and Rainer Osswald. Formalization of RRG syntax. In Delia Bentley, Ricardo Mairal, Wataru Nakamura, and Robert D. Van Valin, Jr., editors, *The Cambridge Handbook of Role and Reference Grammar*. Cambridge University Press, Cambridge, 2023.
- Laura Kallmeyer, Rainer Osswald, and Robert D. Van Valin, Jr. Tree wrapping for Role and Reference Grammar. In Glynn Morrill and Mark-Jan Nederhof, editors, *Proceedings of Formal Grammar 2012 and 2013*, pages 175–190, Berlin & Heidelberg, 2013. Springer.
- Manfred Krifka. A compositional semantics for multiple focus constructions. In Joachim Jacobs, editor, *Informationsstruktur und Grammatik*, Sonderheft der Linguistische Berichte 4, pages 17–53. Springer Fachmedien, Wiesbaden, 1992.
- Manfred Krifka. For a structured meaning account of questions and answers. In C. Féry and W. Sternefeld, editors, *Audiatur Vox Sapientia: A Festschrift for Arnim von Stechow*, pages 287–319. Akademie Verlag, Berlin, 2001.
- Knud Lambrecht. *Information structure and sentence form*. Cambridge University Press, Cambridge, 1994.
- Rainer Osswald and Laura Kallmeyer. Towards a formalization of Role and Reference Grammar. In Rolf Kailuweit, Lisann Künkel, and Eva Staudinger, editors, *Applying and Expanding Role and Reference Grammar*, (NIHIN Studies), pages 355–378. Albert-Ludwigs-Universität, Universitätsbibliothek, Freiburg, 2018.
- Craige Roberts. Information Structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6), 2012.
- Mats Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1, 1992.
- Robert D. Van Valin, Jr. *Exploring the syntax-semantics interface*. Cambridge University Press, Cambridge, 2005.