Right Node Raising in Parsing and Generation

Christian Rohrer Jonas Kuhn Sina Zarrieß
University of Stuttgart, IMS, Germany

We present an implementation of Right Node Raising (RNR) in German in the framework of LFG. Like other types of non-constituent coordination, RNR is often considered notoriously difficult and ignored in grammar writing – although instances of it do occur in real text. Our analysis is inspired by the "rule splitting" technique of (Maxwell and Manning, 1996), but it does not leave the formal framework of standard LFG and at the same time requires few special-purpose rules, as it builds on the regular mechanisms implementing German clausal syntax.

Informally speaking, the term 'Right Node Raising'[1] refers to two coordinated clauses which lack some elements in the first conjunct. The missing parts in the first conjunct have obligatory phonetically overt counterparts in the same structural position in the second conjunct (Féry and Hartmann, 2005). Our implementation of RNR accounts for the following raised constituents: DPs, PPs and ADVPs (example 1), infinitival constructions (example 2), V-final clauses (example 3).

(1) Hans kauft und Maria verkauft Aktien.
Hans buys and Maria sells shares.

(2) Hans versuchte gestern und Emil probiert heute, das
Hans tried yesterday and Emil attempts today, the
Problem zu lösen.
problem to solve.

(3) Hans bedauert und Emil begrüßt, daß Maria geht.
Hans regrets and Emil welcomes, that Maria leaves.

The raised constituent may be "extracted" from different levels of embedding (example 4) and several constituents may be raised (example 5).

(4) Wulff sagte heute und Merkel soll morgen bestätigen,
Wulff said today and Merkel should tomorrow confirm,
daß die Steuern nicht gesenkt werden.
that the taxes not lowered be.

(5) Hans versprach seiner und Eva versprach ihrer Mutter, zu
Hans promised his and Eva promised her mother, to
kommen.
come.

In our analysis, the raised material is adjoined to the coordination of two sentences.

```
(a) ROOT --> CProot: ! $ ^;  and
             CProot: ! $ ^;
             RNR.
(b) RNR --> {DP*|PP*|VP|...}.
```

RNR specifies the possible raised constituents. The raised constituent must be optional in the rule which introduces it. The coordination rule is the familiar LFG-rule for same-constituent coordination. The functional information of the raised contituent(s) is distributed over the set of elements. For distribution to be possible, RNR has to be adjoined to CProot and cannot be part of CProot. The raised constituents are annotated roughly with the function they have in the position from which they are 'extracted'; with some exceptions: For instance, subjects cannot be RNR-ed. The annotation of the RNR-ed constituent follows the same principles as the annotation of topicalized constituents like the relative pronoun in example (9).

(6) Ich kenne den Roman, den Maria schreibt und Erich
I know the novel, which Maria writes and Erich
liest.
reads.

Problems with overgeneration: The implementation sketched so far overgenerates. In generation-mode we get, for instance, the following surface realisation for example (3).

(7) Gestern versuchte Hans und Emil probiert heute, das
Yesterday tried Hans and Emil attempts today, the
Problem zu lösen.
problem to solve.

In this surface realization, the syntactic parallelism of the original sentence (3) is destroyed. The analysis presented so far does not capture two basic features of RNR (Féry and Hartmann, 2005): (a) the raised material is extracted from the right periphery of the first conjunct, (b) the two conjuncts must exhibit a parallel syntactic and focus structure. In order to account for conditions (a) and (b) we introduce a new non-terminal category 'RightPeriphery' into the Middle Field (MF) and use linear precedence rules to enforce its final position in the MF. Furthermore we have to make sure that the constituents in the right periphery have the same function in both conjuncts.[2] Therefore, we introduce a discourse function RNR-FOCUS.

[1] We use the terms 'Right Node Raising' and 'raised constituent' despite the fact that in LFG there is no "extraction" rule.

[2] The grammar also parses sentences like (i) where the DP is 'extracted' from inside a VP.

(i) Der Mitarbeiter wird [DP ...] verfassen und der Chef wird
The colleague will [DP ...] write and the boss will
[DP ...] unterschreiben den Bericht.
[DP ...] sign the report.

```
RightPeriph -->
          e:(^RNR-FOCUS)= ! ;
          { ADVP[std]:! $ (^ADJUNCT)
          | DP[std]: ! = (^ {SUBJ|OBJ|OBJ-TH})
          |...}.
```

We use the term RNR-FOCUS because the constituent dominated by 'RightPeriphery' is Focus-marked. Phonetically speaking, it receives a pitch accent. To capture the notion of functional parallelism, we add a schematic constraint to rule (a) such that the RNR-FOCUS originates from the same grammatical function in both conjuncts. [3]

Syntactic Parallelism: One could argue that sentence (7) is not ungrammatical. We believe that the best way to capture parallelism is by a soft constraint. XLE's log-linear disambiguation component already provides a c-structure feature template 'cs_conj_nonpar' which counts non-parallel conjuncts within depth levels. A similar template could be added for f-structure parallelism.

(Féry and Hartmann, 2005) observed that RNR-construction exhibit a parallel focus structure. If one makes this focus structure accessible to syntax along the lines proposed by (Bögel et al., 2009), then the Principle of Prosodic Preference would prefer RNR-constructions where syntactic constituent boundaries coincide with prosodic boundaries.

Discussion: Maxwell and Manning (1996) propose a treatment of non-constituent coordination requiring a modification of c-structure parsing: coordination may pertain to an incomplete subspan of a c-structure rule's right-hand side. In principle, the LFG parsing algorithm can be modified with a stack-driven mechanism that will allow for coordination of partial constitutents, otherwise following standard LFG assumptions about coordination. The stack is used to keep track of the position in the rule's right-hand side up to which c-structure material was included/excluded in the coordination. To our knowledge, the mechanism has never been implemented, apart from (Zarrieß and Seeker, 2008), who propose a finite-state based rule compilation mechanism that can be combined with XLE.

In spirit however, Maxwell and Manning's analysis can be captured well with moderate extensions applied to a standard broad-coverage grammar of German. This is because for the typical cases of non-constituent coordination, detailed c-structure-level book-keeping over the position up to which material has been included/excluded in the coordination is not required: To capture argument and adjunct placement in the German *Mittelfeld*, no hard grammatical constraints are assumed. Simplifying somewhat, a binary right-branching rule, or, as Forst and Rohrer (2009) propose to capture coordination facts, a flat rule VP → XP* VC can be assumed (where VC is the verbal complex). In addition, Forst and Rohrer (2009) assume an "artificial" category VPargs (on the left edge of VC), which can span two or more of the XP arguments/adjuncts, excluding the verbal complex. By allowing for coordination of this VPargs category, typical conjunction reduction cases (like *John gave Mary apples and Sue bananas*) can be captured. The unconstrained span of XP* in the VP on one hand, and VPargs on the other generates the necessary options for the conjunction reduction phenomenon and is at the same time constrained at the level of f-structure. RNR is not captured by this analysis since it involves argument/adjunct material outside the coordination. But in the present paper, we propose the addition of a dual "artificial" category for this non-coordinated material, the RNR category, which again nicely combines with a flexible-span coordination.[4]

In the RNR analysis, additional f-structural constraints are needed to capture the parallelism constraint on the conjuncts. Note that this would be required in Maxwell and Manning's formalism too, and as we pointed out above, there is clear information structural evidence motivating this.

[3] It is parallelism of f-structure not c-structure parallelism which is required here. Consider example (i) where 'heute' and 'am Montag' belong to different syntactic categories but have the same funtion:

(i) Er kaufte heute und sie kaufte am Montag ein Auto.
He bought today and she bought on Monday a car.

[4] We note that the RNR category can be viewed as a generalized *Nachfeld* category, attached at a high structural level, such that it can distribute over a complete coordinated sentence. Normally, one would not allow for arbitrary functions for a *Nachfeld* category (since extraposed arguments are extremely rare and an extreme amount of local ambiguity is added), but in its interaction with coordination, the construction is sufficiently constrained.

References

Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy Holloway King, and John T. Maxwell III. 2009. Prosodic Phonology in LFG: A New Proposal. In *Proceedings of the LFG09 Conference*.

Forst, Martin, and Christian Rohrer. 2009. Problems of German VP Coordination. In *To appear in the Proceedings of the LFG '09 Conference*, Cambridge, England.

Féry, Caroline, and Katharina Hartmann. 2005. Focus and prosodic structures of German Gapping and Right Node Raising. *The Linguistic Review* 22:67–114.

Maxwell, John, and Christopher Manning. 1996. A Theory of Non-constituent Coordination based on Finite State Rules. In *Proceedings of the First LFG Conference*, Grenoble, France.

Zarrieß, Sina, and Wolfgang Seeker. 2008. Finite-State Rule Deduction for Parsing Non-Constituent Coordination. In *FSMNLP 6th International Workshop,*, Potsdam.