Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

The papers are published under a CC-BY license:
http://creativecommons.org/licenses/by/4.0/

Contents

Editor's note **3**

Gabrielle Aguila-Multner, Berthold Crysmann: The morphology of trapping and climbing in French causatives **4**

Abeer Alsulami, Robert D. Borsley: On exhaustive conditional clauses in Modern Standard Arabic **25**

João Azevedo, Jakob Maché: Between shared expectations and possibilites: The discourse particle *sempre* in European Portuguese **47**

Francis Bond, Joanna Ut-Seong Sio: A Construction-based Approach to Cantonese Classifiers **60**

Carly Crowther, Emily Luedke, Emily M. Bender: The morphological causative in Panãra: A Grammar Matrix implementation **76**

Chenyuan Deng: An HPSG account for German numeral classifiers **88**

Petter Haugereid: Examining Delayed Complements in Norwegian within an Incremental Left-Branching Grammar Framework **104**

János Litzinger: Hybrid Agreement in Bosnian/Croatian/Serbian **118**

Jieun Oh: A new account of the case alternation in Korean complex predicates **132**

Manfred Sailer: Horn clauses and strict NPIs under negated matrix clauses **152**

Berke Şenşekerci: Gradient HPSG **173**

Giuseppe Varaschin, Antonio Machicao y Priemer, Yanru Lu: Topic drop in German: Grammar and usage **193**

Editor's note

The 31st International Conference on Head-Driven Phrase Structure Grammar (8 July–10 July 2024) took place as a hybrid conference at the Palacký University Olomouc, Czech Republic and was organized by Francis Bond and Joanna Sio.

The conference featured one invited talk and 13 papers selected by the program committee (Anne Abeillé (Université Paris Cité), Felix Bildhauer (Institut für Deutsche Sprache), Olivier Bonami (Université Paris Cité), Francis Bond (Palacký University, Olomouc), Rui Chaves (University at Buffalo, chair), Berthold Crysmann (CNRS - LLF & Université Paris Cité), Anthony R. Davis (Southern Oregon University), Anke Holler (University of Göttingen), Jong-Bok Kim (Kyung Hee University), Jean-Pierre Koenig (University at Buffalo, SUNY), Yusuke Kubota (National Institute for Japanese Language and Linguistics), Andy Lücking (Goethe University Frankfurt a. M.), Jakob Maché (University of Lisbon), Nurit Melnik (The Open University of Israel), Stefan Müller (Humboldt Universität zu Berlin), Petya Osenova (Sofia University and IICT-BAS), Rainer Osswald (Heinrich-Heine-Universität Düsseldorf), Gerald Penn (University of Toronto), Frank Richter (Goethe University Frankfurt a. M.), Manfred Sailer (Goethe University Frankfurt a. M.), David Yoshikazu Oshima (Nagoya University), Olga Zamaraeva (University of A Coruña)). There was a workshop on Formal Approaches to Under-resourced Languages (FAUL) with one invited speaker and three regular papers.

We want to thank the program committee for putting this nice program together.

As of 2024, the procedure to publish the proceedings has been changed. A conference submission consists of a five page abstract, which is reviewed by at least two reviewers. After the conference a full version of the paper is submitted to the program committee chair for a further round of reviewing. Papers that are accepted are then published in the proceedings.

To ensure easy access and fast publication we have chosen an electronic format. The proceedings are published by the University Library of Goethe-Universität, Frankfurt am Main.

The morphology of trapping and climbing in French causatives

Gabrielle Aguila-Multner
Romanisches Seminar, Universität Zürich

Berthold Crysmann
Université Paris Cité, Laboratoire de linguistique formelle, CNRS

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 4–24

Keywords: Periphrasis, Information-based Morphology, French, Clitics, Causative

Aguila-Multner, Gabrielle & Berthold Crysmann. 2024. The morphology of trapping and climbing in French causatives. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 4–24. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.1. CC-BY 4.0

Abstract

In this paper[†], we shall discuss how the treatment of morphological periphrasis (Vincent & Börjars, 1996; Popova & Spencer, 2013; Bonami, 2015) can be integrated into the framework of Information-based Morphology (IbM; Crysmann & Bonami, 2015), an inferential-realisational theory of inflection couched in terms of typed feature structures.

French verb morphology provides a rich set of synthetic and analytic forms. Among the latter we find the so-called composed tenses (e.g. *passé composé*) and the near tenses, such as the *passé récent* (='recent past') and the *futur proche* (='near future'). Recently, we have argued that clitic climbing can equally be understood as morphological periphrasis (Aguila-Multner & Crysmann, 2020a,b; Aguila-Multner, 2023). Thus, the morphosyntax of French verbs provides an ideal testing ground to study the interaction of different periphrastic dependencies, which we shall use to illustrate our treatment of periphrasis in IbM.

1 Background: clitic climbing as periphrasis

The division of labour between morphology and syntax in Romance clitic climbing has been the subject of some debate. Based on Zwicky & Pullum (1983)'s lexicalist view of the clitic–affix distinction, the 'clitics' involved have been shown to be lexical affixes (Miller, 1992) of the host verb. Their ability to 'climb' to an auxiliary, as in (1), therefore requires an explanation.

(1) La professeure le lui a écrit.
the professor DO.3SG.M IO.3SG has written
'The professor wrote it to her.'

Miller & Sag (1997) and Abeillé & Godard (1996, 2002) have proposed that such cases involve syntactic raising of the arguments the affixes correspond to: under this view, clitic climbing verbs are argument composition auxiliaries which inherit their complement's argument structure, providing a syntactic explanation for clitic climbing. In Aguila-Multner & Crysmann (2020a), on the other hand, we have proposed that clitic climbing is a consequence of inflectional periphrasis: under this view, clitic climbing verbs are periphrastic auxiliaries able to host part of the lexical verb's morphology, including morphological exponence of its pronominal arguments. Figure 1 illustrates the core mechanisms of the periphrasis approach: essentially, inflectional morphology introduces morphosyntactic requirements (including the auxiliary's *l(exemic)id(entity)*, and, possibly, pronominal affixes – *praf*) on a feature REV(ERSE)-SEL(ECTION), which is percolated in syntax and terminated

[†]We would like to thank the audience at HPSG 2024 for their comments and discussion, in particular Antonio Machicao y Priemer and Emily Bender, as well as the anonymous reviewers for the conference and the proceedings. The research reported here has partially benefited from a public grant overseen by the French National Research Agency (ANR) as part of the program "Investissements d'Avenir" (reference: ANR-10-LABX-0083). It contributes to the IdEx Université Paris Cité (reference: ANR-18-IDEX-0001).

by an ancillary element, such as the tense auxiliary *avoir*. The auxiliary matches the constraints under REV-SEL against its own inflectional properties (its MS set), inheriting any morphological arguments of the participle. Clitic climbing is thus treated as a morphological dependency between the lexical verb and the auxiliary in a periphrastic relation.

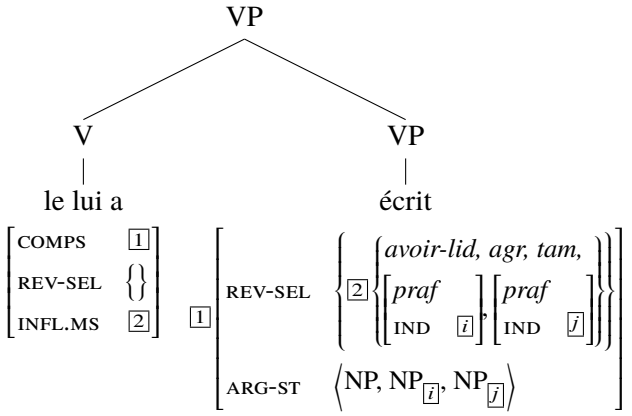


Figure 1: Periphrasis by reverse selection (adapted from Aguila-Multner & Crysmann, 2020a)

Causative constructions provide an interesting testing ground for these competing approaches. Not only are these constructions clitic climbing contexts – clitics may climb to causative *faire* 'make' (2) – but they are also clause union constructions (Aissen & Perlmutter, 1976), displaying various idiosyncratic syntactic properties.

(2) La professeure le lui fait écrire.
the professor DO.3SG.M IO.3SG makes write.
'The professor makes her write it.'

We have shown, however, that the morphological approach based on periphrasis simplifies over the argument composition approach to clause union in a number of ways (Aguila-Multner & Crysmann, 2020b) . In particular, the periphrastic approach places the locus of clause union at the bottom of the construction at the level of the lexical infinitive: this means that decisions regarding argument realisation are made at the most relevant level, where the previous approach required percolation of lexical information in phrase structure to access it from the argument composition auxiliary. This results in a simpler treatment of subject marking, *by*-phrases, and clitic trapping, which depend on argument structure, on lexical semantics, and on the intrinsic status of morphological arguments respectively. The approach further simplifies the associated phrase structure, maintaining a traditional hierarchical structure rather than the flat structure projected by the argument composition auxiliary.

In the next section we discuss the interaction of pronominal affixation, causativisation, reflexivisation and medio-passivisation understood as different processes of grammatical function change interacting with one another.

2 French causatives: processes of grammatical function change interacting

French causative constructions display an interesting interplay of processes of grammatical function change. Firstly, the construction can be conceived of as a periphrastic causative, namely as a two-word form expressing addition of a causer argument (Aguila-Multner, 2023). Thus the lexical infinitive in (3) below is a causativised form, its subject (*Louise*) being demoted to an indirect object (*à*-NP) to make room for the new subject (*la professeure*).

(3) La professeure fait écrire un poème à Louise.
the professor makes write a poem to Louise
'The professor makes Louise write a poem.'

Secondly, as noted in the previous section, causativisation of an infinitive can combine with morphological realisation of arguments. In the standard case, this results in clitic climbing: the morphological arguments of the lexical verb are realised on *faire*, as in (4).

(4) La professeure le lui fait écrire.
the professor DO.3SG.M IO.3SG makes write
'The professor makes her write it.'

As we discussed in Aguila-Multner & Crysmann (2020b), this can be understood as periphrastic realisation of a morphological property. However, a different pattern can also surface in some cases,[1] namely downstairs affixation:

(5) Cela fait lui en vouloir à Louise.
this makes IO.3SG GEN be_mad to Louise
'This makes Louise be mad at her.'

Furthermore, a certain class of French pronominal affixes (which we can call *reflexives* in a broad sense) can mark more complex grammatical function change. This includes not only true reflexives or reciprocals, as in (6), but most interestingly medio-passives, in which *se* marks the promotion of a direct object to subject, the erstwhile subject being generically interpreted, as in (7).

(6) Louise et Matthieu se rencontrent.
Louise and Matthieu REFL.3 meet
'Louise and Matthieu meet each other.'

[1] See Aguila-Multner (2023) for discussion of so-called trapping and of the factors that license such downstairs realisation in French causative constructions.

(7) Ce problème se remarque facilement.
this problem REFL.3 notices easily
'This problem is noticed easily.'

Again, the combination of these processes with causativisation can result in two different patterns: upstairs realisation as in (8), or downstairs realisation, as in (9). This time, a meaning alternation obtains from the difference: as shown by the translation, upstairs realisation corresponds to a medio-passivised causative, while downstairs realisation expresses a causativised medio-passive.

(8) Ce problème peut se faire remarquer à des néophytes.
this problem can REFL.3 make notice to INDEF.PL neophytes
'This problem can be shown to neophytes.'

(9) Son ampleur fait se remarquer facilement le problème.
its scale makes REFL.3 notice easily the problem
'Its scale makes the problem be noticed easily.'

As proposed by Aguila-Multner (2023),[2] the dataset in (3)–(9) can be treated as a permutation in the order of three processes of grammatical function change: mapping of arguments, causativisation, and medio-passivisation. The iconic ordering of the latter two is evidenced by the contrast between medio-passivised causatives and causativised medio-passives, while the contrast between upstairs and downstairs realisation of pronominal arguments can be explained by ordering between the first two. Aguila-Multner proposes to treat this permutation in terms of freely ordered lexical rules:[3] Figure 2 illustrates the implementation of the causativisation process as a lexical rule, augmenting the argument structure of a verb with a causer argument; Figure 3 illustrates the reflexivisation process as a lexical rule, coindexing one NP element of the argument structure to the first element; Figure 4 illustrates a medio-passivisation lexical rule, promoting a direct object to subject.

However, the details of the morphological interface necessary for this line of analysis have not been developed until now. Crucially, to account for the possibility of downstairs realisation, a non-trivial assumption has to be made: that the causativisation process can apply to a fully inflected word form. In order to license examples as in (10), the description to which the causativisation lexical rule applies needs to be that of the affixed word *vous en vouloir* – failing that, the morphology will receive as its input a combination of causativity and pronominal affixes, for

[2] See also Villalba (1994) for an earlier treatment within generative theory.

[3] We use description-level lexical rules, as proposed e.g. in Koenig (1999), namely feature structures of type *word* or *lexeme* whose daughter (DTR) is again lexical sign, i.e. a feature structure of type *word* or *lexeme*.

[4] The attached relational constraint bars *expl(etive)* or *ana(phoric)* affixal synsems (*aff-ss*) from being present on the daughter's COMPS list, in order to capture trapping of reflexives and inherent clitics, cf. (5). This constraint is satisfied in one of two ways: first, if there is no such argument on ARG-ST, there will trivially not be one on COMPS either. Second, if there is such an argument on ARG-ST, application of argument mapping will ensure it is suppressed on COMPS. Since argument mapping applies to signs of type *word*, this entails downstairs morphological expression.

$$\left[\begin{array}{ll}
\textsc{vform} & \textit{nonfinite} \\
\textsc{subj} & \left\langle \mathrm{NP}_{\boxed{i}} \right\rangle \\
\textsc{comps} & \left\langle \mathrm{NP}_{\boxed{j}} \left[\textit{à/bare/par/de} \right] \right\rangle \oplus \boxed{c} \\
\textsc{infl} & \left[\textsc{ms} \quad \boxed{m} \cup \left\{ \textit{cause-lid}, \boxed{i}\ \textit{agr} \right\} \right] \\
\textsc{dtr} & \left[\begin{array}{ll} \textsc{subj} & \left\langle \mathrm{NP}_{\boxed{j}} \right\rangle \\ \textsc{comps} & \boxed{c} \\ \textsc{infl} & \left[\textsc{ms} \quad \boxed{m} \right] \end{array}\right]
\end{array}\right]$$
$$\wedge \neg \text{ member}\left(\left[\begin{array}{ll} \textit{aff-ss} & \\ \textsc{l}|\textsc{cont} & \textit{expl} \vee \textit{ana} \end{array}\right], \boxed{c} \right)$$

Figure 2: Lexical rule for causativised verbs[4]

$$\left[\begin{array}{ll}
\textsc{subj} & \left\langle \boxed{1}\ \mathrm{NP} \right\rangle \\
\textsc{comps} & \left\langle \mathrm{NP}[\textit{aff-ss}]_{\boxed{i}}\textit{:ana} \right\rangle \oplus \boxed{a} \\
\textsc{dtr} & \left[\begin{array}{ll} \textsc{subj} & \left\langle \boxed{1}\ \mathrm{NP}_{\boxed{i}} \right\rangle \\ \textsc{comps} & \left\langle \mathrm{NP}_{\boxed{i}} \right\rangle \bigcirc \boxed{a} \end{array}\right]
\end{array}\right]$$

Figure 3: Lexical rule for reflexives

which the only output is upstairs realisation, cf. **vous en fait vouloir*. Not only should causativisation be able to target inflected words, its output also needs to undergo inflection itself. This is not only to create the periphrastic dependency, but also to realise any pronominal argument corresponding to the downstairs subject: as exemplified in (10), such pronominals are always realised upstairs, even when the main affix cluster is realised downstairs.

(10) Cela lui fait vous en vouloir.
this IO.3SG makes OBJ.2PL GEN be_mad
'This makes her mad at you.'

In sum, an interface between morphology and syntax needs to be devised that can allow lexical rules of grammatical function change to apply to already inflected words, creating new lexical entries susceptible to further periphrastic inflection. In the following section we propose an IbM analysis of the morphological side of French causative constructions that attempts to meet these requirements.

$$\begin{bmatrix} \textsc{subj} & \left\langle \boxed{1}\,\mathrm{NP}_{\boxed{j}} \right\rangle \\ \textsc{comps} & \left\langle \mathrm{NP}[\textit{aff-ss}]_{\boxed{j}}\textit{:ana} \right\rangle \oplus \boxed{a} \\ \textsc{dtr} & \begin{bmatrix} \textsc{subj} & \left\langle \mathrm{NP}_{i} \right\rangle \\ \textsc{comps} & \left\langle \boxed{1}\,\mathrm{NP}_{\boxed{j}} \right\rangle \oplus \boxed{a} \end{bmatrix} \end{bmatrix}$$

Figure 4: Lexical rule for medio-passives

3 The morphology of French clitics: An IbM analysis

In order to integrate a general treatment of periphrasis into the morphological component, the first and most obvious step is to follow Bonami (2015) and recognise delegation of properties to an ancillary element as a means of expression. To this end, we shall augment realisation rules with a set-valued feature DEL that will serve to represent the morphosyntactic features delegated for expression by the ancillary element in a periphrase (see Figure 9 for sample rule types). As captured in Figure 5, the properties delegated by individual rules (under RR) are unioned together on the word's global DLS feature.

$$\textit{word} \rightarrow \begin{bmatrix} \textsc{ph} & \boxed{p_1} \oplus \ldots \oplus \boxed{p_n} \\ \textsc{infl} & \begin{bmatrix} \textsc{mps} & \boxed{e}\ (\boxed{e_1}\left[\textsc{ph}\ \ \boxed{p_1}\right] \bigcirc \ldots \bigcirc \boxed{e_n}\left[\textsc{ph}\ \ \boxed{p_n}\right]) \\ \textsc{ms} & \boxed{0}\ (\boxed{m_1} \uplus \cdots \uplus \boxed{m_n}) \\ \textsc{rr} & \left\{ \begin{bmatrix} \textsc{mph} & \boxed{e_1} \\ \textsc{mps} & \boxed{e} \\ \textsc{mud} & \boxed{m_1} \\ \textsc{ms} & \boxed{0} \\ \textsc{del} & \boxed{d_1} \\ \textsc{dls} & \boxed{d} \end{bmatrix}, \ldots, \begin{bmatrix} \textsc{mph} & \boxed{e_n} \\ \textsc{mps} & \boxed{e} \\ \textsc{mud} & \boxed{m_n} \\ \textsc{ms} & \boxed{0} \\ \textsc{del} & \boxed{d_n} \\ \textsc{dls} & \boxed{d} \end{bmatrix} \right\} \\ \textsc{dls} & \boxed{d}\ (\boxed{d_1} \cup \cdots \cup \boxed{d_n}) \end{bmatrix} \end{bmatrix}$$

Figure 5: Well-formedness

Pronominal affixation constitutes morphological realisation of syntactic arguments, which preempts their realisation as syntactic complements. Following Miller & Sag (1997), we assume an argument mapping constraint as given in Figure 6 that excludes affixal synsems from valence lists and instead adds appropriate marking and index features to the morphosyntactic property set, which serves as "input" for morphological realisation.

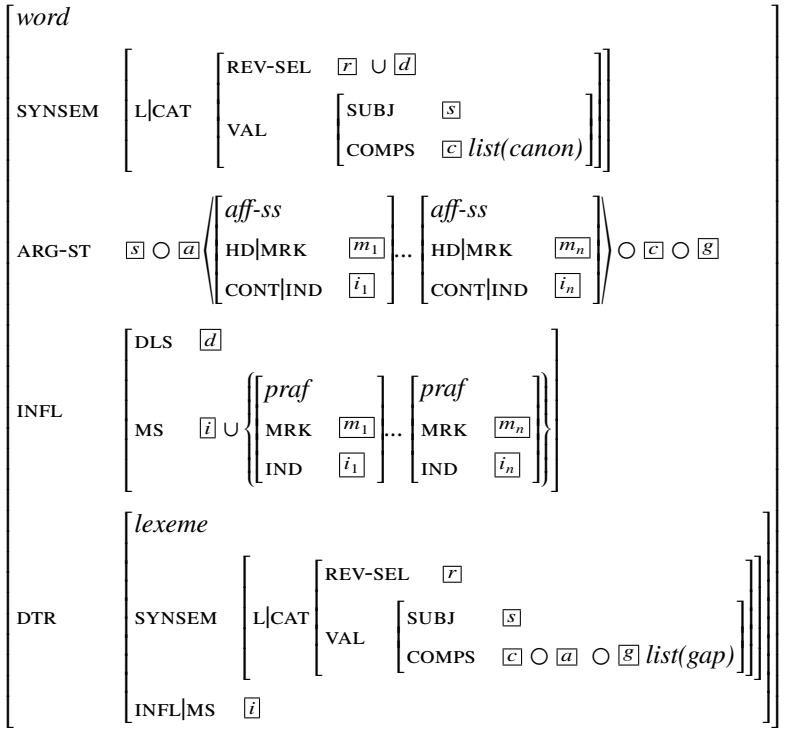


Figure 6: Argument mapping (Aguila-Multner & Crysmann, 2020a)

The morphology of French pronominal affixation provides rules for both local exponence and periphrastic expression. As we have argued above, "clitic climbing" is always dependent on the existence of periphrasis, either tense periphrasis with *avoir*/*être* or a periphrastic causative.

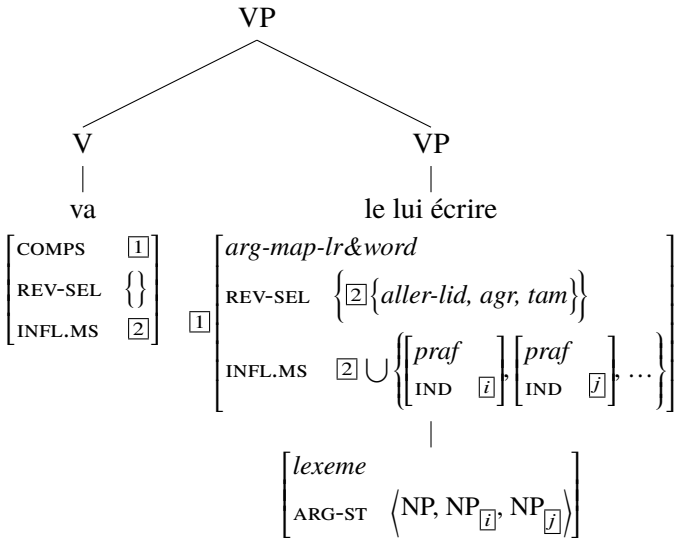


Figure 7: Sample derivation: local realisation of pronominal affixes in the future periphrase

While existence of an independent periphrase can be considered a necessary criterion, it is in itself not sufficient: e.g. the near future *futur proche* (11) is clearly periphrastic, but, nevertheless, it does not give rise to clitic climbing. Instead, we find local exponence of pronominal affixes on the lexical verb. The recent past (*venir de* + Vinf) behaves similarly.

(11) La professeure va le lui écrire.
the professor goes DO.3SG.M IO.3SG write
'The professor is going to write it to her.'

A sample derivation for local realisation in the future periphrase is given in Figure 7. In this case, REV-SEL contains selection of auxiliary *aller*, but not the pronominal affixes, which are realised on the lexical verb.

Figure 8 sketches the rules for local realisation. The rule types in the EXPO dimension provide the necessary pairing of features to be expressed (MUD) with the introduction of an exponent (MPH), complete with phonology and position class information. The COND hierarchy, by contrast, specifies the two situations where local exponence is possible: either if the word does not involve any periphrastic expression at all (empty DLS set), or else, if the periphrasis involves *aller* or *venir de*.

Rules for periphrastic realisation, by contrast, are given in Figure 9. Starting on the right of the hierarchy, we find the rule types for near tenses. As these tenses do

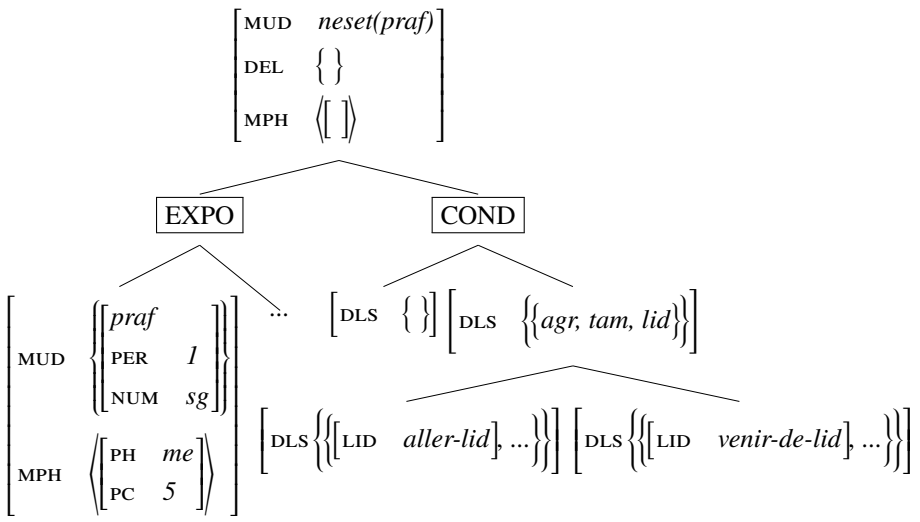


Figure 8: Local realisation

not give rise to clitic climbing, their MUD value is restricted to TAM and agreement properties only, which are expressed analytically by selection of an infinitival stem (STM9) and selection of, e.g., a present tense ancillary element, which also takes on expression of subject agreement properties. The subtypes pair the specific tense value to be expressed (in MUD) with selection of an appropriate ancillary element.

In the middle of the hierarchy, we find rules for tense periphrasis with *avoir/être*. Their MUD value includes a set of *praf* specifications, alongside TAM and subject agreement, capturing the potential for clitic climbing. Analytic expression of TAM involves, again, selection of a specific stem, here: a participial stem (STM12),[5] and selection of an ancillary element. Delegation of *praf* and agreement properties to the ancillary element is captured by reentrancies between MUD and DEL. Subtypes illustrate the two uses of the *passé composé* as a simple past, and as a more compositional perfective.

Clause-union style periphrasis, finally, is represented on the left. In addition to selection of an appropriate ancillary element, e.g. *faire*, periphrastic expression of this type also involves delegation of agreement and *praf* properties. Subtypes of this rule type serve to constrain the range of ancillary elements that can feature in this type of periphrastic construction, which in addition to the causative verbs *faire* and *laisser*, includes the perception verbs.

[5]We assume an analysis of basic French verbal morphology in terms of stem allomorphy plus suffixal exponence such as the one developed in Bonami & Boyé (2002, 2003, 2006). Stem indices follow Bonami & Boyé (2007: 313).

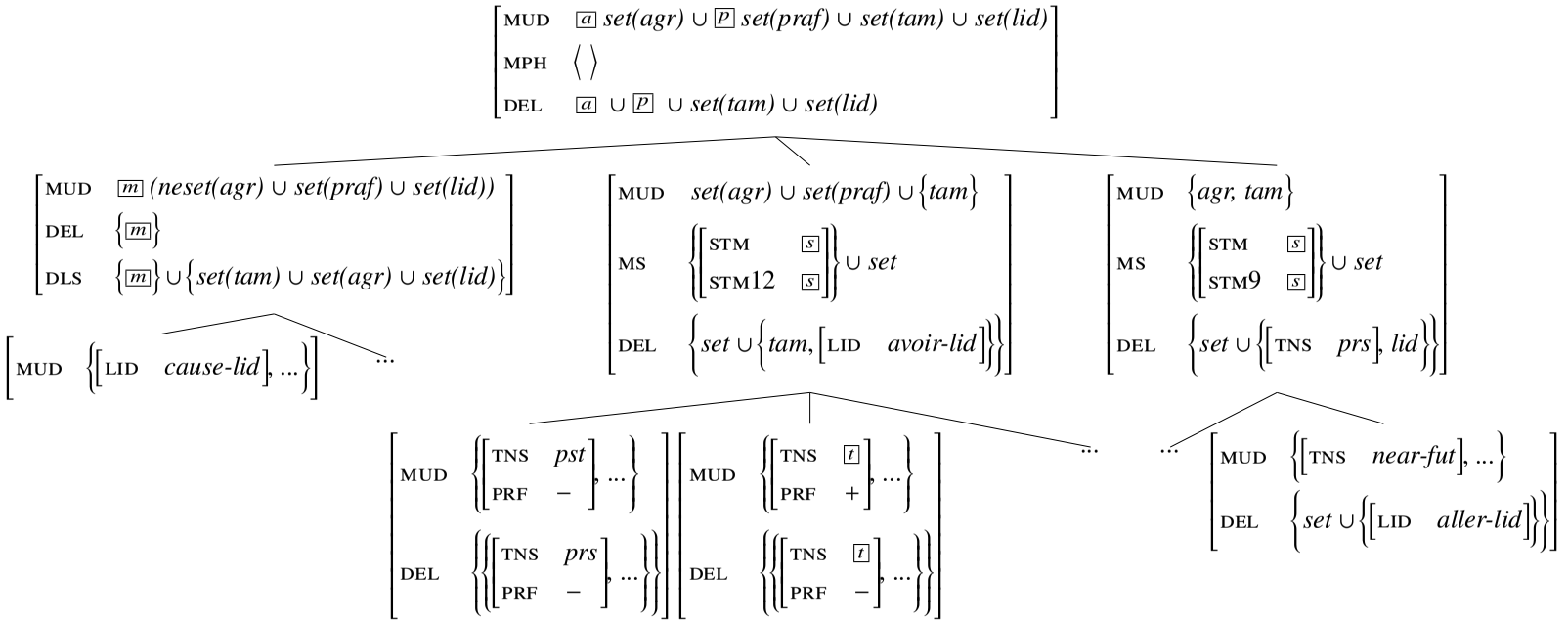


Figure 9: Periphrastic realisation

3.1 Tense periphrasis and clitic climbing

Let us briefly provide some examples as to how some core cases of periphrastic and local realisation can be captured with the rule system we have just set up. To start with, we shall discuss cases of simple tense periphrasis to which we shall add realisation of pronominal arguments.

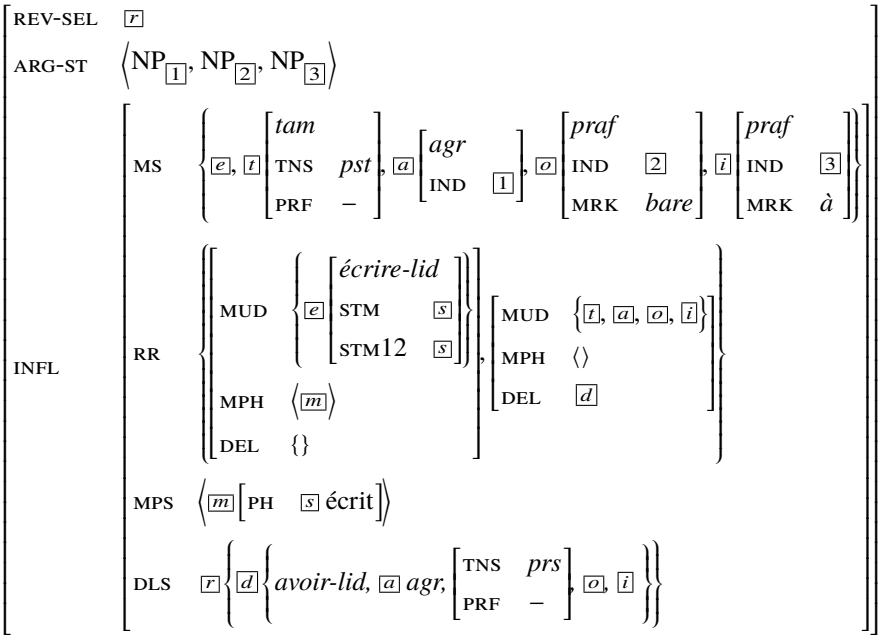


Figure 10: Periphrastic realisation of tense and pronominal affixation (*le lui a écrit*)

To start with, let us consider the morphology of the periphrastic past, showing both tense periphrasis and clitic climbing. We shall use our example from Figure 1 above, and complement the syntactic percolation shown there with an explicit representation as to how the periphrastic dependencies are launched within IbM.

As shown in Figure 10, the morphosyntactic property set (MS) consists of a lexeme identifier (*écrire-lid*), a simple past (non-perfective) TAM specification, an agreement specification (coindexed with the highest argument), and two *praf* specifications (conindexed with the direct and indirect object, respectively).

Realisation rules (RR) specify how these properties are expressed, using two rules: (i) a rule of exponence that realises lexemic identity by inserting a STM into the list of morphs MPH, and (ii) a rule of periphrasis that delegates the expression of agreement, TAM and pronominal affixation (via DEL). Having chosen the rule for periphrastic simple past, note that the TNS specification for the ancillary element (*prs*) differs from that on MS (*pst*), illustrating the non-compositionality. Furthermore, selection for a participial stem (STM12) is also effected by the rule of periphrasis (cf. 9). Periphrastic realisation of pronominal affixes follows from the constraints on local and periphrastic realisation given above, since (i) the constraints for

periphrastic realisation (Figure 9) of the simple past include delegation of pronominal affixation and (ii) the constraints for local realisation do not include (Figure 8) past or perfective periphrasis.

The second case of tense periphrasis concerns near tenses, which display periphrastic realisation of agreement and TAM, yet local realisation of pronominal affixes. Again, Figure 11 provides the morphological derivation of the periphrase illustrated by the tree in Figure 7 above.

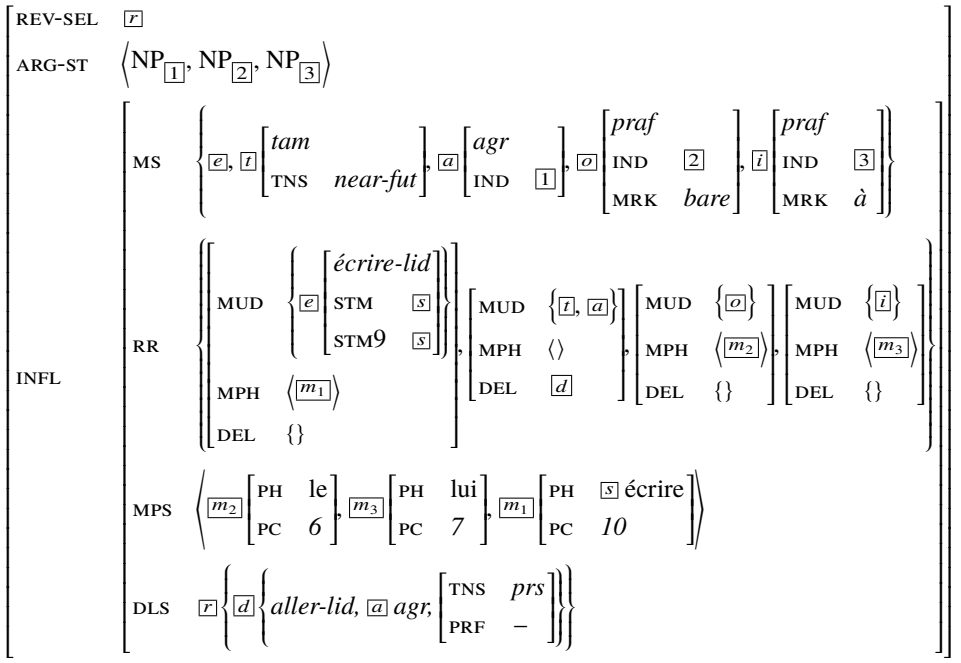


Figure 11: Periphrastic tense and local pronominal affixation (*va le lui écrire*)

As stated by the near future periphrasis rule in Figure 9 (bottom right), this tense category is expressed by selecting the infinitival stem (STM9) and selecting a present tense form of *aller* 'go' as the ancillary element. While expression of (subject) agreement is delegated to the ancillary element as well, pronominal affixation is not included. Conversely, the constraints on local realisation for *praf* values given in Figure 8 do license direct realisation on the lexical verb.

Having seen how interaction between rules of exponence and periphrasis can be captured straightforwardly with IbM, we shall turn to the case of causatives in the next subsection.

3.2 Trapping and climbing with causatives

Having laid out the essentials of local and periphrastic realisation, we shall finally turn to the treatment of trapping vs. climbing. As we have observed above, causativisation and reflexivisation can apply in either order: when reflexivisation applies

before causativisation, the reflexive argument is bound by the downstairs subject. This order of application has the peculiar morphological effect that pronominal affixes pertaining to downstairs arguments cannot climb, with the notable exception of the downstairs subject. When reflexivisation applies after causativisation, the reflexive is bound by the causer. As for morphological realisation, we find climbing of all pronominal affixes in this case, rather than trapping. Taken together, morphological realisation must be able to apply both before (trapping) and after (climbing) causativisation.

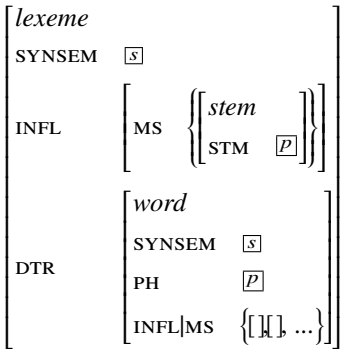


Figure 12: Word-to-Lexeme LR

We therefore conclude that inflected words can be submitted to further rules of grammatical function change, which we take to be lexeme-to-lexeme rules. Building on a recent proposal by Salehi & Koenig (2023), we assume a word-to-lexeme rule as sketched in Figure 12 to resubmit an inflected word to rules such as the causativation rule. As an important side-effect, the rule registers the inflection applied up to this point by inserting it as a stem form to which further morphology may apply.

The possibility of having more than one place where inflection can apply in the derivation of complex words certainly deviates from a more absolute view of the place of realisational morphology as applying only once to fully formed words. Nevertheless, the proposal made here is still far more restrictive than the kind of direct interleaving of lexical rules with morphophonological effects often adopted in HPSG.

Moreover, the need for multiple places to interface derivation with inflection has been forcefully argued for by Koenig & Michelson (2020) on the basis of Oneida. They observe that a class of derived nominals in Oneida can take as their basis verb stems that have already undergone some verbal inflection. Once turned into nouns by a derivational step, they can be further inflected with nominal morphology.

Before we close, we shall briefly discuss how the current setup can account for the interaction between different lexical rules of grammatical function change and its impact on morphological realisation.

Let us start with the simplest case of causativisation (cf. Figure 2) as the only rule of grammatical function change. A sample derivation is given in Figure 13 for

les fait voir aux élèves ‘makes the students see them’. Applying argument mapping (Figure 6) to the output of the causativisation rule will insert morphosyntactic specification for all pronominal arguments, including downstairs pronominal objects and the erstwhile downstairs subject that has been demoted to direct or indirect object by the causativisation rule. Furthermore, the causative LR itself has added a requirement for morphological expression of causation. Since delegation to an ancillary element is the only way to express a causative, the respective rule in Figure 9 will apply, satisfying specification for pronominal affixes equally by delegation.

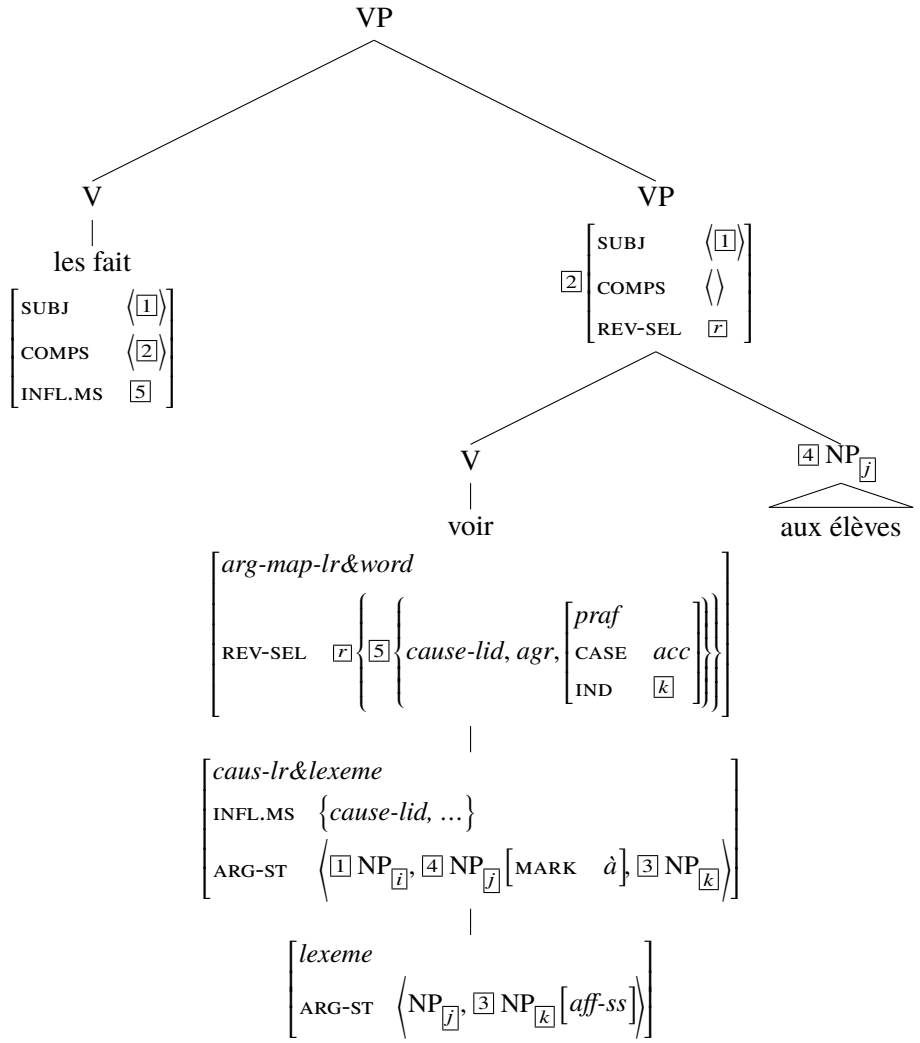


Figure 13: Sample derivation: causative periphrasis

In case either the reflexive LR (Figure 3) or medio-passive LR (Figure 4) has applied to the output of the causative LR, the anaphoric element introduced by these rules will be coindexed with the additional causer subject. This is illustrated in the derivation in Figure 14 for a long medio-passive *se font réparer pour pas cher* ‘get

repaired cheaply'. Morphologically, nothing much will change, with climbing – i.e. delegation of *cause-lid* and all pronominal affixes – as the most straightforward option.

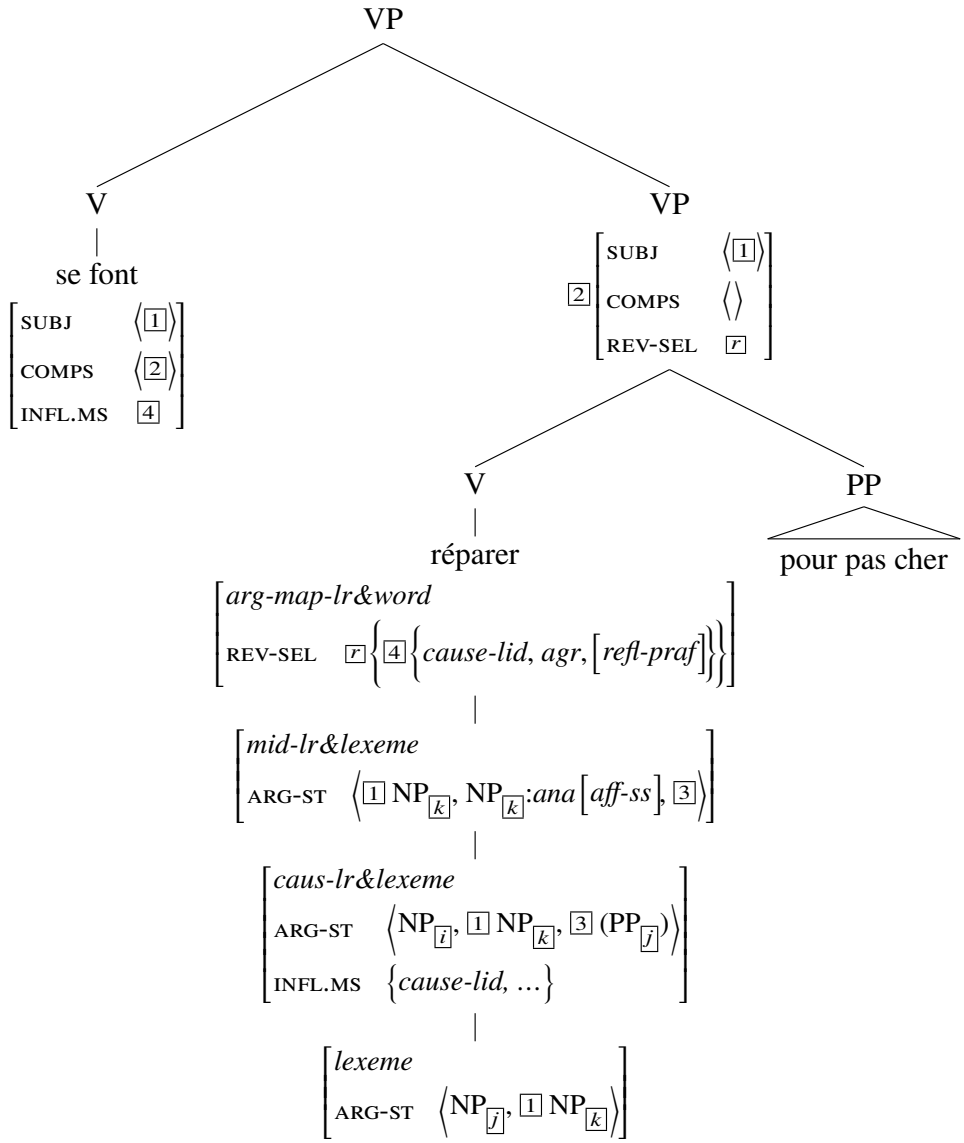


Figure 14: Sample derivation: long medio-passive

Turning now to the opposite order, i.e. reflexive or medio-passive rules feeding causativisation (illustrated in the sample derivation in Figure 15 for a causativised medio-passive *fait se vendre les classiques* 'makes the classics sell well'), we find that direct application of the causative LR (Figure 2) to the output of either Figure 3 or 4 is blocked by the causative LR's constraint on COMPS not to have an intrinsic or anaphoric *aff-ss* complement. If, however, we apply argument mapping (Fig-

ure 6) to the output of reflexivisation, non-canonical synsems will be suppressed on COMPS, since argument conservation (Ginzburg & Sag, 2000) is a constraint on *word* (not *lexeme*). Concomitantly, the morphosyntactic property set of this word will be realised by inflectional morphology. Realisation of pronominal affixes typically will be local, using the rule types in Figure 8, given the causative LR has not applied yet at this point, and therefore there is no requirement for any periphrastic realisation. In order to apply the causative LR to this inflected word, it first needs to be converted into a lexeme by the rule in Figure 12: the most important effect of this rule is that it preserves the inflected form as a stem on which further morphology can operate. Application of the causative LR will essentially proceed as in the simple case given above: since all remaining downstairs complements are of type *canon-ss*, the only clitic that can climb as part of the periphrastic causative is the erstwhile downstairs subject, demoted to COMPS by the causative LR.

Given that we permit inflected words to be fed to rules of grammatical function change, an important question is that of overgeneration. E.g., we predict that downstairs cliticisation should in principle be possible with causatives even if the downstairs verb does not have any intrinsic or reflexive affixal arguments. Indeed, this is what we find in a number of situations: first, if a coordination of verbs is embedded under a causative, we find downstairs realisation when the verbs differ in the type or number of pronominal arguments they take. A most straightforward example is a coordination where the first VP contains the antecedent, and the second one a coreferential pronominal (12). Second, downstairs realisation can be chosen to avoid violations of morphotactic constraints on the clitic cluster: e.g. with causatives of ditransitives, we find the downstairs subject as a dative clitic on the causative verb, while the downstairs dative (and all other clitics) appear on the lexical verb (13). The very strong tendency to use climbing whenever possible can then be understood as a dispreference for splitting inflection without need.

(12) Elle a fait écrire un poème aux enfants et le lire aux parents.
she has made write a poem to.the children and DO.3SG.M read to.the parents
'She made the children write a poem and the parents read it.'

(13) Elle lui a fait leur donner les cadeaux.
she IO.3SG has made IO.3PL give the presents
'She made her give them the presents.'

Before we conclude, we shall briefly illustrate how the morphology can capture split realisation. Recall that downstairs realisation is the result of realising inflectional morphology before causativisation (cf. Figure 15). Therefore, the MS set of the downstairs verb looks just like an ordinary infinitive, possibly including a specification for pronominal affixes. These are realised locally, as shown in Figure 16.

In order to causativise an already inflected verb, we first convert the *word* to a *lexeme*, by way of the word-to-lexeme LR: as shown in Figure 12, this rule will encapsulate the inflected form as a stem (cf. Salehi & Koenig, 2023). Application of the causativisation rule will add a causer argument to the front of ARG-ST and

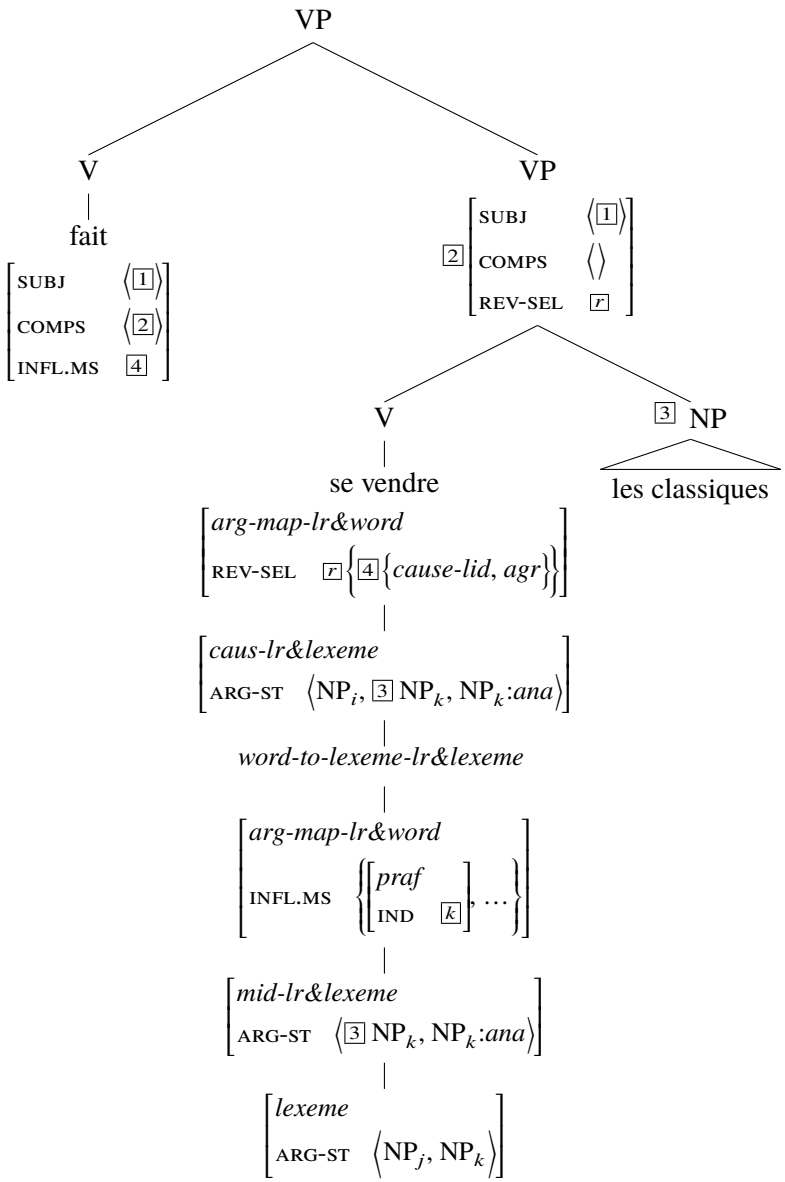


Figure 15: Sample derivation: downstairs medio-passive

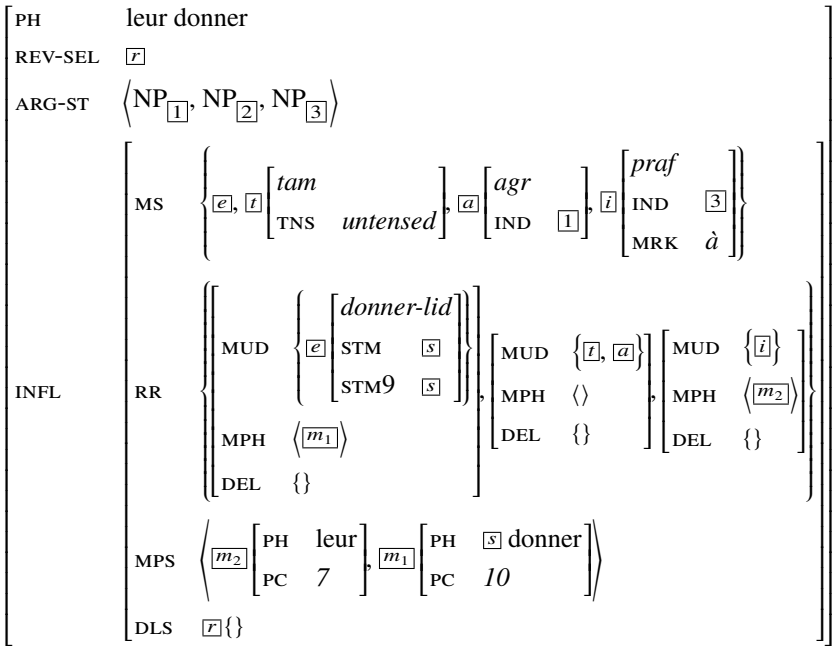


Figure 16: Local pronominal affixation on infinitives (*leur donner les cadeaux*)

thereby demote the erstwhile downstairs subject to (in)direct object. As a result, the subject-demotion effect of causativisation makes available the erstwhile subject of the infinitive for realisation as a pronominal affix. As stated by the causative realisation rule type in Figure 9 (left-most leaf type), realisation of *cause-lid* is delegated to the ancillary element, together with agreement information and any pronominal affixes. This is shown for our concrete case in Figure 17.

4 Conclusion

In this paper we have discussed how the morphosyntax of clitic climbing and trapping in French causatives can be integrated with an inferential-realisational model such as Information-based Morphology (IbM; Crysmann & Bonami, 2015). Building on our earlier work (Aguila-Multner & Crysmann, 2020a; Aguila-Multner, 2023), where we proposed an approach of clitic climbing as periphrastic realisation of pronominal affixes, we have now provided a formalisation of periphrasis rules in IbM. Most specifically, we use a pair of features to capture rule-local delegation (DEL) to an ancillary element, as well as constrain the interaction of multiple periphrastic dependencies (DLS) at the global word-level. This distinction mimics that between MUD and MS for the expression and conditioning of morphosyntactic properties, as well as the more recent distinction between rule-local morphotactic contribution (MPH) and word-wide morphotactic constraints (DLS), cf. Crysmann (2023).

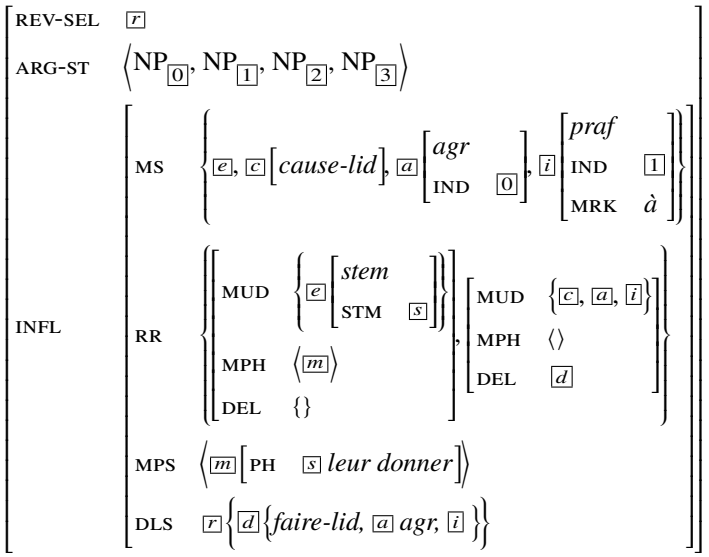


Figure 17: Split realisation: climbing of downstairs "subject" (*lui faire leur donner les cadeaux*)

Finally, we addressed the interaction between causatives, reflexives and mediopassives, arguing that inflectional morphology can apply in two steps, thereby capturing the concomitant effects of clitic trapping vs. clitic climbing. By allowing inflected words to be submitted to further rules of grammatical function change, we provide for a clean separation between lexical rules and inflectional morphology, keeping the benefits of an inferential-realisational model while providing sufficient flexibility to address complex interactions of grammatical function change.

References

Abeillé, Anne & Danièle Godard. 1996. La complémentation des auxiliaires français. *Langages* 30(122). 32–61.

Abeillé, Anne & Danièle Godard. 2002. The syntactic structure of French auxiliaries. *Language* 78(3). 404–452.

Aguila-Multner, Gabrielle. 2023. *The morphosyntax of French complex predicates: Clitic climbing and periphrasis*: Université Paris Cité dissertation.

Aguila-Multner, Gabrielle & Berthold Crysmann. 2020a. French clitic climbing as periphrasis. *Lingvisticæ Investigationes* 43. 23–61.

Aguila-Multner, Gabrielle & Berthold Crysmann. 2020b. An inside-out approach to French causatives. In *Proceedings of the 27th International Conference on Head-Driven Phrase Structure Grammar*, 5–25. doi:10.21248/hpsg.2020.1. `https://proceedings.hpsg.xyz/article/view/397`.

Aissen, Judith & David M. Perlmutter. 1976. Clause reduction in Spanish. In *Annual meeting of the Berkeley Linguistics Society*, vol. 2, 1–30.

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25(1). 63–110. doi:10.1007/s11525-015-9254-3.

Bonami, Olivier & Gilles Boyé. 2002. Suppletion and dependency in inflectional morphology. In *The proceedings of the HPSG 2001 conference*, Stanford: CSLI Publications.

Bonami, Olivier & Gilles Boyé. 2003. Supplétion et classes flexionnelles. *Langages* 152. 102–126.

Bonami, Olivier & Gilles Boyé. 2006. Deriving inflectional irregularity. In *Proceedings of the 13th international conference on HPSG*, 39–59.

Bonami, Olivier & Gilles Boyé. 2007. French pronominal clitics and the design of paradigm function morphology. In *Mediterranean morphology meetings*, vol. 5, 291–322.

Crysmann, Berthold. 2023. Morphotactic competition in Murrinh-Patha: Rule composition and rule interaction in Information-based Morphology. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 27–44. Frankfurt/Main: University Library. doi:10.21248/hpsg.2023.2.

Crysmann, Berthold & Olivier Bonami. 2015. Variable morphotactics in Information-based Morphology. *Journal of Linguistics* 52(2). 311–374. doi: 10.1017/S0022226715000018.

Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations. the form, meaning, and use of English interrogatives*. Stanford, CA: CSLI Publications.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI publications.

Koenig, Jean-Pierre & Karin Michelson. 2020. Derived nouns and structured inflection in Oneida. *Lingue e linguaggio* 19(1). 9–33.

Miller, Philip H. 1992. *Clitics and constituents in phrase structure grammar*. Garland, New York.

Miller, Philip H. & Ivan A. Sag. 1997. French clitic movement without clitics or movement. *Natural Language & Linguistic Theory* 15(3). 573–639.

Popova, Gergana & Andrew Spencer. 2013. Relatedness in periphrasis: A paradigm-based perspective. In Marina Chumakina & Greville G. Corbett (eds.), *Periphrasis: The role of syntax and morphology in paradigms*, 191–225. Oxford: Oxford University Press.

Salehi, Ali & Jean-Pierre Koenig. 2023. Argument realization and argument ref- erencing in Soranî Kurdish. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 100–120. Frankfurt/Main: University Library.

Villalba, Xavier. 1994. Clitic climbing in causative constructions. *Catalan working papers in linguistics* 3(2). 123–152.

Vincent, Nigel & Kersti Börjars. 1996. Suppletion and syntactic theory. In *Proceedings of the 1st LFG colloquium*, 448–62.

Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English n't. *Language* 502–513.

On exhaustive conditional clauses in Modern Standard Arabic

Abeer Alsulami
King Saud University, Riyadh

Robert D. Borsley
University of Essex and Bangor University

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 25–46

Keywords: Modern Standard Arabic, exhaustive conditional, universal exhaustive conditional, alternative exhaustive conditional

Alsulami, Abeer & Robert D. Borsley. 2024. On exhaustive conditional clauses in Modern Standard Arabic. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 25–46. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.2.
CC-BY 4.0

Abstract

Simple conditional clauses identify a single condition under which a modified clause is true. In contrast, exhaustive conditionals (ECs) identify a set of conditions under all of which the clause is true. Two binary distinctions give four possible types of EC. Three of these are found in Modern Standard Arabic (MSA). Unlike English ECs, MSA ECs have essentially the same distribution as simple conditionals. Three rather different analyses seem appropriate for the three types, but they share a number of properties allowing the similarities between the three types to be captured.

1. Introduction

Simple conditional clauses, exemplified by English *if*-clauses, have been a focus of research by logicians and linguists for a very long time. (See e.g. Huddleston & Pullum 2002: 738-760 and Bhatt & Pancheva 2006 for useful discussion of many of the issues.) Over the last two decades, what are known as exhaustive conditionals (ECs) (or unconditionals), exemplified in English by examples like the following have also received some attention. (See e.g. Huddleston & Pullum 2002: 761-5, 985-91, Rawlins 2008, 2013, Arnold & Borsley 2014.)

(1) a. whatever you say
    b. whether you speak or not

Whereas simple conditional clauses identify a condition under which the main clause they modify is true, ECs identify a set of two or more conditions under all of which the clause is true. Simple conditionals can also refer to more than one condition, e.g. with *any*, as in *if you say anything*, or with *or*, as in *if you go to Paris or Rome*, but there is no requirement that they do so.

There is an important semantic distinction between universal ECs and alternative ECs. The former, exemplified by (1a) and the examples in (2), refer to all conditions of a certain form.

(2) a. whatever you read (all conditions of the form *you read x*)
    b. wherever you go (all conditions of the form *you go to x*)

* We are grateful to a number of colleagues for useful discussion of the issues addressed here, including Doug Arnold, Dan Flickinger, and Jacob Maché, and also to two referees for helpful comments. We alone are responsible for what appears here.

The latter, exemplified by (1b) and the following examples, essentially list the conditions.

(3) a. whether you go or not
    b. whether you go to Paris or to Berlin
    c. whether you go to Paris or to Berlin or Rome

While alternative ECs often involve *or not*, (3b) shows that they don't have to, and (3c) shows that they may identify more than two conditions.

Separate from this semantic distinction is a formal distinction, highlighted in Huddleston and Pullum (2002: 761-765), between ungoverned ECs and governed ECs. The former, exemplified by all the examples presented so far, involve just a clause of some kind. The latter involve a clause which is a dependent of an element such as *no matter*. The following illustrate:

(4) a. no matter what you read
    b. no matter where you go

These two distinctions give four types of ECs, as follows:

(5) Ungoverned universal ECs
    Ungoverned alternative ECs
    Governed universal ECs
    Governed alternative ECs

Of course, a language may not have all these types.

It is clear from Haspelmath & König (1998) that these two distinctions are relevant to many languages. This includes Modern Standard Arabic (MSA), but we will show below that MSA only has three of the four types identified above. Moreover, in MSA, it is not just the internal structure of ECs that is of interest but also their distribution. This is more like that of simple conditionals than it is with their English counterparts. Like simple conditionals, they can appear in both an ordinary head-adjunct clause and in an MSA counterpart of an English *if-then* clause.

The aim of this paper is to explore both the internal structure and the distribution of MSA ECs, and develop analyses within HPSG. Our main focus will be on syntax, but we will also consider semantics.

2. The basic data

MSA has **ungoverned universal ECs**, involving just a clause and referring to all conditions of a certain form, which are broadly similar to their English counterparts:

(6) [mahmɑ faʕala-t l-llajnat-u] sa-taðʕallu
whatever do.PAST.3SGF DEF-committee-NOM] will-continue
l-ʔintiqɑːdat-u tuwajjah ʔilay-hɑ
DEF-criticisms-NOM direct.PASS to-it.3SGF
'Whatever the committee does, criticisms will still be directed at it.'

The initial constituent may be nominal, as in (6), or adverbial, as in (7):

(7) [matamɑ takun l-ħaflat-u] ʔaðhab ʔilay-haa
whenever be.JUSS.3SGM the-party-NOM go.JUSS.1SG to-3SGF
'Whenever the party is, I'm going to it.'

Like their English counterparts, they appear to be head-filler phrases with one of a small set of lexical items in the filler. In addition to the items already illustrated, they may contain the following:

(8) *ʔayy* 'whoever'
ʔaynama 'wherever'
ħayθuma 'wherever'
kullama 'whenever'
kayfamaː 'however'

They may also have more complex NP or PP fillers, as the following show:

(9) a. [[ʔayy-a kitɑːb-in] taqraʔ] lan tastafiːda min-hu
whichever-ACC book-GEN read.2SGM NEG benefit.2SGM from-it
'Whichever book you read; you won't benefit from it.'
b. [[min ʔayy-i dawlat-in] qadim-ta] ʔanta
from whichever-GEN country-GEN came-2SGM 2SGM
muraħab-un bi-ka
welcome-NOM with-2SGM
'Whichever country you come from; you are welcome.'

In English, ungoverned universal ECs look like free relatives and it has sometimes been proposed that that is what they are. (See Rawlins 2008: 2.1.3 for critical discussion). In MSA, some free relatives look like ECs:

(10) saʔaʃtarii la-ka [mahmɑ turiidu]
will-buy.1SG.M/F for-2SGM whatever want.2SGM
'I will buy for you whatever you want.'

But free relatives are often quite different:

(11) saʔaχtɑːru [ʔallað turiːdu / turiːdu-hu].
will-choose.1SGM COMP want.2SGM / want.2SGM-it
'I will choose whatever you want.'

The free relative here is identical to an ordinary relative clause.

(12) saʔaχtɑːru l-kitɑːb-a [ʔallað turiːdu / turiːdu-hu]
will-choose.1SGM DEF-book-ACC COMP want.2SGM / want.2SGM-it
'I will choose the book you want.'

It is essentially a relative clause without a visible antecedent, and there is evidence that the element that introduces it is a complementizer (Alqurashi 2012). There are no ECs like this. Hence, there is no reason to consider a free relative analysis for ungoverned universal ECs in MSA.

In English, it has been argued by Huddleston and Pullum (2002: 761-765) and Rawlins (2008: 2.1.3, 2013: 3.1) that ungoverned universal ECs are *wh*-interrogatives. In MSA, ungoverned universal ECs cannot be *wh*-interrogatives because they have a different set of lexical items in the filler. Thus, the following are not possible interrogatives:

(13) a. *mahmɑ faʕala-t l-llajnat-u
whatever do.PAST.3SGF DEF-committee-NOM
'Whatever does the committee do?'
b. *matamɑ takunu l-ħaflat-u
whenever be.3SGM DEF-party-NOM
'Whenever is the party?'

Instead MSA has the following:

(14) a. maɑː faʕala-t l-llajnat-u
what do.PAST.3SGF DEF-committee-NOM
'What does the committee do?'
b. matɑː takunu l-ħaflat-u
when be.3SGM DEF-party-NOM
'When is the party?'

But although MSA ungoverned universal ECs are not *wh*-interrogatives, they are like *wh*-interrogatives in identifying a set of possible situations, and they indicate that all the situations are ones in which the modified clause is true.

MSA also has **ungoverned alternative ECs**, which look quite like their English translations:

(15) a. [ʔa-ðahab-ta ʔilaɑ:baris ʔam lam taðhab]
(Q)-go.PAST-2SGM to Paris or not go.PRES.2SGM
sa-taqdˤii: waqt-an mumtiʕ-an
will-have.2SGM time-ACC good-ACC
'Whether you go to Paris or not, you'll have a good time.'

b. [ʔa-ðahab-ta ʔilaɑ:baris ʔam ʔilaɑ:ru:mɑ:] (Q)-
go.PAST-2SGM to Paris or to Rome
sa-taqdˤii: waqt-an mumtiʕ-an
will-have.2SGM time.ACC good-ACC
'Whether you go to Paris or Rome, you'll have a good time.'

c. [ʔa-ðahab-ta ʔilaɑ:baris ʔam ʔilaɑ:ru:mɑ: ʔam ʔilaɑ:
(Q)-go.PAST-2SGM to Paris or to Rome or to
berli:n] sa-taqdˤii: waqt-an mumtiʕ-an
Berlin will-have.2SGM time.ACC good-ACC
'Whether you go to Paris or Rome or Berlin, you'll have a good time.'

These ECs are in fact identical to alternative interrogatives, which have the same form in both main clauses and complement clauses:

(16) a. (ʔa)-ðahab-ta ʔilaɑ:baris ʔam lam taðhab
(Q)-go.PAST.2SGM to Paris or not go.PRES.2SGM
'Did you go to Paris or not?'

b. saʔaluu:n-i: [(ʔa)-ðahab-ta ʔilaɑ:baris ʔam lam
ask.PAST.3PLM-1SGM/F Q-go.PAST-2SGM to Paris or not
taðhab]
go.PRES.2SGM
'They asked me whether you went to Paris or not.'

Ungoverned alternative ECs identify two or more possible situations in the same way as alternative interrogatives and indicate that all the situations are ones in which the modified clause true.

MSA also has **governed alternative ECs**, involving *sawa:ʔ-un* 'same' followed by an alternative interrogative:

(17) a. [sawɑ:ʔ-un [(ʔa)-ðahab-ta ʔilaɑ:baris ʔam lam
same-NOM Q-go.PAST-2SGM to Paris or not
taðhab]] sa-taqdˤii: waqt-an mumtiʕ-an
go.PRES.2SGM will-have.2SGM time-ACC good-ACC
'No matter whether you go to Paris or not, you'll have a good time.'

b. [sawɑːʔ-un [(ʔa)-ðahab-ta ʔilɑɑː baris ʔam ʔilɑɑː
same-NOM (Q)-go.PAST-2SGM to Paris or to
ruːmɑː]] sa-taqdˁiː waqt-an mumtiʕ-an
Rome will-have.2SGM time.ACC good-ACC
'No matter whether you go to Paris or Rome, you'll have a good time.'

These ECs look rather like English ECs with *no matter*. They look even more like certain Finnish ECs which also involve a word meaning 'same', e.g. the following from Haspelmath & König (1998: 618):

(18) [Ihan sama [mitä hän sanoo]], mies psyy vaiti.
quite same what she says man stays silent
'No matter what she says, he keeps quiet.'

The appearance of a word meaning 'same' clearly reflects the fact that the main clause is true in all of the situations identified by the EC. Hence, they are all equally good, or the same. Haspelmath & König also give similar examples from Polish (19) and Romani (20).

(19) [Wszystko jedno [czy pójdziemy do teatruczy czy
all one whether go.1PL to theatre whether
zostaniemy w domu]], chciałbym spędzicć ten wieczór
stay.1PL at home want.SUBJ.1SG spend.INF this evening
z tobą
with you
'Whether we go to the theatre or spend the evening at home, I would like to spend the evening with you.' (1998: 601)

(20) [Sa jekh [kaj voj ža-l-a]] vov šoha či mekh-el-a
all one where she go-3SG-FUT he never not leave-3SG-FUT
la
he
'No matter where she goes, he will never leave her.' (1998: 618)

In both, 'all one' indicates that all the conditions are equally good.

The examples in (4), (18) and (20) contain a *wh*-interrogative. The example in (19) contains an alternative interrogative, but a *wh*-interrogative is also possible after *wszystko jedno*, as (21) illustrates.[1]

(21) [wszystko jedno [gdzie pójdziemy]]
all one where go.1PL
'wherever you go'

[1] We are grateful to Ewa Jaworska for advice on Polish.

In contrast, the MSA construction can only contain an alternative interrogative. Thus, the following with a *wh*-interrogative are ungrammatical:

(22) a. *[sawɑːʔ-un [maɑː faʕala-ta]], sa-taqdˁii: waqt-an
same-NOM what do.PAST-2SGM will-have.2SGM time.ACC
mumtiʕ-an]
good-ACC
'No matter what you do, you will have a good time.'

b. *[sawɑːʔ-un [matɑː ðahab-ta]], sa-taqdˁii: waqt-an
same-NOM when go.PAST-2SGM will-have.2SGM time.ACC
mumtiʕ-an]
good-ACC
'No matter when you go, you will have a good time.'

These examples would also be ungrammatical with *mahma* and *matama* instead of *maaː* and matɑː. Thus, *sawaːʔ-un* cannot combine with an ungoverned universal EC any more than a *wh*-interrogative. It seems, then, that MSA has no governed universal ECs.

Turning to the **distribution of MSA ECs**, they are like simple conditional clauses and other adjunct clauses in modifying an ordinary clause that can stand alone. The following show that the second clauses in (6) and (15a) are ones that can stand alone:

(23) sa-taðˁallu l-ʔintiqɑːdat-u tuwajjah ʔilay-hɑ
will-continue DEF-criticisms-NOM direct.PASS to-it.3SGF
'Criticisms will still be directed at it.'

(24) sa-taqdˁii: waqt-an mumtiʕ-an
will-have.2SGM time-ACC good-ACC
'You'll have a good time.'

The following show that simple conditional clauses and other adjunct clauses can modify an ordinary clause that can stand alone.

(25) a. [ʔiðɑː ʔaχtˁaʔa -t l-llajnat-u]
if makes a mistake.PAST.3SGF DEF-committee-NOM
sa- tuwajjah l-ʔintiqɑːdat-u ʔilay-hɑ
will-direct.PASS DEF-criticisms-NOM to-it.3SGF
'If the committee makes mistake, criticisms will be directed at it.'

b. sa- tuwajjah l-ʔintiqɑːdat-u ʔilay-hɑ
will-direct.PASS DEF-criticisms-NOM to-it.3SGF
'Criticisms will be directed at it.'

(26) a. [ħiːna / ħiːnamɑː tuχtˁiʔu
when when makes a mistake.PRES.3SGF
l-llajnat-u] tuwajjahu l-ʔintiqɑːdat-u
DEF-committee-NOM direct.PASS DEF-criticisms-NOM
ʔilay-hɑ
to-it.3SGF
'When the committee makes a mistake, criticisms are directed at it.
b. tuwajjahu l-ʔintiqɑːdat-u ʔilay-hɑ
direct.PASS DEF-criticisms-NOM to-it.3SGF
'Criticisms are directed at it.'

As one might expect, the adjunct clause can precede or follow the clause it modifies. These are alternative versions of (6) (with an ungoverned universal EC), (25a) (with a simple conditional clause), and (26a) (with a 'when' clause):

(27) sa-taðˁallu l-ʔintiqɑːdat-u tuwajjah ʔila
will-continue DEF-criticisms-NOM direct.PASS to
l-llajnat-i [mahmɑ faʕala-t]
DEF-commit()tee-GEN whatever do.PAST.3SGF
'Criticisms will still be directed at the committee, whatever it does.'
(28) sa-taðˁallu l-ʔintiqɑːdat-u tuwajjah ʔila
will-continue DEF-criticisms-NOM direct.PASS to
l-llajnat-i [ʔiðɑː ʔaχtˁaʔa -t]
DEF-committee-GEN if makes a mistake.PAST.3SGF
'Criticisms will still be directed at the committee, if it makes a mistake.'
(29) tuwajjahu l-ʔintiqɑːdat-u ʔila l-llajnat-i
direct.PASS DEF-criticisms-NOM to DEF-committee-GEN
[ħiːna / ħiːnamɑː tuχtˁiʔu]
when makes a mistake.PRES.3SGF
'Criticisma are directed at the committee, when it makes a mistake.'

In MSA, as in English, simple conditionals can also modify a clause with a special marking which cannot stand alone. We have examples like (30a), where, as (30b) shows, the modified clause cannot stand alone:

(30) a. ʔiðɑː ʔaχtˁaʔa -t l-llajnat-u]
if makes a mistake.PAST.3SGF DEF-committee-NOM
fa-sa-tuwajjah l-ʔintiqɑːdat-u ʔilay-hɑ
then-will-direct.PASS DEF-criticisms-NOM to-it.3SGF
'If the committee makes mistake, criticisms will be directed at it.'

b. *fa-sa-tuwajjah l-ʔintiqɑːdat-u ʔilay-hɑ
then-will-direct.PASS DEF-criticisms-NOM to-it.3SGF

In English, ECs cannot modify a marked clause:[2]

(31) Whatever the committee does, (*then) criticisms will be directed at it.

But this is possible in MSA. The following illustrates for ungoverned universal ECs:

(32) [mahmɑɑ: faʕala-t l-llajnat-u]
whatever do.PAST-3SGF DEF-committee-NOM
fa-sa-tuwajjah l-ʔintiqɑːdat-u ʔilay-hɑ
then-will-direct.PASS DEF-criticisms-NOM to-it.3SGF
'Whatever the committee does, criticisms will be directed at it.'

Ungoverned alternative ECs and governed alternative ECs are the same.

(33) [ʔa-ðahab-ta ʔilɑɑ: baris ʔam lam taðhab]
(Q)-go.PAST-2SGM to Paris or not go.PRES.2SGM
fa-sa-taqdʕii: waqt-an mumtiʕ-an
then-will-have.2SGM time-ACC good-ACC
'Whether you go to Paris or not, you'll have a good time.'

(34) [sawɑːʔ-un [(ʔa)-ðahab-ta ʔilɑː baris ʔam lam taðhab]]
same-NOM Q-go.PAST-2SGM to Paris or not go.PRES.2SGM
fa-sa-taqdʕii: waqt-an mumtiʕ-an
then-will-have.2SGM time-ACC good-ACC
'No matter whether you go to Paris or not, you'll have a good time.'

Whereas both simple conditionals and ECs can follow as well as precede an unmarked clause, they can only precede a marked clause, as the following (ungrammatical versions of (30a), (32), (33) and (34)) show:

[2] Bhatt & Pancheva (2006: 4.1.1) highlight a number of situations in which *then* is unacceptable in an English conditionals, among them situations where 'the antecedent explicitly exhausts all possibilities', which they illustrate with the following (where '#' indicates unacceptability):
(i). If John is dead or alive, (# *then*) Bill will find him.
On the face it, this would exclude *then* after an exhaustive conditional. It seems that MSA *fa-* is not restricted in this way.

(35) *fa-sa-taðˤallu l-ʔintiqɑːdat-u tuwajjah ʔila
then-will-continue DEF-criticisms-NOM direct.PASS to
l-llajnat-i [ʔiðɑː ʔaχtˤaʔa-t]
DEF-committee-GEN if makes a mistake.PAST.3SGF
'Criticisms will still be directed at the committee, if it makes a mistake.'

(36) *fa-sa-tuwajjah l-ʔintiqɑːdat-u ʔila
then-will-direct.PASS DEF-criticisms-NOM to
l-llajnat-i [mahmɑɑː faʕala-t]
DEF-committee-GEN whatever do.PAST-3SGF
'Criticisms will be directed at the committee, whatever it does.'

(37) *fa-sa-taqdˤiiː waqt-an mumtiʕ-an
then-will-have.2SGM time-ACC good-ACC
[ʔa-ðahab-ta ʔilaɑː baris ʔam lam taðhab]
(Q)-go.PAST-2SGM to Paris or not go.PRES.2SGM
'You'll have a good time, whether you go to Paris or not.'

(38) *fa-sa-taqdˤiiː waqt-an mumtiʕ-an
then-will-have.2SGM time-ACC good-ACC
[sawɑːʔ-un [(ʔa)-ðahab-ta ʔilaɑː baris ʔam lam taðhab]]
same-NOM Q-go.PAST-2SGM to Paris or not go.PRES.2SGM
'You'll have a good time, no matter whether you go to Paris or not.'

3. Analyses

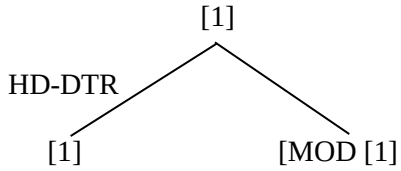
We will look first at **the distribution of ECs** and then consider their internal structure. The central fact here is that they can modify either an ordinary clause that can stand alone or a clause marked by *fa-*. The first situation is a simple matter. The second is more challenging.

Combinations of simple conditional or EC and an ordinary clause can be analysed as head-adjunct structures just like other combinations of adverbial clause and main clause. We assume the following coinstraint:

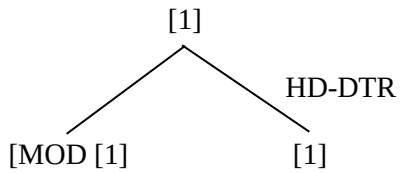
(39) *hd-adj-ph* $\Rightarrow \begin{bmatrix} \text{DTRS} < [1][\text{SS}\ [2]], [\text{HEAD}\ [\text{MOD}\ [2]]] > \\ \text{HD-DTR}\ [1] \end{bmatrix}$

Assuming that some general constraint, e.g. the Generalized Head Feature Principle of Ginzburg & Sag (2000), requires a phrase and its head to normally have the same syntactic and semantic properties, this will give structures of the following form (where the daughters may appear in either order):

(40)

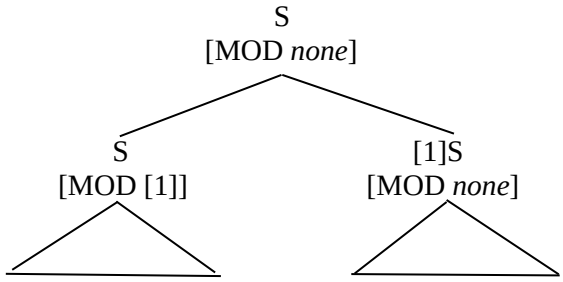


(41)



The examples in (6), (7), (9), and (15) will have structures of the following form:

(42)



The example in (27), in which the adjunct clause comes second will have a structure of this form with the order of the daughters reversed. We will propose below that governed alternative ECs introduced by *sawa:ʔ-un* are NPs. This means that the examples in (17), where EC contains *sawa:ʔ-un* will have a structure like (42) in which the modifier is an NP.

Combinations of simple conditional or EC and a clause marked by *fa-* are different. They cannot be analysed as ordinary head-adjunct structures. If they were, they would have the same SYNSEM value as the *fa*-clause, which would leave us without an explanation for the fact that such combinations are ordinary main clauses, which, unlike *fa*-clauses, can stand on their own. Like simple conditionals, they are one of a number of types of correlative clause, discussed Alqurashi & Borsley (2014), in which an adverbial clause and a main clause both have some distinctive marking and the main clause cannot appear on its own. The following illustrate:

(43) a. [bimaa ʔannka taqraʔu ʔakθar] [ʔiðann
as/since COMP.2SGM read-IMPF.2SGM more so
sa-tafhamu ʔakθar]
will-understand.IMPF.2SGM more
'As/since you read more, so you will understand more.'

b. *ʔiðann sa-tafhamu ʔakθar
so will-understand.IMPF.2SGM more

(44) a. [biqadri-maa taqraʔ] [biqadri-maa tafham]
as-much-as read-IMPF.2SGM as-much-as understand. IMPF.2SGM
'As much as you read, so much you understand.'

b. *biqadri-maa tafham
as-much-as understand. IMPF.2SGM

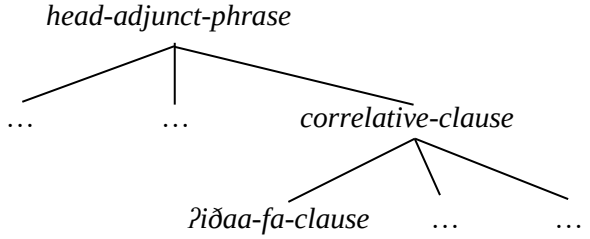
(45) a. [kullamã qaraʔta ʔakθar] [kullamã fahimta ʔakθar]
whenever read.PERF.2SGM more whenever understand.PERF.2SGM more
'Whenever you read more, you understood more.'
'The more you read, the more you understood.'

b. *kullamã fahimta ʔakθar]
whenever understand.PERF.2SGM more

The type of analysis proposed by Alqurashi & Borsley (2014) for these examples can also be applied to ECs combining with a clause marked by *fa-*.

These clause types are unproblematic if general constraints can be overridden by more specific constraints since this means a constraint can require a phrase and its head to differ in some respects. Following Alqurashi & Borsley (2014), we assume that a number of types of clause with a distinctive form have a value other than *none* for a feature CORREL, while ordinary clauses which can stand alone are [CORREL *none*].[3] We propose that there is a subtype of *head-adjunct-phrase* called *correlative-clause*, and that it has a number of subtypes, including *ʔiðaa-fa-clause*, giving the following type hierarchy:

(46)



We propose that *correlative-clause* and *ʔiðaa-fa-clause* are subject to the following constraints:

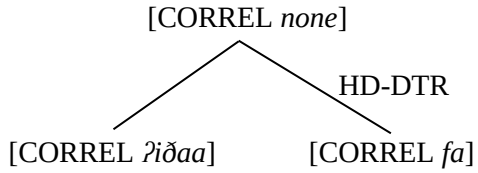
[3] Essentially the same CORREL feature was assumed in Abeillé, Borsley, Espinal 2006) and Borsley (2011). It is also assumed more recently in Abeillé & Chaves 2021: 3.3).

(47) *correlative-cl* $\Rightarrow \begin{bmatrix} \text{CORREL } none \\ \text{DTRS} < [\text{CORREL } \neg none], [\text{CORREL } \neg none] > \end{bmatrix}$

(48) *ʔiðaa-fa-cl* ⇒ [DTRS <[CORREL *fa*], [CORREL *ʔiðaa*]>]

Together they give clauses with structures with following form for of (30a), (32), (33) and (34):

(49)



If both simple conditionals and ECs are [CORREL *ʔiðaa*], they will appear in these clauses.

The analysis needs one further component: a constraint to ensure that the main clause, marked with *fa-*, comes second in correlative clauses, including *ʔiðaa-fa* clauses. The following seems appropriate:

(50) *correlative-cl* $\Rightarrow \begin{bmatrix} \text{PHON } [1] \oplus [2] \\ \text{DTRS} < [\text{PHON } [2]], [\text{PHON } [1] > \end{bmatrix}$

With this, we have a fairly simple account of the distribution of ECs, which captures their similarity to simple conditionals and some other types of clause which can appear in correlative clauses.

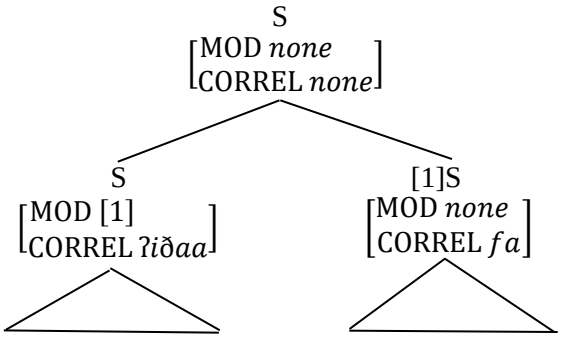
An important feature of this analysis is that [CORREL *ʔiðaa*] clauses do not always contain the lexeme *ʔiðaa*. But the following suggests that English [CORREL *if*] clauses do not always contain the lexeme *if*:

(51) Had I been there, then I would have seen you.

Thus, there is no obvious reason why clauses which do not contain *ʔiðaa* should not be [CORREL *ʔiðaa*].

Within this analysis, (32) and (33) will have structures of the following form:

(52)



(34) will have a similar structure in which the modifier is an NP.

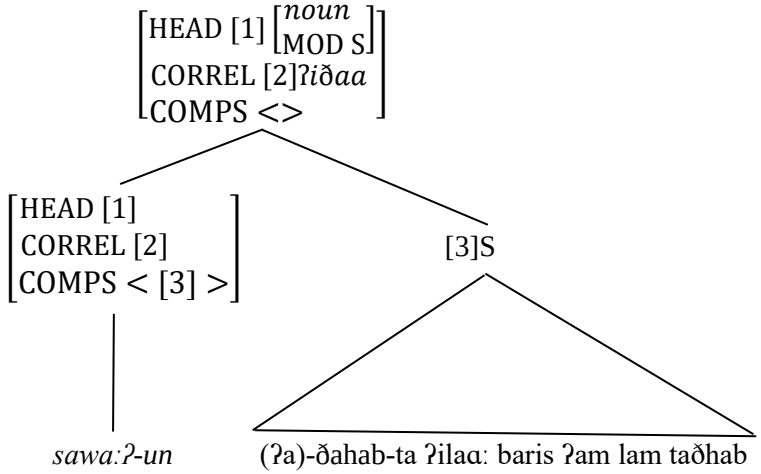
We can turn now to **the internal structure of ECs**. We will begin with **governed alternative ECs** such as those in (17). Essentially, all that is needed here is an appropriate analysis for *sawa:ʔ-un*.

Like *no matter*, as discussed in Arnold and Borsley (2014), *sawa:ʔ-un* can be analysed as a head which takes an interrogative and derives a conditional meaning from it, but unlike *no matter*, it only takes an alternative interrogative. Given the approach just proposed to the distribution of ECs, ECs and hence *sawa:ʔ-un*, must be [CORREL *ʔiðaa*]. This suggests an analysis of the following form for *sawa:ʔ-un*:

(53)
$$\begin{bmatrix} \text{SS|LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix} noun \\ \text{MOD S: [1]} \end{bmatrix} \\ \text{CORREL } ʔiðaa \end{bmatrix} \\ \text{CONT } ex\text{–}cond\ ([2], [1]) \end{bmatrix} \\ \text{ARG–ST } \left\langle \begin{bmatrix} \text{LOC} \begin{bmatrix} \text{CAT S} \\ \text{CONT [2]} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Following Arnold and Borsley (2014: 33), *ex-cond* ([2], [1]) is a condition which holds just in case [1] holds in every situation identified by [2]. Nothing here ensures that the complement is an alternative interrogative. This should probably be done with an appropriate CONT value, perhaps drawing on the analysis of Yoo (2000). With this analysis for *sawa:ʔ-un*, we will have a structure of the form in (54) for the EC in (17a) ('No matter whether you go to Paris or not').

(54)



There is no need for (52) to specify what the modified S can be. The grammar will allow either an S[CORREL *none*] in an ordinary head-adjunct clause or a *fa*-clause in an *ʔidaa-fa* clause.

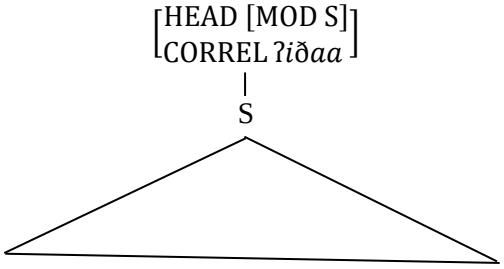
We can turn next to **ungoverned alternative ECs**. One possibility for ungoverned alternative ECs would be an analysis involving a phonologically null counterpart of *sawa:ʔ-un*. But if one shares the standard HPSG preference to avoid empty elements, the obvious alternative is a unary branching analysis in which the daughter has an interrogative meaning just like the complement of *sawa:ʔ-un* and the mother derives a conditional meaning from it in essentially the same way as *sawa:ʔ-un* does. We propose a phrase type *ungoverned-alternative-ec* subject to the following constraint:

(55) *ungoverned-alternative-ec* ⇒

$$\begin{bmatrix} \text{SS|LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD [MOD S: [1]]} \\ \text{CORREL } ʔiðaa \end{bmatrix} \\ \text{CONT } ex\text{–}cond\ ([2],[1]) \end{bmatrix} \\ \text{DTRS} \left\langle \begin{bmatrix} \text{LOC} \begin{bmatrix} \text{CAT S} \\ \text{CONT [2]} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

As with the complement in (45), it needs to be specified that the daughter is an alternative interrogative, probably with an appropriate CONT value. This will give a structure of the following form for the EC in (15a) (‘whether you go to Paris or not’):

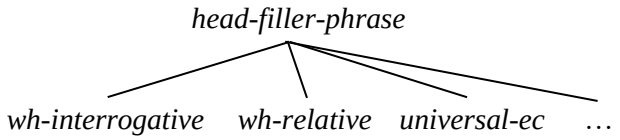
(56)



Finally, we can consider **ungoverned universal ECs** (which, as we have noted, are the only type of universal EC). We have seen that they involve head-filler phrases in which the filler contains one of a small number of EC words. If they were identical to *wh*-interrogatives like their English counterparts, it would be reasonable to propose a unary branching analysis like that we have proposed for ungoverned alternative ECs.[4] It is clear that they are not *wh*-interrogatives, but the analysis of *wh*-interrogatives is still of some relevance. As we noted above, they are like *wh*-interrogatives in identifying a set of possible situations, but unlike *wh*-interrogatives in indicating that all the situations are ones in which the modified clause is true.

We propose that ungoverned universal ECs involve a special subtype of *head-filler-phrase*, which we will call the subtype *universal-ec*. In other words, we assume a type hierarchy of the following form:

(57)



This subtype needs to have the following properties:

- It has a filler with one of a small number of EC words.
- It modifies a clause.
- It is [CORREL *ʔidaa*].
- It has conditional semantics.

We attribute these properties to the following constraint:

[4] However, not all English *wh*-interrogatives can be ECs. *Whatever he did* and *whether he did it or not* can be ECs, but *what he did* and *whether he did it* can't. Arnold & Borsley (2014: 34) note this, but do not discuss how overgeneration could be avoided.

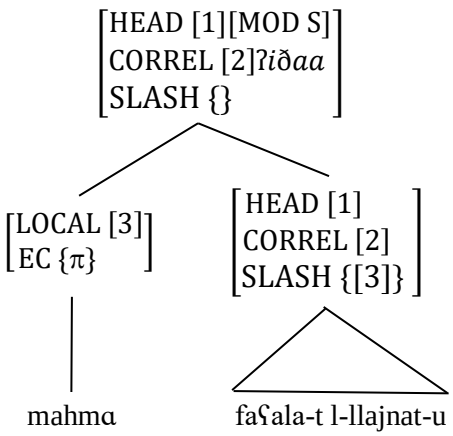
(58) *universal-ec* ⇒

$$\begin{bmatrix} \text{SS|LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD [MOD S: [1]]} \\ \text{CORREL } \textit{ʔidaa} \end{bmatrix} \\ \text{CONT } \textit{ex–cond } (\lambda\{\pi, \ldots\}[\lambda\mathbf{X}[\mathbf{Y}](\mathbf{Z})], [1]) \end{bmatrix} \\ \text{DTRS } \langle \begin{bmatrix} \text{EC } \{\pi\} \\ \text{CONT } \mathbf{Z} \end{bmatrix}, \begin{bmatrix} \text{SLASH } \{[\text{CONT } \mathbf{X}]\} \\ \text{CONT } \mathbf{Y} \end{bmatrix} \rangle \end{bmatrix}$$

We have an EC feature here where *wh*-interrogatives have WH, and we assume that its value, like that of WH, is a set containing a single parameter, a combination of an index and a restriction. Building on Sag's (2010: 5.3) analysis of *wh*-interrogatives, we assume that the semantics involves a propositional abstract constructed from the semantics of the daughters, but unlike with *wh*-interrogatives, this is the first argument of *ex-cond*, and the modified clause is the second argument as before. This is also somewhat like Sag's (2010: 5.4) analysis of *wh*-relatives, in which a modifying semantics is based on a clausal semantics.

With this analysis, we will have a structure of the following form for the EC in (6):

(59)



With this we have an analysis of all three types of EC.

4. Concluding remarks

In this paper, we have outlined analyses of both the distribution of MSA ECs and the internal structure and interpretation of the three different types. We have argued for the following positions:

- Like simple conditionals, ECs can be the adjunct in both an ordinary head-adjunct clause and a *ʔiðaa-fa* subtype of correlative clause.

- Governed alternative ECs are head-complement phrases, in which a head takes an alternative interrogative as its complement and derives a conditional meaning from it.

- Ungoverned alternative ECs have a unary branching analysis in which the daughter is an alternative interrogative and the mother derives a conditional meaning from it.

- Ungoverned universal ECs involve a subtype of head-filler phrase, which derives a conditional meaning from its daughters.

We have developed three rather different analyses here for the three types of ECs that occur in MSA. We think this is justified by their rather different properties. However, the analyses share certain features reflecting the shared properties. All have a MOD value allowing them to modify a clause, all have the [CORREL *ʔidaa*] specification allowing them to be the adjunct daughter in an either an ordinary head-adjunct clause or an *ʔidaa-fa* clause, and all have conditional semantics, based on EX-COND. We think, then, that the analyses capture both the differences and the similarities among MSA ECs.

Naturally, there are further issues that merit discussion here. We have emphasized similarities between ECs and simple conditionals, which are greater in MSA than English. There are, however, important differences in MSA, as in Engish. One involves *faqatˁ* 'only'. This can be added to a simple conditional, as the following illustrate:

(60) a. saʔaraː-ka ʔiðaː kunta fi bariːs ʔaw
will-see.1SG.F/M-2SGM if was.2SGM in Paris or
ruːma
Rome
'I will see you if you are in Paris or Rome.'

b. saʔaraː-ka faqatˁ ʔiðaː kunta fi bariːs
will-see.1SG.F/M-2SGM only if was.2SGM in Paris
ʔaw ruːma
or Rome
'I will see you only if you are in Paris or Rome.'

But *faqatˁ* cannot be added to an EC. Hence, only (a) is acceptable in the following:

(61) a. saʔaraː-ka kunta fi bariːs ʔaw ruːma
will-see.1SG.F/M-2SGM was.2SGM in Paris or Rome
'I will see you whether you are in Paris or Rome.'

b. *saʔaraː-ka faqatˤ kunta fi bariːs ʔaw ruːma
will-see.1SG.F/M-2SGM only was.2SGM in Paris or Rome
'*I will see you only whether you are in Paris or Rome.'

There is nothing in our analysis that suggests there should not be differences as well as similarities between ECs and simple conditionals. As the translations indicate, the same contrast is found in English, and we assume that it will be explained in the same way in the two languages.

There is at least one important limitation of our discussion. We have focused throughout on ECs modifying declarative clauses. But in MSA, as in English, both simple conditionals and ECs can also modify interrogatives and imperatives, as the following show:

(62) a. [ʔiðaː ðahab-ta hunaːk]maðaː sa-tafʕal?
if go.PAST.2SGM there, what will-do.2SG.M?
'If you go there, what will you do?'

b. [ʔiðaː ðahab-ta hunaːk] sallim ʕalay-him
if go.PAST.2SGM there greet.2SGM to-him.
'If you go there, greet him.'

(63) a. [ʔaynmaː ðahab-ta] hal sa-taʕuːdu sariːʕ-an?
wherever go.PAST.2SGM Q will-back quickly-ACC
'Wherever you go, will you come back quickly?'

b. [ʔaynmaː ðahab-ta] badir bi-tˤrħ-I
wherever go.PAST.2SGM initiate.2SGM with-ask-GEN
l-ʔsʔilat-i
DEF-questions-GEN
'Wherever you go, ask questions.'

The approach to conditionals assumed here seems to deal well with declaratives, but on the face of it, it needs to be revised or extended in some way to accommodate interrogatives and imperatives. But this is an issue that is not specific to MSA. The facts are essentially the same in English, and no doubt other languages. We assume, therefore, that whatever approach seems appropriate elsewhere could be extended to MSA.[5]

[5] For some discussion, see e,g. Isaacs & Rawlins (2008) and Kaufmann & Schwager (2009).

REFERENCES

Abeillé, Anne, Robert D. Borsley & Maria-Teresa Espinal. 2006. The syntax of comparative correlatives in French and Spanish. In Stefan Müller (ed.), *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, Varna, 6–26. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2006.1.

Abeillé, Anne & Rui P. Chaves. 2021. Coordination. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-driven Phrase Structure Grammar: The Handbook*, 725–776. Berlin: Language Science Press.

Alqurashi, Abdulrahman. 2012. An HPSG approach to free relatives in Arabic. In Stefan Müller (ed.), *Proceedings of the 19th International Conference on Head-driven Phrase Structure Grammar*, 6–25. Stanford, CA: CSLI Publications.

Alqurashi, Abdulrahman & Robert D. Borsley. 2014. The comparative correlative construction in Modern Standard Arabic. In Stefan Müller (ed.), *Proceedings of the HPSG14 Conference*, 6–26. Stanford, CA: CSLI Publications.

Arnold, Doug & Robert D. Borsley. 2014. On the analysis of English exhaustive conditionals. In Stefan Müller (ed.), *Proceedings of the HPSG14 Conference*, 27–47. Stanford, CA: CSLI Publications.

Bhatt, Rajesh and Roumyana Pancheva. 2006. Conditionals. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, Vol. 1, 638–687. Malden, MA: Blackwell.

Borsley, Robert D. 2011. Constructions, functional heads and comparative Correlatives. In Olivier Bonami & Patricia Cabredo Hofherr (eds.), *Empirical Issues in Syntax and Semantics* 8, 7–26. http://www.cssp.cnrs.fr/eiss8

Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*. Stanford, CA: CSLI Publications.

Haspelmath, Martin & Ekkehard König. 1998. Concessive conditionals in the languages of Europe. In Johan van der Auwera (ed.), *Adverbial Constructions in the Languages of Europe*, 563–640. Berlin: Mouton de Gruyter.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Isaacs, James & Kyle Rawlins. 2008. Conditional Questions. *Journal of Semantics* 25: 269–319

Kaufmann, Stefan & Magdalena Schwager. 2009. A unified analysis of conditional imperatives *Semantics and Linguistic Theory* 19, 239–256.

Rawlins, Kyle. 2008. *An Investigation in the Syntax and Semantics of Conditional Structures*. PhD dissertation, University of California, Santa Cruz.
Rawlins, Kyle. 2013. (Un)conditionals. *Natural Language Semantics* 21(2), 111–178.
Sag, Ivan A. 2010. English filler–gap constructions. *Language* 86, 486–545.
Yoo, Eun-Jung. 2000. A constraint-based approach to alternative questions in English. *Language Research* 36, 19–42.

Between shared expectations and possibilites: The discourse particle *sempre* in European Portuguese

João Azevedo
Universidade de Lisboa

Jakob Maché
Centro de Linguística da Universidade de Lisboa

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 47–59

Keywords: keywords

Azevedo, João & Jakob Maché. 2024. Between shared expectations and possibilites: The discourse particle *sempre* in European Portuguese. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 47–59. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.3.
CC-BY 4.0

Abstract

This paper examines the syntactic and semantic properties of the confirmative use of *sempre* in European Portuguese. Unlike its temporal counterpart which carries the meaning of 'always', confirmative *sempre* is restricted to the pre-verbal position, disqualifying it as a prototypical adverb. In terms of its semantic contribution, the confirmative discourse particle *sempre* marks the proposition as given by suspending the stress on phonological constituents in the clause. Consequently, the nuclear stress falls on *sempre*. Contrary to the analysis proposed by Amaral & Del Prete's (2014), givenness is not equated with being part of the shared knowledge/common ground. Instead, there could have been disagreement about the validity of the embedded proposition between the speaker and the addressee at some earlier stage. Givenness, therefore, will be understood as being part of at least one party's discourse commitments, following the framework of Farkas & Bruce (2010).

1 Introduction

This paper addresses the question of whether there are genuine 'modal' particles in Portuguese. Although there have been proposals slightly fewer than twenty candidates over the past thirty years (cf. Franco 1990: 175, 1998: 147–149, Macário-Lopes 1998: 8–10, Meisnitzer 2012: 344–353, Pinto de Lima 1997, Mendes & Lejeune 2022), these candidates lack essential syntactic properties. The present study focuses on the items *afinal*, *cá*, *lá*, *sempre* and *nem*, (i) all of which are restricted to the position preceding the finite verb and (ii) all of which make reference to either previously shared expectations attributed to the speaker, the addressee, or some third party.

To date, it remains contested whether these items qualify as belonging to the syntactic category *particle* and to what extent they exhibit 'modal' semantics. This debate largely stems from terminological confusion inherited from the early pioneering work on German modal particles by Weydt (1969). Already Thurmair (1989: 3) observed that, in previous studies on these particles in German, the term 'modal' only is used to express that these markers convey extra-propositional meaning and therefore does not contribute much in the characterisation of these elements. As a consequence, the most recent literature on German replaced the misleading term *modal* by the descriptively more adequate term *discourse*, as in the handbook articles by

[†]First, we would like to express my gratitude to the three anonymous reviewers during the initial reviewing process and the fourth one in the final reviewing process for their comments and suggestions. They helped to significally improve the quality of the present paper. Furthermore, we are indebted to Marco Coniglio, Jonathan Ginzburg, Elena Karagjosova, Andy Lücking, António Machicao y Priemer, Ana Maria Martins, Rui Marques, Amalia Mendes, Benjamin Meisnitzer, Eva Remberger, Oliver Schallert and Giuseppe Varaschin for ther comments and support before and after the presentation.

Zimmermann (2011) and Grosz (2020). Thus the more appropriate term *discourse particle* is used in this paper whenever these items do not clearly reflect the basic characteristics of modal semantics.

The paper aims to bring clarity to the discussion by drawing on corpus data and on recent findings on discourse semantics. It concludes that at least the confirmative use of *sempre*, which is the main focus of this study, exhibits all the relevant characteristics comparable to its German counterparts. Furthermore, regarding its semantics, mirativity and the dimension of discourse are much more relevant for its description than is the dimension modality in the narrow sense. The dimension discourse is understood here as referencing assertions or other types of commitments, beliefs or expectations that were previously shared between the speaker and the addressee. Specifically, the particle *sempre* makes reference to previous *Discourse Commitments* and/or *Common Ground*, in the sense of Farkas & Bruce (2010: 84–90). Finally, a definition of the syntactic category of discourse particles will be suggested, which applies at least to Germanic and Romance languages.

2 Syntactic properties of *sempre*, *nem*, *lá*, *cá* and *afinal*

Adopting the more developed insights from decades of research on German modal particles as discussed by Thurmair (1989: 25–29, 36) and Coniglio (2008: 14–16, 121–126, 148, 159), this paper shows that the term *particle* is essentially a syntactically motivated category. Specificially, it refers to a type of defective sentence adverbial, which has almost completely lost its capacity for displacement within a clause and no longer is part of a question domain of any question pronoun, in the sense of Ginzburg (2012: 122), and often not of a focus domain either. The question domain of a question pronoun like *who*, *what* or *why* is the set of constituents which are possible answers. For instance, the question domain of *who* contains all NPs that denote referents with the feature [+HUMAN]. Similarly, the question domain of *why* encompasses all the PPs and adverbials, or adverbial clauses that can denote a cause of an event or proposition. Likewise, the focus domain contains all the possible alternatives by which a focused item could be replaced. It should be noted, however, that confirmative *sempre* in declarative clauses obligatorily attracts nuclear stress, just as its german congate stressed *DOCH* which acts as an exponent of verum (focus), as shown by Gutzmann (2010) and Egg & Zimmermann (2012: 230–233).

Before examining the corpus data, it is important to note that in Indo-European languages most of the discourse particles represent just one single use among several uses of polyfunctional or heteronymous lexemes, as illustrated by Thurmair (1989: 21), Helbig 1994, and Coniglio (2008: 8–9). In European Portuguese, *sempre* is primarly used as a temporal adverb that

universally quantifies over time intervals, similar to its English counterpart *always*. In that usage, it preferably occurs in the post-verbal position as demonstrated in example (1).

(1) O banho de imersão é$_V$ sempre demorado.[1]
DET.M bath of immersion is always take.time-PPP
'The full bath is always time consuming.'

(2) Vocês ontem sempre foram$_V$ ao cinema?[2]
you.PL yesterday after.all were at.the.M cinema
'Did you go to the cinema after all?'

In contrast, the more gramaticalised discourse particle use of *sempre*, also referred to as confirmative *sempre*, is restricted to the preverbal position as illustrated in example (2). Its semantics is more intricate. Confirmative *sempre* in example (2) requires there to be a previously shared commitment of the addressee to go the cinema, which was later doubted by the speaker.

In its limitation to the preverbal position, confirmative *sempre* fulfils the main criterion for particles introduced above. Apart from *sempre* there are many other core candidates for discourse particles competing for the same slot (cf. Franco 1990: 175, 1998: 147, 150; Macário-Lopes 1998: 7, Brito 2001: 66, Ambar et al. 2004: 2–5, Fiéis 2010 and Amaral & Del Prete 2014: 137). Amaral & Del Prete (2014: 137) and Ambar et al. (2004: 3) even attempt to correlate the semantic interpretation of the modifier *sempre* with its syntactic position, claiming that the confirmative use of *sempre* is only available in the preverbal position whereas the temporal use is rejected in the preverbal positions by many native speakers. However, in the corpora investigated here, there are numerous instances of temporal *sempre* in preverbal position, falsifying the second part of the authors' claim.

Additionally, Macário-Lopes (1998: 9) and Amaral & Del Prete (2014: 146–147) argue that confirmative *sempre* cannot co-occur with negation, instead another particle *afinal* has to be chosen. In contrast, Franco (1998: 148) discusses an example where the negation *não* appears within the scope of confirmative *sempre*, though he notes that negation can never take scope over confirmative *sempre*.

We conducted two independent corpus studies, based on data from the *DiLeB* corpus (*Discurso Informal de Lisboa e Braga*) and the *CRPC* corpus (*Reference Corpus of Contemporary Portuguese*).[3]

[1] *CRPC-ORAL* pf1202pu.txt.

[2] *CRPC-ORAL* pfamcv06.txt.

[3] The *DiLeB*-corpus (*Discurso Informal de Lisboa e Braga*) is an online corpus of informal conversations with speakers from Lisbon and Braga in sociolinguistic interviews from the last decade of the 20th century.

http://teitok.clul.ul.pt/dileb/index.php?action=home

The *CRPC* corpus (*CRPC-ORAL* (*Reference Corpus of Contemporary Portuguese*)

The frequencies are represented in Tables 1 and 2. It should be noted that one defining feature of discourse particles is that they often represent just one usage of polyfunctional or heteronym lexemes, which are notoriously difficult to distinguish (Thurmair 1989: 21, Helbig 1994, Coniglio 2008: 8–9). In most cases, there are also less grammaticalised, more lexical adverbial uses. For instance, Portuguese *sempre* primarly functions as a temporal adverb meaning 'always', which universally quantifies over time intervals.

The figures presented are simple counts of the lexemes without differentiating the particle uses. However, the discourse-oriented uses of these adverbs are rare, and almost exclusively in the preverbal position. The situation with *afinal* is less clear, as it always conveys discourse meaning but it can occur post-verbally or clause-initially ,and unlike *sempre*, it can appear with varying placements of the nuclear stress (cf. Amaral & Del Prete 2014: 141 for a similar observation for its Italian counterpart *alla fine*). In the *DiLeB* corpus, *afinal* has a strong preference for the preverbal position, in the *CRPC* corpus, however, it is attested mostly in clause initial position. Similarly, the discourse-oriented uses of *nem* have a strong preference for the preverbal position.

position	*nem*	*sempre*	*afinal*	*lá*	*cá*
preverbal	181	177	13	505	132
postverbal	99	503	2	1018	259
other pos.	283	109	0	345	115
total	563	789	15	1868	506

Table 1: The placement of discourse modifiers and their polyfunctional variants – *CRPC* corpus

position	*nem*	*sempre*	*afinal*	*lá*	*cá*
preverbal	395	288	7	1232	133
postverbal	152	1009	5	2358	378
other pos.	523	297	25	867	334
total	1070	1594	37	4457	845

Table 2: The placement of discourse modifiers and their polyfunctional uses –*DiLeB* corpus

Being limited to the preverbal position, confirmative *sempre* behaves

is another online corpus composed mostly of spoken European Portuguese from various contexts and periods of recent time.

http://teitok.clul.ul.pt/crpcoral/index.php?action=home

similarly to other recognised modal particles in other Romance, such as Italian *mai* and particles in Germanic languages including German and Scandinavian (cf. Coniglio 2008: 14–16, 95–101, Coniglio 2023: 11–15). In other words, these particles defined by their placement restrictions to some position at the left edge of the extended VP or TP—corresponding to the the preverbal position in Romance languages, and some position in the middle field in German.[4]

It is quite plausible to assume that the extremely limited freedom of movement for these particles arises from more general requirements. As observed by Hentschel (1983: 48, 1986: 210–213, 232–238) Thurmair (1989: 25–37) and Coniglio (2008: 102–108), the dimensions of givenness and definiteness play essential roles in the placement of discourse particles in German. These authors argue that discourse particles mark the boundary between the theme (given information) and rheme (discourse new information). Their observations suggest that the nuclear stress must follow the particles, and that it may be the focus—rather than the rheme—that comes after the particle.

Despite Franco's (1998: 144, 150) observation that discourse particles in Portuguese do not mark the left boundary of the constituent of the rheme, the Portuguese confirmative particle *sempre* is also sensitive to focus. Amaral & Del Prete (2014: 139–140, 148–149) argue that the confirmative use of *sempre* requires the entire prejacent proposition to be destressed, while *sempre* itself attracts nuclear stress. Thus, it behaves semantically very similarly to the stressed version of the German particle *doch* (cf. Egg & Zimmermann 2012), as already noticed by Franco (1998: 153).

3 Modal or discourse oriented – the semantics of *sempre*

Thurmair (1989: 3) noted that previous studies on particles used the term 'modal' loosely to indicate that these markers convey extra-propositional meaning, following Palmer's (1986: 1) very vague definition of modality, which posits that modal modifiers encompass any type of modifier that take scope over the proposition. As a result, this term contributed little to the precise characterisation of these elements. In subsequent research, Portner

[4]However, there are no discrete boundaries between discourse particles and sentential adverbs as regards to their semantics. Confirmative *sempre* in Italian is fairly common in postverbal position, which is only possible for adverbs under the definition defended here. But at the same time, Italian *sempre* lacks past related interpretations in as demonstrated by Amaral & Del Prete (2014: 135, 137, 140–149), which is a clear indicator that it is grammaticalised to a lesser degree. Apart from that, the Italian marker is less frequently used than its Portuguese cognate. In a similar vein, Portuguese *afinal* and its Italian counterpart *alla fine* occur in positions typical for adverbs but nevertheless these itmes display a meaning related to *sempre*, indicating an epistemic change or conflict in the discourse.

(2009: 1) offered a clearer definition of modality describing it as making statements about situations in non-actual possibilities/possible worlds. His work is heavily inspired by earlier work by Kratzer (1978, 1981), who focused only on selected modal verbs in German and English but did not explicitly defne of modality as a phenomenon itself. However, among the 20–35 markers considered as modal particles by Thurmair (1989: 49) and Durell (2011) only a few, such as the epistemic particles *wohl* refer to non-actual possibilities. Most of these elements instead refer to shared or individual beliefs, convictions or statements by discourse participant or third participants, some of them but not all also refer to events in non-actual worlds.

The dimension of discourse seems more relevant for many of these items, as they reference propositions to which at least one speech participant is publicly committed too. Macário-Lopes (1998: 8–9) observed that the confirmative particle *sempre* expresses the speaker's expectations and doubts regarding the truth of the prejacent propostion p. Amaral & Del Prete (2014: 135–140, 2016: 1135–1137, 2020: 5–7) propose a more specific and detailed description of the semantic contribution of confirmative *sempre*. They argue that it is "only felicitous in a context where the truth of the prejacent is presupposed to have been under discussion by the interlocutors" (p. 140).

In their analysis, Amaral & Del Prete (2014: 149–150) take the particle *sempre* to be an epistemic modal operator that takes a proposition p and returns the confirmation of the truth of that proposition, and referencing three different temporal points: an initial point t_1 prior to utterance time when the prejacent p_4 was considered true in all the best epistemic worlds, a subsequent point t_2 when it became possible that p could be false, and utterance time t_0 when p is confirmed to be true. Crucially, they align with Kratzer's (1978, 1981), view that epistemic modal operators are evaluated with respect to collective knowledge, rather than individual knowledge.

While Amaral & Del Prete's (2014: 149–150) analysis captures many relevant aspects of *sempre*, it has at least two essential short comings. Firstly, it fails to distinguish between propositions to which both speech participants commit (hence part of the common ground) and propositions to which are only known to the speech participants, in the sense of that the speech participants know that one of them has publically committed to them without that the other speech participants shares that commitment. In the example (3), it is possible that the speaker never believed the proposition p that CHEGA would win more than 15% but addressee always insisted that p would be true. In such a scenario p was never part of the common ground, of the shared knowledge between speaker and addressee. However, what was mutually shared knowledge is that the addressee wanted to add p to the common ground. So there are two different senses for p of being 'known'.

The most common case are propositions to which only one speech participant committed publicly by asserting it, but whose truth is contested by the other party. This conflict can be resolved assuming that there is another

resource in the discourse representation besides the common ground, which represents shared beliefs, namely *discourse commitments*, in the sense of Gunlogson (2001: 146–150) and Farkas & Bruce (2010). Discourse commitments are propositions to which only one of the party is unilaterally committed. The importance of commitments is increasingly being acknowledged in current research, Geurts (2019) proposes an entire theory of speech acts and discourse representation grounded on commitments.

The second shortcoming concerns the conception of epistemic modality. Lasersohn (2005: 277), Stephenson (2007: 489) and Maché (2019: 517–540) observed that epistemic modal operators are not evaluated with respect to what is "generally known in the world", as suggested in Kratzer's early work but their interpretation is always dependent on a specific epistemic judge.

Amaral & Del Prete (2014: 149–150) encounter difficulties in explaining the most common occurrences of *sempre*, especially when there is a disagreement at t_2 between the speaker and addressee or third party about the validity of the proposition p. For the sake of clarity, the different roles will be defined as follows: the *endorser* refers to the participant who initially endorses the proposition p, and as the *questioner*, the participant who later challenges the validity of p. A scenario where *sempre* signals disagreement between a endorser and a questioner is illustrated in example (3). The endorser of the expectation p can be explicitly encoded by phrases like 'you were right' and as shown below, speaker and addressee can have different convictions:

(3) Tinha/Tinhas/A minha vizinha tinha
have.IMP.1S/have.IMPF.2S/DET.F my.F neighbour have.PST.3S
razão, o CHEGA sempre ganhou mais de 15%.
right DET.M Chega sempre win.PST.3S more than 15%
'I was/You were/My neighbour was right, the CHEGA party won more than 15% after all.'

The proposition 'CHEGA wins more than 15%' in the example above was never part of the *Common Ground*, because it was never a belief shared between all the parties involved. A similar test can be applied to identify the questioner, who holds the belief at t_2 that p could be false. The follow up *T'as a ver?* 'Do you see now?' presupposes that the addressee has not committed to p, and thus cannot be the endorser, but must have been the questioner in previous discourse. In other words, the roles of the endorser and the questioner can be assigned to either speaker and addressee or vice versa, depending on the context. But both roles may be performed by the speaker simultaneously, indicating a double change of their epistemic state at t_2 and at utterance time.

4 Analysis

The analysis presented here is implemented in HPSG_{TTR}/KoS (based on *Type Theory with Records*) as suggested by Ginzburg (2012) and more recently in the HPSG handbook in Lücking et al. (2021). This particular version of HPSG includes semantics that is specialised for representing dialogues with their intricate semantic relations. The discourse particles under discussion make reference to discourse commitments or their suspension that have been previously shared between the speaker and the addressee. Given the large body research within HPSG_{TTR}, which specialises in the treatment of the semantic relations within dialogue moves, it is more appropriate than alternative implementations of semantics within the framework of HPSG. As will be shown below, a dialogue game board including the feature MOVE is incredibly helpful for tracking discourse commitments previously made by the speaker or the addressee.

Confirmative *sempre* is treated as a defective sentential adverb with the denotation of a propositional modifier of the type $\langle t, t \rangle$. It takes a proposition, returns the same proposition and adds the requirement that two specific previous moves must have had occured for its truth conditions to be met: (i) A move m_1, a public commitment to the truth of p by some speech participant (endorser) x and (ii) another subsequent move m_2, a commitment by the same or another speech participant (questioner) y that p is/or may no longer be valid. The semantic contribution is summarised as follows. Confirmative *sempre* makes references three times: t_1, which precedes t_2, which in turn precedes t_{utt}. There is an epistemic attitude holder, the endorser x, who publicly committed to the validity of p at t_1 or that it would become true in future. At some subsequent moment t_2, the questioner y commits to the possibility that p could be false. This is expressed by means of a modal operator anchored to some attitude holder or modal judge, the questioner y, in the sense of Stephenson (2007: 501).

In cases where the endorser x and the questioner y are identical, the sudden commitment m_2 to the possibility that p is no longer valid, entails a retraction of their previous commitment m_1, and m_2 will be interpreted as a concession. This strict separation between the two roles is necessary to model scenarious in which there is an epistemic disagreement between speaker and addressee, one insisting in the truth of p, the other in the truth of $\neg p$. Importantly, both the expectation and the concession that p might be false are prior dialogue moves. Eventually the speaker asserts p. Remember that the speaker can be identical to x or y or both of them. This is ensured by the two auxiliary clauses c_1 and c_2.

This analysis parallels the question bias with low negative polar questions with low negation reading, as analysed by Sudo (2013: 276–284). Speakers of such questions express an initial belief that p was true (epistemic bias) but have encountered compelling evidence suggesting that p is false (evidential

bias) shortly before utterance time. Turning to confirmative *sempre*, the shift in epistemic state is modeled by introducing an epistemic possibility at t_2, without detailing the specific evidence behind this change.

Implemented in HPSG_{TTR}, the lexicon entry for confirmative *sempre* is illustrated in Figure 1. The assumption is that in a previous move m_1, a speech participants x–who could be either the speaker or the addressee—publicly committed to the validity of p by asserting it. In a subsequent move m_2, the same or a different speech participant raises doubts about the validity of p into question committing to the possibility that $\neg p$. If the endorser and the questioner are the same speech participant ($y = x$), the previous commitment m_1 will be retracted by conceding the possibility that p may not be valid. However if y is instantiated by another speech participant, the retraction of the initial commitment m_1 is unnecessary. As the expectation that p will remain valid is modeled as discourse commitment by means of some move made earlier in the dialogue, the analysis here is compatible with p being part of the *Common Ground* in some scenarios.

The move m_2 contains a possibility operator following Cooper's (2023: 247) adaptation of Kratzer's (1981) concept of modal operators into TTR. In the present study, the operator is a predicate of the arity $\langle$*Ind*, *Type*, *Type*, *Type*$\rangle$, taking as arguments: an epistemic judge (here modeled as *Individual*), a proposition (here modeled as *Type*), an epistemic modal base B (here modeled as *Type*) and a stereotypical ordering source of ideals I (likewise modeled as *Type*).

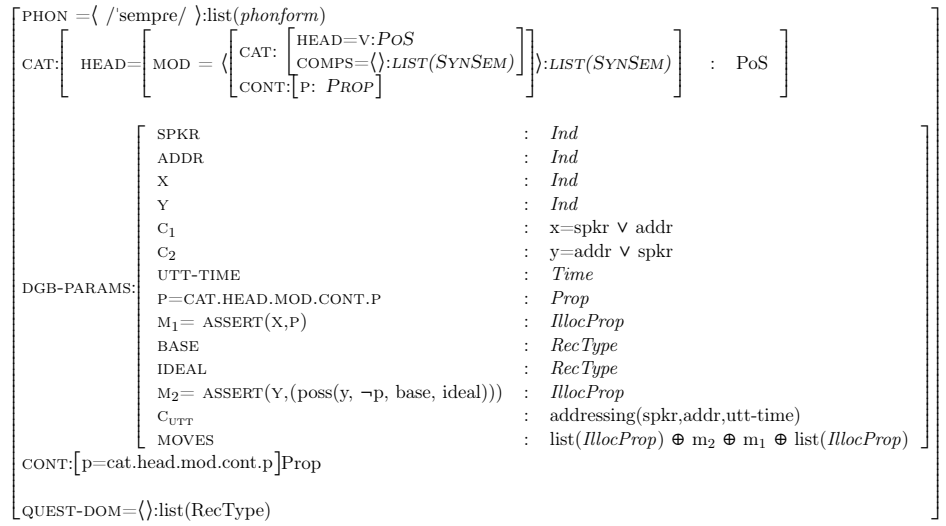


Figure 1: Lexicon entry for the confirmative discourse particle *sempre*

In summary, confirmative *sempre* resembles stressed *doch* in German, which references to a previously negated proposition that was earlier present in the common ground, as illustrated by Karagjosova (2009), Egg & Zim-

mermann (2012: 227–228) and Döring (2016). Returning to the question, whether European Portuguese has items that deserve to be considered as discourse particles of the Germanic type, the answer is clearly yes.

References

Amaral, Patrícia & Fabio Del Prete. 2014. On truth persistence: A comparison between European Portuguese and Italian in relation to *sempre*. In Marie-Hélène Côté & Eric Mathieu (eds.), *Variation within and across Romance languages. Selected papers from the 41st Linguistic Symposium on Romance Languages* (Current Issues in Linguistic Theory 333), 135–154. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/cilt.333.11ama.

Amaral, Patrícia & Fabio Del Prete. 2016. On truth unpersistence: At the crossroads of epistemic modality and discourse. *Natural Language & Linguistic Theory* 34(4). 1135–1165. DOI: 10.1007/s11049-015-9325-5.

Amaral, Patrícia & Fabio Del Prete. 2020. Predicting the end: Epistemic change in Romance. *Semantics and Pragmatics* 13(11). 1–50. DOI: 10.3765/sp.13.11.

Ambar, Manuela, Manuela Gonzaga & Esmeralda Negrão. 2004. Tense, quantification and clause structure in European Portuguese and Brazilian Portuguese: Evidence from a comparative study on *sempre*. In Reineke Bok-Bennema, Bart Hollebrandse, Brigitte Kampers-Manhe & Petra Sleeman (eds.), *Romance Languages and Linguistic Theory 2002. Selected papers from* Going Romance*, Groningen 28–30 November 2002* (Current Issues in Linguistic Theory 256), 1–16. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/cilt.256.

Brito, Ana Maria. 2001. Clause structure, subject positions and verb movement: About the position of *sempre* in European Portuguese and Brazilian Portuguese. In Yves d'Hulst, Johan Rooryck & Jan Schroten (eds.), *Romance Languages and Linguistic Theory 1999*, 63–85. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/cilt.221.

Coniglio, Marco. 2008. *Die Syntax der deutschen Modalpartikeln: Ihre Distribution und Lizenzierung in Haupt und Nebensätzen*. Università Ca' Foscari Venezia. (Doctoral dissertation).

Coniglio, Marco. 2023. *On a syntactic definition of modal particles: Synchronic and diachronic reflections*. Presentation given at the Workshop *Modal particles in Romance languages* at the Universidade de Lisboa 1st–2nd June 2023.

Cooper, Robin. 2023. *From perception to communication: A type theory for action and meaning* (Oxford Studies in Semantics and Pragmatics 16). Oxford University Press. DOI: 10.1093/oso/9780192871312.001.0001.

Döring, Sophia. 2016. *Modal particles, discourse structure and common ground management: Theoretical and empirical aspects*. Berlin: Humboldt Universität zu Berlin. (Doctoral dissertation). DOI: 10.18452/19449.

Durell, Martin. 2011. *Hammer's German grammar and usage*. 5th edn. London: Routledge.

Egg, Markus & Malte Zimmermann. 2012. Stressed out! Accented discourse particles: the case of *doch*. In Ana Aguilar Guevara, Anna Chernilovskaya & Rick Nouwen (eds.), *Proceedings of Sinn und Bedeutung 16*, 225–238. Konstanz: Universität Konstanz. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/422/355.

Farkas, D. F. & K. B. Bruce. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27(1). 81–118. DOI: 10.1093/jos/ffp010.

Fiéis, Alexandra. 2010. On the position of *sempre* in Medieval Portuguese and in Modern European Portuguese. *The Linguistic Review* 27(1). 75–105. DOI: 10.1515/tlir.2010.004.

Franco, António. 1990. Partículas modais do Português. *Revista da Faculdade de Letras : Línguas e Literaturas* 7. 176–196.

Franco, António. 1998. Partículas modais da língua Portuguesa: Relances contrastivos com as partículas alemãs. *Revista da Faculdade de Letras : Línguas e Literaturas* 5(1). 137–156.

Geurts, Bart. 2019. Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics* 45(1-2). 1–30. DOI: 10.1515/tl-2019-0001.

Ginzburg, Jonathan. 2012. *The interactive stance: Meaning for conversation*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199697922.001.0001.

Grosz, Patrick G. 2020. Discourse particles. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann & Thomas Ede Zimmermann (eds.), *The companion to semantics*, 1–34. Oxford: Wiley. DOI: 10.1002/9781118788516.sem047.

Gunlogson, Christine. 2001. *True to form: Rising and falling declaratives as questions in English*. Santa Cruz: University of California. (Doctoral dissertation). http://hdl.handle.net/1802/231.

Gutzmann, Daniel. 2010. Betonte Modalpartikel und Verumfokus. In Elke Hentschel & Theo Harden (eds.), *40 Jahre Partikelforschung*, 119–138. Tübingen: Stauffenburg.

Helbig, Gerhard. 1994. *Lexikon deutscher Partikel*. Leipzig: Langenscheidt.

Hentschel, Elke. 1983. Partikeln und Wortstellung. In Harald Weydt (ed.), *Partikeln und Interaktion* (Reihe Germanistische Linguistik 44), 46–53. Tübingen: Max Niemeyer. DOI: 10.1515/9783111661643.46.

Hentschel, Elke. 1986. *Funktion und Geschichte deutscher Partikeln.* Ja*,* doch*,* halt *und* eben (Reihe Germanistische Linguistik 63). Tübingen: Max Niemeyer. DOI: 10.1515/9783111371221.

Karagjosova, Elena. 2009. A unified DRT-based account of accented and unaccented middle field *doch*. *Sprache & Datenverarbeitung* 33(1–2). 77–93.

Kratzer, Angelika. 1978. *Semantik der Rede: Kontexttheorie – Modalwörter – Konditionalsätze*. Königstein/Ts.: Scriptor.

Kratzer, Angelika. 1981. The notional category of modality. In Hans Jürgen Eikmeyer & Hannes Rieser (eds.), *Words, worlds and contexts. New approaches in World Semantics*, 38–74. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110842524-004.

Lasersohn, Peter. 2005. Context dependence, disagreement and predicates of personal taste. *Linguistics and Philosophy* 28. 643–686. DOI: 10.1007/s10988-005-0596-x.

Lücking, Andy, Jonathan Ginzburg & Robin Cooper. 2021. Grammar in dialogue. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 1155–1199. Berlin: Language Science Press. DOI: 10.5281/ZENODO.5599870.

Macário-Lopes, Ana Cristina. 1998. Contribução para o estudo dos valores discursivos de *sempre*. In Maria Antónia Mota & Rita Marquilhas (eds.), *Actas do XIII Encontro Nacional da APL*. 3–14. Lisboa: Colibri.

Maché, Jakob. 2019. *How epistemic modifiers emerge* (Trends in Linguistics 292). Berlin: De Gruyter. DOI: 10.1515/9783110411027.

Meisnitzer, Benjamin. 2012. Modality in the Romance languages: Modal verbs and modal particles. In Werner Abraham & Elisabeth Leiss (eds.), *Theory of Mind elements across languages* (Trends in Linguistics 243), 335–360. Berlin: de Gruyter. DOI: 10.1515/9783110271072.335.

Mendes, Amália & Pierre Lejeune. 2022. *Marcadores discursivos e marcadores modais: Uma análisa contrastiva de* claro *e seus equivalentes funcionais em Francês*. Talk given at Associação Portuguesa de Linguística.

Palmer, Frank Robert. 1986. *Mood and modality* (Cambridge Text Books in Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139167178.

Pinto de Lima, José. 1997. Caminhos semânticos-pragmáticos da gramaticalização: O case de *embora*. In Ana Maria Brito, Fátima Oliveira, Isabel Pires de Lima & Rosa Maria Martelo (eds.), *Sentido que a Vida Faz. Estudos para Óscar Lopes*, 643–655. Porto: Campo das Letras.

Portner, Paul. 2009. *Modality*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780199292424.001.0001.

Stephenson, Tamina. 2007. Judge dependence, epistemic modals and predicates of personal taste. *Linguistics and Philosophy* 30(4). 487–525. DOI: 10.1007/s10988-008-9023-4.

Sudo, Yasutada. 2013. Biased polar questions in English and Japanese. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Exploration in use-conditional meaning*, 275–295. Leiden: Brill. DOI: 10.1163/9789004183988_009.

Thurmair, Maria. 1989. *Modalpartikeln und ihre Kombinationen* (Linguistische Arbeiten 223). Max Niemeyer. DOI: 10.1515/9783111354569.

Weydt, Harald. 1969. *Abtönungspartikel. Die deutschen Modalwörter und ihre französischen Entsprechungen*. Bad Homburg: Gehlen.

Zimmermann, Malte. 2011. Discourse particles. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics. An International Handbook of Natural Language Meaning*, vol. 2 (Handbücher zur Sprach- und Kommunikationswissenschaft 33), 2011–238. Berlin: De Gruyter. DOI: 10.1515/9783110255072.2012.

A Construction-based Approach to Cantonese Classifiers

Francis Bond
Palacký University

Joanna Ut-Seong Sio
Palacký University

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 60–75

Keywords: classifier, Cantonese, HPSG, specifier, cognitive status, givenness, computational grammar

Bond, Francis & Joanna Ut-Seong Sio. 2024. A Construction-based Approach to Cantonese Classifiers. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 60–75. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.4.
CC-BY 4.0

Abstract

This paper focuses on the structure and interpretation of Cantonese NPs. We first map different cognitive statuses (Gundel et al. 1993) to different forms of Cantonese noun phrases, following the hierarchy proposed in Borthen & Haugereid (2005). We then provide an HPSG analysis for Cantonese noun phrases. We account for the differences between classifiers appearing with and without a numeral, where classifiers with no numeral are interpreted as having a cardinality of 'one'. We propose a Classifier Head Rule where the noun first takes a specifier containing the classifier, and the output further takes a determiner as its specifier. The analysis is implemented in an open-source Cantonese HPSG.

1 Introduction

Cantonese, a variety of Yue, belongs to the Sinitic branch of the Sino-Tibetan language family. Originating from southern China, it is named after Canton (Guangzhou), the capital city of the Guangdong province. Cantonese is spoken in Guangdong China, and the two Special Administrative Regions, Hong Kong and Macao, as well as in diaspora communities (e.g., Singapore, Malaysia, Australia, the United Kingdom and North America). There are over 82.4 million Cantonese native language speakers (Wikipedia contributors 2024).

This paper provides an HPSG analysis for Cantonese noun phrases with the following three implementations. First, we assign cognitive status to different types of Cantonese NPs, following the hierarchy proposed in Borthen & Haugereid (2005). Second, we account for the differences between classifiers appearing with and without a numeral. Classifiers with no numerals are interpreted as having a cardinality of 'one'. Third, we propose a Classifier Head Rule where the noun first takes a specifier containing the classifier, and the output further takes a determiner as its specifier. This is similar to a double specifier analysis except that the locus of the 'special treatment' (having to take both the classifier and then the determiner as specifiers) is built around the classifier. The analysis is implemented in an open-source Cantonese HPSG.[1]

2 Cantonese NPs

Cantonese NPs (unmodified) have the 4 schematic forms shown in Table 1 and they have different definiteness interpretations (Cheng & Sybesma 1999). In (1), we give example sentences illustrating the different types of NPs in the object position.

[†]We would like to thank the reviewers, Dan Flickinger, Emily Bender and Luis Morgado da Costa for their helpful comments and discussion.

[1]The implementation, using the DELPH-IN tools, is available at ⟨`https://github.com/neosome/yue`⟩.

Table 1: Definiteness (after Cheng & Sybesma 1999)

Type	Example
D-(X)-C-N	definite
X-C-N	indefinite
C-N	(in)definite
N	indefinite

D: demonstrative, X: numeral,* C: classifier, N: noun

*X can in fact be a numeral phrase or one of a small set of quantifiers, in this paper, we only discuss X being a numeral.

(1) Cantonese (yue)

a. D-(X)-C-N

明恩	食咗	呢	*(一)*	個	蘋果。
Ming4jan1	sik6-zo2	nei1	jat1	go3	ping4gwo2
Ming-Jan	eat-PERF	this	one	CL	apple

'Ming-Jan ate this apple.'

b. X-C-N

明恩	食咗	一	個	蘋果。
Ming4jan1	sik6-zo2	jat1	go3	ping4gwo2
Ming-Jan	eat-PERF	one	CL	apple

'Ming-Jan ate one apple.'

c. C-N

明恩	食咗	個	蘋果。
Ming4jan1	sik6-zo2	go3	ping4gwo2
Ming-Jan	eat-PERF	CL	apple

'Ming-Jan ate an/the apple.'

d. N

明恩	食咗	蘋果。
Ming4jan1	sik6-zo2	ping4gwo2
Ming-Jan	ate-PERF	apple

'Ming-Jan ate an apple/apples.'

In Chinese (Cantonese included), only definite NPs can appear in the subject or topic position in a sentence (Li & Thompson 1989), though not without exceptions (Li 1998).[2] Thus, it is important to include the definiteness information of the NPs when modeling Cantonese grammar. Definiteness is understood as the grammatical encoding of the pragmatic concept of identifiability (Chen 2004). Identifiability is related to the assumptions made by the speaker on the cognitive status of a referent in the mind of the addressee in the context of an utterance (Gundel et al. 1993).

[2]Li (1998) argues that when the interpretation of a Chinese [X-C-N] phrase (indefinite) has only a quantity reading rather than an individual reading, it can appear in the subject position.

Gundel et al. (1993)'s analysis has been implemented as a type-hierarchy in (Borthen & Haugereid 2005) and adequately describes the distinctions needed for Cantonese. We adopt Gundel et al. (1993)'s model in this paper.

Gundel et al. (1993) proposes six cognitive statuses: 'type-identifiable', 'referential', 'uniquely identifiable', 'familiar', 'activated', and 'in focus'. Each cognitive status can be expressed with different forms of noun phrases in different languages. The different forms serve as processing signals to the addressee.[3] Each of the cognitive statuses will be discussed below with Cantonese example sentences.

'Type-identifiable' refers to cases where the addressee is able to access a representation of the type of objects described by the expression (Gundel et al. 1993). The English indefinite article is used in such cases. In Cantonese, the same cognitive status can be expressed by a bare noun, or a [(X)-C-N] phrase, with the numeral being optional.

(2) 我 去 買 *(一)* *(個)* 西瓜 。
ngo5 heoi3 maai5 (jat1) (go3) sai1gwaa1
1SG go buy (one) (CL) watermelon

'I go buy a watermelon/watermelons.'

The only requirement on the addressee is that they understand the noun *sai1gwaa1* 'watermelon' to understand what it is to be bought. Note that when only a bare noun is used, it can be interpreted either as singular or plural.

'Referential' refers to cases where the speaker intends to refer to a particular object (or objects). The addressee needs to access a appropriate type-representation, plus either retrieving an existing representation of the speaker's intended referent or construct a new representation the time the sentence is uttered (Gundel et al. 1993). Borthen & Haugereid (2005) argue that cognitive status is speaker-oriented, and so

[3] One reviewer suggested other models of givenness, in particular, the four distinctions: (i) discourse-old and hearer-old; (ii) discourse-old and hearer-new; (iii) discourse-new and hearer-old, and (iv) discourse-new and hearer-new, as discussed in Birner (2021). We believe these four categories can be captured by Gundel et al. (1993)'s hierarchy, e.g., 'discourse-old and hearer-old' can be subsumed under 'familiar'; 'discourse-new and hearer-new' can be subsumed under 'type-identifiable'; 'discourse-new and hearer-old' are cases like e.g., 'the sun', 'the President', which can be subsumed under 'uniquely identifiable' (as these referents are unique in any particular context without prior introduction). 'Discourse-old and hearer-new' are cases where the referent is 'inferrable' (Prince 1981, Schwarz 2009), as in e.g., 'John put away all his grooming tools. *The combs* he put into the top drawer' (Birner 2021: 263). This can also be subsumed under 'uniquely identifiable', due to the prior introduction of an 'anchor' ('grooming tools' in the example), which makes the referent ('the combs') unique. In both 'discourse-new and hearer-old' and 'discourse-old and hearer-new' cases, [C-N] phrases are used in Cantonese, in other words, a distinction in forms is not made in Cantonese in these two cases. Given that the same forms of Cantonese noun phrases are often used for multiple categories of givenness, with the trend that the more 'given' (the more accessible in the addressee's mental representation) is associated with the use of the demonstrative (rather than clearly demarcated categorical 'forms to givenness' matching), we believe that Gundel et al. (1993)'s hierarchy provides a sufficient and fitting model for our purpose. Furthermore, as noted in the main text, Gundel et al. (1993) is adopted mainly because it has been implemented as a type-hierarchy.

'referential' is taken out of the hierarchy. They treat 'referential' as specificity which is speaker-oriented and introduce a separate feature (SPECI bool) that can be cross-classified with the different cognitive statuses (both definite and indefinite). We follow them to exclude 'referential' from our hierarchy of cognitive status.

'Uniquely identifiable' refers to cases where the addressee can identify referent on the basis of the nominal alone (Gundel et al. 1993). Identifiability does not have to be familiarity if enough descriptive content is provided (Gundel et al. 1993). In these situations, a [C-N] phrase is used in Cantonese, as shown in the example below (PERF = perfective marker; SFP = sentence-final-particle). The sentence can be used when there is only one open window in the non-linguistic context.

(3) 閂咗 個 窗 佢 丫 。
saan1-zo2 go3 coeng1 keoi5 aa1
close-PERF CL window 3SG SPF

'Close the window.'

'Familiar' is when the addressee is able to uniquely identify the intended referent because they already has a representation of it in memory (in long-term memory if it has not been recently mentioned or perceived, or in short-term memory if it has) (Gundel et al. 1993). In these cases, both a [C-N] phrase and a [D-(X)-C-N] phrase can be used in Cantonese. For example, in a context where a particular student has been mentioned earlier on in the discourse, using *(go)2 go3 hok6saang1* 'the/that student' to refer to her/him would be appropriate.

(4) (嗰) 個 學生 去咗 邊 呀？
(go2) go3 hok6saang1 heoi3-zo2 bin1 aa3
(that) CL student go-PERF where SPF

'Where does the student go?'

'Activated' is defined as a referent being represented in current working memory; it can be retrieved from long term memory, or they may arise from the immediate linguistic or extra-linguistic contexts (Gundel et al. 1993). In these cases, a demonstrative has to be used, as in *go2 di1 seng1* 'that noise' below:

(5) 嗰 啲 聲 攪 到 我 成 晚 都 瞓唔着。
go2 di1 seng1 gaau2 dou3 ngo5 seng4 maan5 dou1 fan3-m4-zoek6
that CL noise make to.the.extent 1SG whole night also sleep-NEG-fall

'That noise made me unable to sleep the whole night'

In-foc is represented not only in short-term memory, but is also at the current center of attention' (Gundel et al. 1993). In these cases, again, a demonstrative needs to be used, as in *go2 tiu4 jyu2* 'that fish' below:

(6) 嗰 條 魚 實在 太 好味 啦 。
go2 tiu4 jyu2 sat6zoi6 taai3 hou2mei6 laa3
that CL fish indeed too yummy SPF

'That fish is indeed too yummy.'

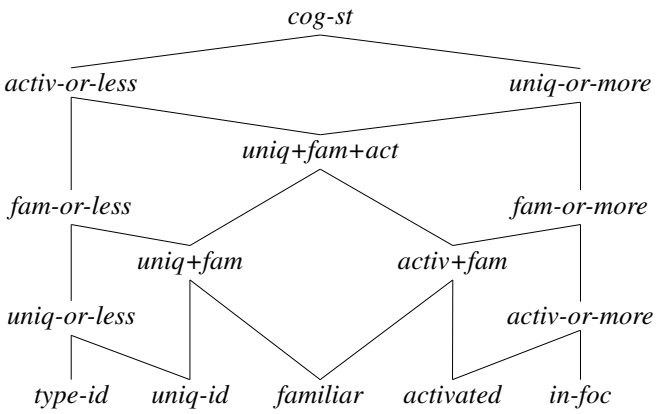


Figure 1: Cognitive Status Hierarchy

Borthen & Haugereid (2005) provide an HPSG-based type hierarchy of cognitive status, which was then refined by Bender & Goss-Grubbs (2008), as shown in Figure 1. Different languages have different inventories of referring expressions that can be used for different cognitive statuses. In Cantonese, we propose the interpretations in Table 2.

Table 2: Cognitive status

Type	**Example**	**cog-st**	Definiteness
D-(X)-C-N	呢 (一) 個蘋果	*fam-or-more*	Definite
X-C-N	一個蘋果	*type-id*	Indefinite
C-N	個蘋果	*fam-or-less*	In/Definite
N	蘋果	*type-id*	Indefinite (or Generic)

In Sio & Song (2015), D-(X)-C-N covers all cognitive statuses except *type-id* in Figure 1, i.e., *uniq-or-more*. In this paper, we restrict D-(X)-C-N to *fam-or-more*. D-(X)-C-N is not used in cases of *uniq-id*. *Uniq-id* (uniquely identifiable) is defined as the addressee being able to identify the referent on the basis of the nominal alone. We believe this covers cases which Schwarz (2009) calls *larger situation definites* (e.g., the moon), *immediate situation definites* (in a room with one door clearly open,

e.g., close the door, please.) and part-whole bridging definites (e.g., *I bought a shirt yesterday. The buttons are too big.*). In these situations, C-N rather than D-(X)-C-N is used in Cantonese. In Sio & Song (2015), C-N is totally under-specified, compatible with all *cog-st*. In this paper, we restrict it to *fam-or-less*, excluding it from *activated* and *in-foc*.

The general tendency of the mapping between the cognitive statuses to the Cantonese NPs is such that the demonstrative, D-(X)-C-N, is required when the cognitive status reaches a certain level of prominence (i.e., 'familiar', 'activated', 'in-focus') while C-N spans over some less prominent cognitive status (i.e., 'type-identifiable', 'uniquely-identifiable' and 'familiar'). D-(X)-C-N and C-N overlap in covering 'familiar' cases. A note of caution is required here. The mapping between a certain cognitive status to a particular NP form is not always easy to determine, we follow the coding guidelines from the protocol for each cognitive status in Gundel et al. (1993) to the best of our understanding. It is possible that in some situations, the choice could just be a preference.[4]

3 Analysis

Following the majority of HPSG analyses on Chinese NPs (Wang & Liu 2007: and references therein), we adopt an NP analysis, where the numeral forms a constituent together with the classifier (Her 2016). We treat both the demonstrative and classifier as specifiers, following the analysis of Mandarin by Ng (1997) and Wang & Liu (2007). However, instead of the nouns selecting two specifiers and modifying the HEAD-SPECIFIER rule, we add a new classifier construction (*cl-head*: §**??**) which requires another specifier after consuming the classifier. Empirical data from a wide range of languages does not require two specifiers for an adequate description of noun phrases and it is the classifier that is special in the Cantonese noun phrase, thus, we make the classifier-construction the locus of the unusual syntax. Currently, we have found no data that differentiates clearly between our one-specifier analysis and the two-specifier analysis. In future work, we will attempt to discover if there are different predictions from the two approaches.

In Cantonese, when the numeral is omitted, both X-C-N and C-N have a cardinality of 'one'. However, in answering the question 'how many', only X-C-N can be used. This is, in part, similar to the contrast between 'one N' and 'a/an N' in English. The semantics represents this with the *card* relation, with a value of '1'. In addition, the well-formed semantics must have a quantifier for every referential index, if there is no explicit demonstrative, the grammar must supply this from a construction.

Our analysis requires one new lexical type (for sortal classifiers); one new feature used on classifier phrases to mark if they have been explicitly enumerated or not and

[4]It is not easy to determine whether [C-N] phrases can cover 'activated'; similarly, it is not easy to decide whether [D-(X)-C-N] phrases can cover 'uniq-id'. At any rate, the mapping can be easily adjusted for future work (e.g., for D-(X)-C-N, from *fam-or-more* to *uniq-or-more*.). We have chosen a more restrictive approach in this paper.

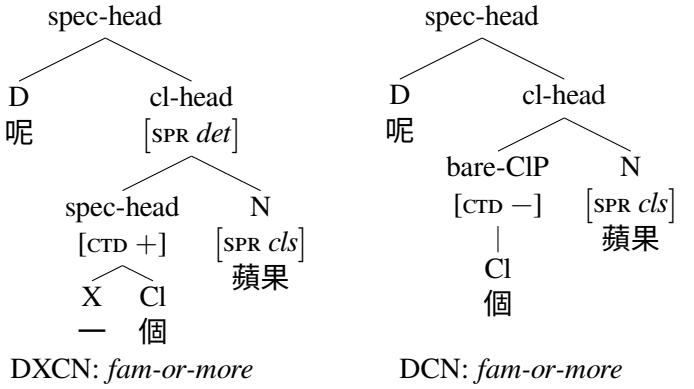


Figure 2: NPs with demonstratives

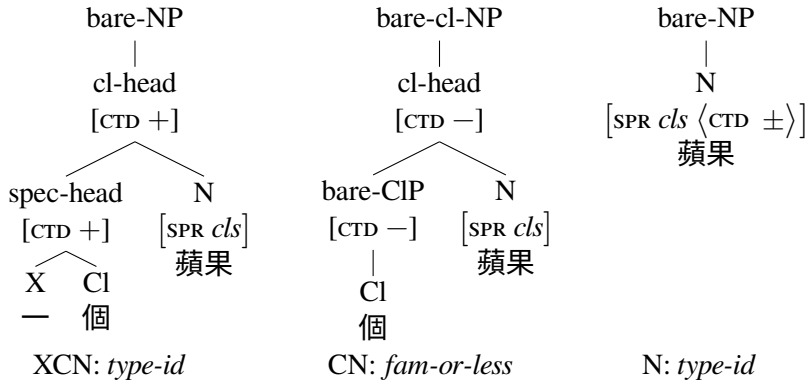


Figure 3: NPs without demonstratives

three new constructions (classifier-head, bare-classifier and bare-classifier-np) as well as changes to the existing lexical types for numerals, and the head-specifier and bare-np rules. Derivation trees are shown for the two NP types with demonstratives in Figure 2 and the three types without in Figure 3. The descriptions given below are all only partial, we omit information we consider not relevant to the discussion at hand. Paths may also be shortened for clarity.

3.1 Lexical types

3.1.1 Classifier lexical type

The sortal classifier lexical type is shown in (7). The category is *cls* for classifier. The *cl-pred* shows where the predicate would be for an actual entry of a word. They optionally take a number as their specifier. The head-specifier rule will link the XARG to the INDEX of the specified constituent.

The sortal classifier lexical type doesn't say anything about cognitive status, nom-

inals containing the classifier are compatible with all cognitive status. The ultimate cognitive status of a nominal containing a classifier is determined by (i) whether it is preceded by a numeral; (ii) whether the nominal contains a demonstrative.

(7)
$$\begin{bmatrix} \textit{classifier-lex} & \\ \textsc{syn} & \begin{bmatrix} \textsc{head} & \textit{cls} \\ \textsc{val.spr} & \left\langle \begin{bmatrix} \textsc{syn.head} & \textit{num,} \\ \textsc{opt} & + \end{bmatrix} \right\rangle \end{bmatrix} \\ \textsc{sem} & \begin{bmatrix} \textsc{hook} & \begin{bmatrix} \textsc{index} & \boxed{1} \\ \textsc{xarg} & \boxed{2} \end{bmatrix} \\ \textsc{rels} & \left\langle \begin{bmatrix} \textsc{reln} & \textit{cl-pred} \\ \textsc{arg}0 & \boxed{1}\ e \\ \textsc{arg}1 & \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Because sortal classifiers do not introduce a referent, their type is *e* (non-indexical). This contrasts with referential noun-phrases, which will be *ref-ind* (referential index). The types we use are shown in Figure 4.

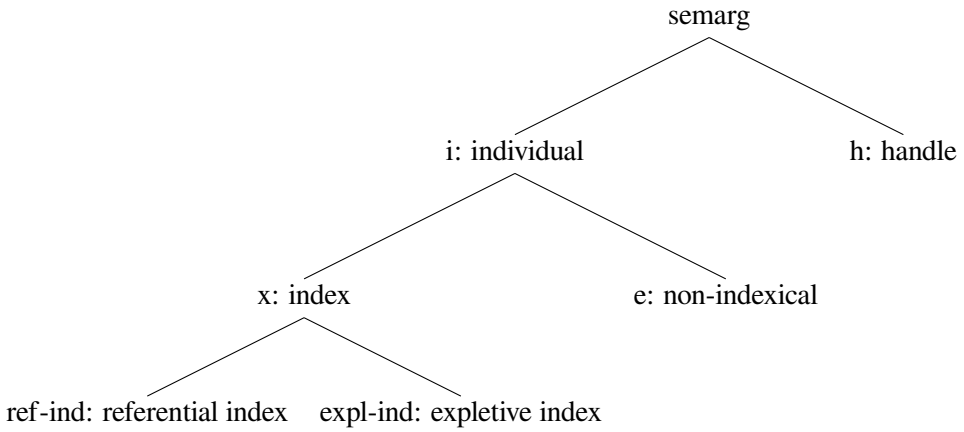


Figure 4: Types of semantic objects

3.1.2 Numeral lexical type

Their semantics is somewhat special, using CARG (Constant Argument) to introduce the value of the number. The index of the thing it will specify over (the classifier) is the same as ARG1 on the relation it introduces. That is, it counts the classifier. Further, it sets it's head to CTD +: it has been explicitly counted.

In the implemented grammar, rather than defining a new feature, we reuse the PRON feature. This makes the size of the feature structure smaller. Because PRON was originally only used on NPs and we only use it here on ClPs, its interpretation is never ambiguous.

(8)
$$\begin{bmatrix} \textit{numeral-lex} \\ \textsc{syn} & \begin{bmatrix} \textsc{head} & \textit{numeral} \\ \textsc{val.spec} & \left\langle \begin{bmatrix} \textsc{syn.head.ctd} & + \\ \textsc{sem.index} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \\ \textsc{sem} & \begin{bmatrix} \textsc{hook} & \left[\textsc{index}\ \boxed{2}\right] \\ \textsc{rels} & \left\langle \begin{bmatrix} \textsc{reln} & \textit{card-relation} \\ \textsc{carg} & \textit{?} \\ \textsc{arg}0 & \boxed{2}\ \textit{non-ref} \\ \textsc{arg}1 & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

3.1.3 Noun lexical type

The Cantonese *noun-lex* (9) sets its specifier to be a classifier, not a determiner. This means it must either pick up a specifier, or have the specifier discharged by the bare NP rule.

(9)
$$\begin{bmatrix} \textit{noun-lex} \\ \textsc{syn} & \begin{bmatrix} \textsc{head} & \textit{noun} \\ \textsc{val.spr} & \left\langle \begin{bmatrix} \textsc{syn.head} & \textit{cls,} \\ \textsc{opt} & + \end{bmatrix} \right\rangle \end{bmatrix} \\ \textsc{sem} & \begin{bmatrix} \textsc{hook} & \left[\textsc{index}\ \boxed{1}\right] \\ \textsc{rels} & \left\langle \left[\textsc{arg}0\ \boxed{1}\ \textit{ref-ind}\right] \right\rangle \end{bmatrix} \end{bmatrix}$$

3.1.4 Demonstrative

A demonstrative (10) constrains the index of the noun it specifies to be *fam-or-more*, it does not care about the CTD value of its specifier.

(10)
$$\begin{bmatrix} \textit{dem-lex} \\ \textsc{syn} & \begin{bmatrix} \textsc{head} & \textit{det} \\ \textsc{val.spec} & \left\langle \left[\textsc{local.sem.hook.index}\ \boxed{1}\right] \right\rangle \end{bmatrix} \\ \textsc{sem} & \begin{bmatrix} \textsc{hook} & \left[\textsc{index}\ \boxed{1}\ \left[\textsc{cog-st}\ \textit{activ+fam}\right]\right] \\ \textsc{rels} & \left\langle \left[\textsc{arg}0\ \boxed{1}\ \textit{ref-ind}\right] \right\rangle \end{bmatrix} \end{bmatrix}$$

3.2 Rules

3.2.1 Classifier Head Rule (*cl-head*)

This rule is the main new construction. It takes two daughters. The left-hand, non-head daughter (NHD) takes a classifier phrase as its daughter. The right-hand, head daughter (HD), takes a noun or nominal that requires a classifier as its specifier. Crucially, the parent also requires a specifier, this time a determiner: in this way a noun

phrase can effectively have two specifiers, so long as the first is a classifier, and the second a determiner, even though the noun has only one specifier. The value of CTD is passed from the non-head daughter (the specifier) to the new specifier slot, making it visible to the bare NP rules. In most other ways it is identical to the *spec-head* rule (and thus can inherit from a common super-type).

(11)
$$\begin{bmatrix} \textit{cl-head-phrase} \\ \text{SYN} \quad \begin{bmatrix} \text{VAL.SPR} & \langle [\text{HEAD } \textit{det} \; [\text{CTD } \boxed{0}]] \rangle \\ \text{SEM.INDEX} & \boxed{1} \end{bmatrix} \\ \text{NHD } \boxed{2} \begin{bmatrix} \text{VAL.SPR} & \langle [\text{HEAD } \textit{cls} \; [\text{CTD } \boxed{0}]] \rangle \\ \text{SEM.XARG} & \boxed{1} \end{bmatrix} \\ \text{HD} \quad \begin{bmatrix} \text{VAL.SPR} & \langle \boxed{2} \rangle \\ \text{SEM.INDEX} & \boxed{1} \end{bmatrix} \end{bmatrix}$$

3.2.2 Head Specifier Rule (*spec-head*)

The head specifier rule has one change: we do not allow a classifier as specifier — in this case, the classifier head rule should be used instead.

3.2.3 Bare NP rules (*bare-NP*, *bare-cl-NP*)

We introduce two bare NP rules, for the two different cognitive statuses we want. The first (12) is a headed unary rule, which makes an NP with the specifier satisfied, if the head daughter's specifier is *cls-or-det* and *ctd* +. This will be true for nouns with a numeral and classifier as input, or just for a noun, as its CTD is unspecified. The *cog-st* of the resulting NP is set to *type-id*.

(12)
$$\begin{bmatrix} \textit{bare-np-phrase} \\ \text{SYN} \quad \begin{bmatrix} \text{VAL.SPR} & \langle\rangle \\ \text{SEM.INDEX} & \boxed{1} \; [\text{COG-ST } \textit{id-type}] \end{bmatrix} \\ \text{HD} \quad \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{VAL.SPR} & \langle [\text{HEAD } \textit{cls-or-det} \; [\text{CTD } +]] \rangle \\ \text{SEM.INDEX} & \boxed{1} \end{bmatrix} \\ \text{C-CONT} \quad \begin{bmatrix} \text{RELS} & \left\langle \begin{bmatrix} \text{RELN} & \textit{exist\_q} \\ \text{ARG0} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

The second (13) restricts the value of the head daughter's spec to a determiner (DET) with CTD −, and the NP's *cog-st* is set to *fam-or-less*. This excludes bare nouns, whose specifier is *cls* and nouns specified with a classifier and no numeral, which will be CTD +.

(13)
$$\begin{bmatrix} \textit{bare-cl-np-phrase} \\ \textsc{syn} & \begin{bmatrix} \textsc{val.spr} & \langle\rangle \\ \textsc{sem.index} & \boxed{1}\ \left[\textsc{cog-st}\ \textit{fam-or-less}\right] \end{bmatrix} \\ \textsc{hd} & \begin{bmatrix} \textsc{head} & \textit{noun} \\ \textsc{val.spr} & \left\langle \left[\textsc{head}\ \textit{det}\ \left[\textsc{ctd}\ -\right]\right]\right\rangle \\ \textsc{sem.index} & \boxed{1} \end{bmatrix} \\ \textsc{c-cont} & \begin{bmatrix} \textsc{rels} & \left\langle \begin{bmatrix} \textsc{reln} & \textit{exist\_q} \\ \textsc{arg}0 & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

In the implemented grammar, both of these rules inherit from a single supertype *bare-np-super* which contains the shared structure.

3.2.4 Bare Classifier Rule

This non-branching rule (14) takes a classifier, and creates a classifier phrase. As the interpretation is always that there is one thing being classified, the rule adds a *card-relation* with CARG of *1*. It also sets CTD to − so that the classifier phrase will pass through the Bare NP Rule for bare classifiers (3.2.3). The rule is similar to the NO-SPR-CL-RULE proposed by Sio & Song (2015: 189), but differs in two important ways. The first is that it explicitly models the cardinality. The second is that it marks the head so that the cognitive status can be restricted.

(14)
$$\begin{bmatrix} \textit{bare-cl-phrase} \\ \textsc{syn} & \begin{bmatrix} \textsc{val} & \begin{bmatrix} \textsc{spr} & \langle\rangle, \\ \textsc{spec} & \boxed{3}\ \left\langle \left[\textsc{index}\ \boxed{0}\right]\right\rangle \end{bmatrix} \\ \textsc{sem} & \begin{bmatrix} \textsc{index} & \boxed{0}, \\ \textsc{xarg} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \textsc{hd} & \begin{bmatrix} \textsc{head}\ \textit{cls}\ \left[\textsc{ctd}\ -\right] \\ \textsc{val} & \begin{bmatrix} \textsc{spr} & \left\langle \textsc{unexpressed}\right\rangle \\ \textsc{spec} & \boxed{3} \end{bmatrix} \\ \textsc{sem} & \begin{bmatrix} \textsc{index} & \boxed{2} \\ \textsc{xarg} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \textsc{c-cont} & \begin{bmatrix} \textsc{rels} & \left\langle \begin{bmatrix} \textsc{reln} & \textit{card-relation} \\ \textsc{arg}1 & \boxed{2} \\ \textsc{carg} & \textit{1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

3.3 Semantics

The semantics for the 5 types are given in Table 3 (using indexed MRS: Copestake et al. 2005). We use the jyutping transliteration for the predicate names, in the actual grammar they are written with Chinese characters. In all cases save the bare noun, there are four predicates: a quantifier (either from the determiner or the Bare Noun

Type	Indexed MRS	**cog-st** of x_1
D-X-C-N	nei1_q(x_1, h_2, h_3); card(e_4, x_1, 'x'), go3_x(e_5, x_1); ping4gwo2_n(x_1)	*fam-or-more*
D-C-N	nei1_q(x_1, h_2, h_3); card(e_4, x_1, '1'), go3_x(e_5, x_1); ping4gwo2_n(x_1)	*fam-or-more*
X-C-N	exist_q(x_1, h_2, h_3); card(e_4, x_1, 'x'), go3_x(e_5, x_1); ping4gwo2_n(x_1)	*type-id*
C-N	exist_q(x_1, h_2, h_3); card(e_4, x_1 '1'), go3_x(e_5, x_1), ping4gwo2_n(x_1)	*fam-or-less*
N	exist_q(x_1, h_2, h_3), ping4gwo2_n(x_1)	*type-id*

Table 3: MRS for the various combinations

Phrase rule), the head noun, the classifier (which takes the head noun as its external argument (ARG1) and the cardinality relation (which has two arguments, the classifier as its external argument and the amount as a value). The semantics expresses the situation where the noun is being measured out in units of the classifier, to the amount of the number. If the cardinality is not explicitly given, then the default value of one comes from the Bare Classifier Rule. For the bare noun, there is no measurement, so the noun appears with just the default determiner.

This is compatible with the analysis of Takao (2005) for Japanese, where he combines the cardinality and classifer into a single *measure* relation, as in (15). We choose to encode it as two different relations to retain compatibility with the analyses of other languages in the DELPH-IN framework (Uszkoreit 2002). However, it would be trivial to transform one to the other.

(15)
$$\begin{bmatrix} \text{RELN} & \textit{measure} \\ \text{ARG1} & x_1 \\ \text{NUM} & \textit{1} \\ \text{DIMENSION} & \text{個}\_x\ (\text{IN OUR CASE GO3\_X}) \end{bmatrix}$$

It differs from the analysis of Bender & Siegel (2004) who co-index the classifier and noun, and have no representation for the classifier. This models the intuition that they are not separate referents, but loses the opportunity to represent the classifier semantics. As the choice of classifier has some effect on the interpretation of the meaning, this is undesirable. We model this intuition by making the index (ARG0) of the classifier (and cardinality) relation non-referential.

Kim & Yang (2007) co-index the classifier and noun, and have a representation for the classifier. This means that two referential predicates share an ARG0. We avoid this to retain compatibility with the characteristic variable property of Dependency MRS (Copestake 2009). Again this could be converted easily as our analysis captures the same intuition (that we need two predicates and that they have only one quantifier).

Our grammar is bi-directional, it can parse from a string to the semantics or from the semantics to a string (or strings). In the fully specified semantics, the classifier

is given. In many applications (such as machine translation), it would need to be generated, which could be done with a generative language model or an ontology such as wordnet (Mok et al. 2012, Morgado da Costa et al. 2016).

4 Conclusion and future work

In this paper, we presented our preliminary attempt in generating different nominal types in Cantonese (with construction-specific rules) as well as mapping them to different cognitive statuses in HPSG. In the future, we want to expand our investigation in the following directions. Our analysis does not investigate the effects of modification on the semantics or cognitive status, nor the anaphoric use of the classifier (in the absence of the head noun). We also have only looked at sortal classifiers, not mensural or kind. With the inclusion of cognitive statuses, we would like to model the restriction on banning indefinite NPs (i.e., *type-id*) appearing in subject and topic position in Chinese (Li & Thompson 1989). We would like to extend the analysis to cover these, and test against naturally occurring texts. Finally, although we have focused on Cantonese here, we would like to compare our analysis in more depth to those of other classifier languages, especially those with computational analyses like Indonesian (Moeljadi et al. 2015), Japanese (Siegel et al. 2016), Korean (Kim et al. 2011) and Mandarin (Müller & Lipenkova 2013, Fan et al. 2015).

References

Bender, Emily M. & David Goss-Grubbs. 2008. Semantic Representations of Syntactically Marked Discourse Status in Crosslinguistic Perspective. In *Proceedings of the 2008 conference on semantics in text processing*, 17–29.

Bender, Emily M. & Melanie Siegel. 2004. Implementing the syntax of Japanese numeral classifiers. In *Proceedings of the IJC-NLP-2004*.

Birner, Betty J. 2021. *Pragmatics: a slim guide*. Oxford University Press, USA.

Borthen, Kaja & Petter Haugereid. 2005. Representing Referential Properties of Nominals. *Research on Language and Computation* 3(2-3). 221–246.

Chen, Ping. 2004. Identifiability and definiteness in Chinese. *Linguistics* 42(6). 1129–1184. DOI: 10.1515/ling.2004.42.6.1129.

Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic inquiry* 30(4). 509–542.

Copestake, Ann. 2009. Slacker semantics: why superficiality, dependency and avoidance of commitment can be the right way to go. In *Proceedings of the 12th conference of the european chapter of the ACL (EACL 2009)*, 1–9. Athens.

Copestake, Ann, Dan Flickinger, Ivan A. Sag & Carl Pollard. 2005. Minimal Recursion Semantics. An introduction. *Research on Language and Computation* 3(4). 281–332.

Fan, Zhenzhen, Sanghoun Song & Francis Bond. 2015. Building Zhong [|], a Chinese HPSG meta-grammar. In *Proceedings of the 22nd international conference on Head-Driven Phrase Structure Grammar (HPSG 2015)*, 97–110. DOI: 10.21248/hpsg.2015.6.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2). 274–307.

Her, One-Soon. 2016. Structure of numerals and classifiers in Chinese. *Historical and typological perspectives and cross-linguistic implications*. 28–29.

Kim, Jong-Bok & Jaehyung Yang. 2007. Syntax and semantics of Korean numeral classifier constructions. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford Department of Linguistics and CSLI's LinGO Lab*, 163–172. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2007.10.

Kim, Jong-Bok, Jaehyung Yang, Sanghoun Song & Francis Bond. 2011. Processing of Korean and the development of the Korean resource grammar. *Linguistic Research* 28(3). 635–672.

Li, Charles N & Sandra A Thompson. 1989. *Mandarin Chinese: a functional reference grammar*. Univ of California Press.

Li, Yen-hui Audrey. 1998. Argument determiner phrases and number phrases. *Linguistic Inquiry* 29(4). 693–702.

Moeljadi, David, Francis Bond & Sanghoun Song. 2015. Building an HPSG-based Indonesian resource grammar (INDRA). In *ACL 2015 workshop on grammar engineering across frameworks (GEAF 2015)*.

Mok, Hazel Shuwen, Eshley Huini Gao & Francis Bond. 2012. Generating numeral classifiers in Chinese and Japanese. In *Proceedings of the 6th global wordnet conference (GWC 2012)*. 211-218. Matsue.

Morgado da Costa, Luís, Francis Bond & Helena Gao. 2016. Mapping and generating classifiers using an open Chinese ontology. In *Proceedings of the 8th Global Wordnet Conference (GWC 2016)*, 248–254.

Müller, Stefan & Janna Lipenkova. 2013. Chingram: a TRALE implementation of an HPSG fragment of Mandarin Chinese. *Sponsors: National Science Council, Executive Yuan, ROC Institute of Linguistics, Academia Sinica NCCU Office of Research and Development*. 240.

Ng, Say Kiat. 1997. *A double specifier account of Chinese NPs using head-driven phrase structure grammar*. University of Edinburgh: Department of Linguistics. (MSc Thesis).

Prince, Ellen F. 1981. *Toward a taxonomy of given-new information*.

Schwarz, Florian. 2009. *Two types of definites in natural language*. University of Massachusetts Amherst.

Siegel, Melanie, Emily Bender & Francis Bond. 2016. *Jacy — an implemented grammar of Japanese*. CSLI Publications.

Sio, Joanna Ut-Seong & Sanghoun Song. 2015. Divergence in expressing definiteness between Mandarin and Cantonese. In Stefan Müller (ed.), *Proceedings of*

the 22nd International Conference on Head-Driven Phrase Structure Grammar, Nanyang Technological University (NTU), Singapore, 177–194. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2015.11.

Takao, GUNJI. 2005. Measurement and quantification revisited. 8(8). 21–36.

Uszkoreit, Hans. 2002. New chances for deep linguistic processing. In *19th international conference on computational linguistics: COLING-2002*, XIV–XXVII. Taipei.

Wang, Lulu & Haitao Liu. 2007. A description of Chinese NPs using Head-Driven Phrase Structure Grammar. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford Department of Linguistics and CSLI's LinGO Lab*, 287–305. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2007.17.

Wikipedia contributors. 2024. *Cantonese — Wikipedia, the free encyclopedia*. [Online; accessed 24-March-2024]. https://en.wikipedia.org/w/index.php?title=Cantonese&oldid=1214658644.

The morphological causative in Panãra: A Grammar Matrix implementation

Carly Crowther
University of Washington

Emily Luedke
University of Washington

Emily M. Bender
University of Washington

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 76–87

Keywords: HPSG, Morphotactics, Valence-Changing Operations, Panãra, LinGO Grammar Matrix

Crowther, Carly, Emily Luedke & Emily M. Bender. 2024. The morphological causative in Panãra: A Grammar Matrix implementation. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 76–87. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.5. CC-BY 4.0

Abstract

In this paper, we describe the modeling of the causative morpheme in an implemented HPSG grammar for the Panãra language. The Panãra causative morpheme appears on the verb between the agreement markers for the ergative and absolutive arguments, increasing the verb's valency to two core arguments. Due to the linear ordering of inflectional morphemes (position classes) in Grammar Matrix grammars, the position of the causative morpheme would require the verb to have information about its valence before the morpheme is attached. We model the descriptive data with a new lexical rule that introduces the causative morpheme and changes the verbal inflection path from intransitive to transitive using the existing Valence Change library. This analysis is validated in an implemented grammar fragment for Panãra.

1 Introduction

In this paper, we ask whether and how the approach to morphosyntax encoded in the Grammar Matrix can accommodate constructions where the choice and interpretation of a morpheme closer to the root depends on morphology further away from the root. Specifically, we look into the interaction of agreement markers and causative morphology in Panãra (Jê, Brazil).

Panãra exhibits complex verbal morphology with a strictly ordered set of polypersonal agreement affixes. The verb stem is inflected first with either the object (O) prefix[1] for transitive constructions, or the subject (S) prefix for intransitive constructions. The transitive agent (A) prefix attaches after. A causative morpheme can appear with an intransitive verb, increasing its valence to two core arguments. This morpheme is only compatible with intransitive verbs, as is the S argument inflectional position class. Contrastively, the A and O inflectional position classes are only compatible with transitive verb stems. The causative morpheme appears further from the stem than the O position class, which poses a problem as the order of inflection would not allow the verb to select an object before 'knowing' that the valence is to be increased with the causative morpheme. However, Panãra exhibits ergative-absolutive alignment in its verb agreement (as well as its case system), meaning that the S and O agreement prefixes happen to be homophonous. Furthermore, the interaction with the Valence Change

*Authors Carly Crowther and Emily Luedke contributed equally to this paper.

†We are grateful to Myriam Lapierre for guidance on understanding the grammar of Panãra, to the HPSG audience and reviewers for their insightful feedback, and to the Panãra people for their collaboration with linguistic researchers.

[1]Throughout this paper we use the term prefix to refer to the morphemes Bardagil (2018) refers to as clitics. This reflects the analysis we pursued of the morphemes as both syntactically and phonologically dependent, even if less phonologically integrated than other affixes.

library provided by the Grammar Matrix preserves agreement information about an intransitive subject in the resulting transitive complement. We model the causative morpheme to take the S agreement affix as input, and allow the A agreement prefix to further inflect the verb to produce the desired fully inflected causative verb construction.

2 Background

2.1 Panãra

Panãra [ISO 639-3: kre] is a Jê language spoken by about 630 native speakers, between the states of Pará and Mato Grosso in Brazil (Lapierre 2023). Data for this project comes from Bardagil's (2018) dissertation, which is based on primary field work and description of the Panãra language. This paper stems from a class project in which we used the Grammar Matrix customization system (Bender et al. 2010) and hand-edited TDL (Type Description Language) files to build an implemented grammar of Panãra.

2.2 The LinGO Grammar Matrix

The LinGO Grammar Matrix (Bender et al. 2002, 2010, Zamaraeva et al. 2022) is a grammar engineering framework for creating implemented grammars using Head-Driven Phrase Structure Grammar (HPSG) (Pollard & Sag 1994, Müller et al. 2024) and Minimal Recursion Semantics (MRS) (Copestake et al. 2005). The grammar fragments are written in TDL formalism (Copestake 2002a) and can be interpreted by other DELPH-IN software, including the Linguistic Knowledge Builder (LKB) grammar development environment (Copestake 2002b).

The Grammar Matrix includes various libraries developed over the years to address individual phenomena such as morphotactics (Goodman 2013) and valence-change operations (Curtis 2018). The morphotactics library creates a model of morphological relations via strictly ordered position classes. This library adopts an approach to morphology that fits into Stump's (2001) inferential-incremental category. Roots are inflected by adding morphemes via lexical rules, with each morpheme contributing its own morphosyntactic properties, and inflection is constrained by a strict linear ordering, with each position class taking one or more others as possible inputs. The valence-change library generates lexical rules to increase or decrease a verb's valency tailored to a language's grammar.

3 Data

3.1 Argument roles, agreement, and case

Panãra has an ergative-absolutive syntactic alignment system. The A argument of a transitive verb is marked with the ergative morpheme *hẽ*.[2] The transitive verb's O argument and intransitive verb's sole S argument do not receive case marking. (1), (2), and (3) illustrate the first person singular pronoun serving as each of the three types of core arguments.[3]

(1) Jyrawâ inkjẽ.
Jy-ra-wâ inkjẽ
INTR-1SG.S-born 1SG
'I was born.' (Bardagil 2018: 103)

(2) Karân kamẽrânpun inkjẽ.
Ka-rân ka-mẽ-r-ânpun inkjẽ
2SG-DU.ERG 2SG.A-DU-1SG.O-see 1SG
'You two saw me.' (Bardagil 2018: 121)

(3) Inkjẽ hẽ rêsunpa nãkãã.
Inkjẽ hẽ rê-s-unpa nãkãã
1SG ERG 1SG.A-3SG.O-fear snake
'I'm scared of snakes.' (Bardagil 2018: 59)

As shown above, the first person pronoun only receives ergative case marking when it is the A argument of a transitive construction; the S and O arguments pattern together in the unmarked absolutive case.[4] This ergative/absolutive alignment extends to the verbal agreement prefix paradigm as well, described in Tables 1 and 2.

For Panãra intransitive verbs, the prefix that agrees with the sole S argument appears directly adjacent to the left edge of the verb root. For transitives, the O argument agreement prefix occurs in this same location; the A argument agreement prefix precedes it. Panãra also has a dual marker *mẽ-*, which marks agreement with a dual number value on the A, O, both A and O, or S. Intransitive verbs receive an additional verbal prefix attached to the left edge of the verb, which indicates the intransitivity of the verb. This pattern is summarized in Tables 3 and 4.

[2]Dual and plural personal pronouns are case marked with an ergative suffix rather than *hẽ*, as seen in (2) (Bardagil 2018).

[3]The gloss line in all following IGT examples has been changed slightly from the reference material to reflect the morpheme's agreement with the syntactic role (S/O/A) of the argument rather than the argument's case, as syntactic roles are more relevant to our paper.

[4]This pattern is identical for both pronouns and full NPs.

Person	SG	DU	PL
1	rê	rê...mẽ	nẽ
2	ka	ka...mẽ	ka rê
3	ti	ti...mẽ	nẽ

Table 1: Ergative agreement prefixes

Person	SG	DU	PL
1	ra (r)	mẽ...ra (r)	ra (p)
2	a (k)	mẽ...a (k)	rê...a (rê...k)
3	ø (s/j)	mẽ...ø (s/j)	ra (r)

Table 2: Absolutive agreement prefixes
(Allophones for vowel-initial verbs in parentheses)

A agr-	Dual-	O agr-	Verb root

Table 3: Transitive Verbs

Intrans-	Dual-	S agr-	Verb root

Table 4: Intransitive Verbs

3.2 The causative morpheme

Panãra causatives are formed with the verbal prefix *ho-*, which attaches to intransitives only,[5] demoting the S to O and providing a new A argument. Causatives, being derived transitives, don't take the intransitive *jy-* prefix. This pattern is illustrated in (4)–(6).

(4) Ka jõpãã jysõti.
Ka jõpãã jy-ø-sõti
2SG child INTR-3SG.S-sleep
'Your child sleeps.' (Bardagil 2018: 108)

(5) Ka hẽ kahosõti ka jõpãã.
Ka hẽ ka-ho-ø-sõti ka jõpãã
2SG ERG 3SG.A-CAUS-3SG.O-sleep 2SG child
'You made your child sleep.' (Bardagil 2018: 108)

(6) *Inkjẽ hẽ rêhokuri inkjẽ jõpãã suasĩra jĩ.
Inkjẽ hẽ rê-ho-ø-kuri inkjẽ jõpãã suasĩra jĩ
1SG ERG 1SG.A-CAUS-3SG.O-eat 1SG child peccary meat
'I made my child eat peccary meat.' (Bardagil 2018: 174)

[5]Transitive verbs require a periphrastic construction for creating a causative semantic relation (Bardagil 2018).

4 Analysis

4.1 The Panãra verb

To model the patterns in Tables 3 and 4 with the Grammar Matrix customization system's morphotactics library (Goodman 2013), we created a position class for each of these 'slots' in a verb's inflection pattern. The position classes serve to determine the order that the morphemes appear in relation to the verb root. Figure 1 illustrates these chained position classes for both transitive and intransitive verbs. Within each of these position classes are multiple lexical rule types (LRTs), one for each person/number combination distinguished in Panãra, with each instantiated by a lexical rule instance.

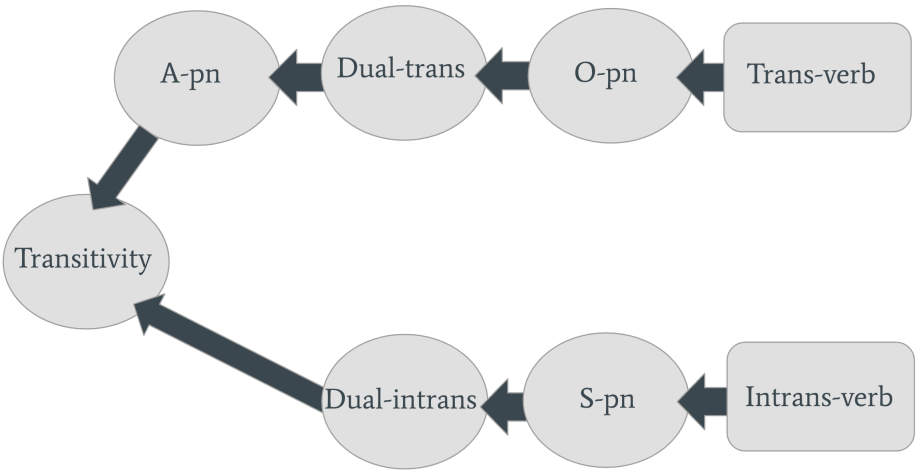


Figure 1: Chained verbal position classes

Although in descriptive work, the S and O verbal prefixes are grouped together in the absolutive, it is necessary for us to model them with separate position classes here. The LRTs in the S-pn position class constrain agreement information on the *subject* while those in the O-pn position class do so for the *object*. Therefore, in an HPSG analysis, they are not the same morpheme.

The Transitivity position class at the end of the chain accepts both the output of the A-pn[6] and dual-intransitive position classes as its input. An LRT for transitive verbs requires a transitive argument structure with ergative case on the first argument and absolutive on the second, and contributes no affix.[7] For intransitive verbs, the LRT applies the intransitive prefix *jy-* and contributes an argument-structure constraint of a single, absolutive argument.

[6]A-pn is the name of the position class for the verbal morpheme that agrees with the agent in person and number. The same naming pattern extends to the O-pn and S-pn position classes, which agree with the person and number information of the object and subject, respectively.

[7]In this paper we describe a grammar as built; however, there is another possible

4.2 Towards a concise implementation of the causative

In analyzing the verbal inflection position classes implemented for the intransitive and transitive, we saw potential for a concise and accurate representation of the causative operation as a change from an intransitive verb to a transitive.

In the implementation described thus far, the S-pn and A-pn position classes will never apply to the same verb, since they have mutually exclusive requirements for their inputs. The S-pn position class takes only uninflected intransitive verb stems as input, while the A-pn position class takes inflected transitive verbs from the dual-transitive position class as input. Although they fill different semantic roles in this language, S and O arguments share a number of properties. They have the same orthographic forms for both full noun phrases and pronouns, as well the same absolutive verbal agreement prefix paradigm across person and number, as seen in Tables 1 and 2. The S-pn and O-pn position classes are also the first that the verb stem goes through — for intransitive and transitive verbs, respectively.

We analyzed the Panãra causative as a 'switch' midway through the verbal inflection from the intransitive verb's chain of position classes to the transitive verb's chain. This switch is triggered by the *ho-* morpheme, which occurs between the S-pn and dual position classes. Figure 2 offers a visual representation of this analysis.

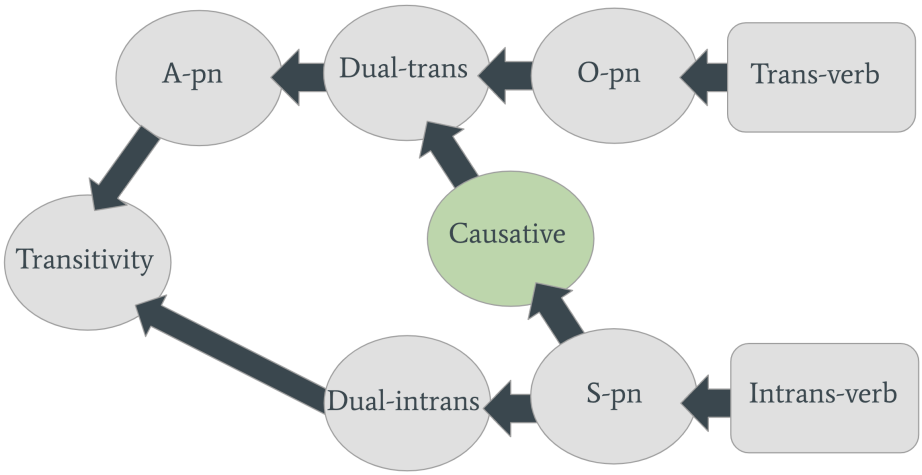


Figure 2: Panãra's causative as a 'switch' midway through the verbal inflection

analysis of the Transitivity position class. In this analysis, constraints on the case of arguments are provided by the verb lexical types, and only intransitive verb roots would need to be inflected with the (In)Transitivity position class, since transitive verbs remain unmarked. The motivation for both types of verbs to go through the Transitivity position class is an additional mood inflection in development that both verb types are subject to. In the alternative analysis, the mood position class could have instead taken as an input the (In)Transitivity and Dual-Transitive position classes.

To model this phenomenon, we added a new optional[8] verbal inflection position class called Causative. This position class takes as its input the output of the S-pn position class, and appears as the prefix *ho-* on the current left edge of the verb complex in the inflection process. There is one LRT in this position class, which specifies that the subject must have ergative case. To employ the work of the valence-changing operation library (Curtis 2018), we added a subject-adding valence-changing operation to the lexical rule. The feature structure in Figure 3 illustrates the constraints describing the subject-adding operation, as provided by Curtis's library.

$$\left[\begin{array}{l} \textit{causative-to-arg2-itr-op-lex-rule} \\ \text{C-CONT} \mid \text{RELS} \mid \text{LIST} \left\langle \left[\text{PRED} \quad \text{cause\_rel}\right] \right\rangle \\ \text{SYNSEM..VAL} \left[\begin{array}{ll} \text{SUBJ} & \left\langle \left[\begin{array}{ll} \text{CASE} & \text{erg} \\ \text{SPR} & \langle\,\rangle \\ \text{COMPS} & \langle\,\rangle \end{array}\right] \right\rangle \\ \text{COMPS} & \left\langle \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{VAL} & \boxed{2} \end{array}\right] \right\rangle \end{array}\right] \\ \text{DTR} \mid \text{SYNSEM..VAL} \left[\begin{array}{ll} \text{SUBJ} & \left\langle \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{VAL} & \boxed{2} \end{array}\right] \right\rangle \\ \text{COMPS} & \langle\,\rangle \end{array}\right] \end{array}\right]$$

Figure 3: Subject-adding valence-changing inflectional rule

Information for both the valency and index is copied from the daughter's SUBJ list to the output's COMPS list. The resulting subject must have ergative case, as is consistent with transitive subjects. The non-empty SUBJ and COMPS lists in the output ensure the desired valency. Finally, this rule contributes a PRED value of caus_rel in the MRS (Copestake et al. 2005). When defining a rule that will contribute semantic information to the sentence, a Grammar Matrix user can choose the PRED value; we chose caus_rel because the new semantic information is contributed by the causative morpheme.

[8]The morphotactics library allows position classes to either be optional or obligatory. Obligatory position classes must be instantiated in all forms that pass through their inflection path (though possibly by a non-affixing rule). Optional position classes may be skipped and should not include non-affixing rules. Since not all intransitive verbs will be causativized, this position class is optional.

4.3 Implementation

Using the LinGO Grammar Matrix (Bender et al. 2002) as a starting point, we modeled the grammar as described by Bardagil (2018) with the Grammar Matrix customization system (Bender et al. 2010). The analysis and corresponding implementation detailed above produces the desired behavior for causative constructions, while ruling out ungrammatical structures.

After going through the Causative position class, and switching to the transitive path of position classes, the argument marked by the prefix closest to the root can no longer be interpreted as an S, but rather must be O. Consequently, the inflected verb is prevented from incorrectly taking the intransitive prefix *jy-*, which can only appear when the subject is absolutive, per the constraints of the LRTs in the Transitivity position class. The parse tree in Figure 4 illustrates a successful implementation of our analysis, using sentence (7)[9] as an example.

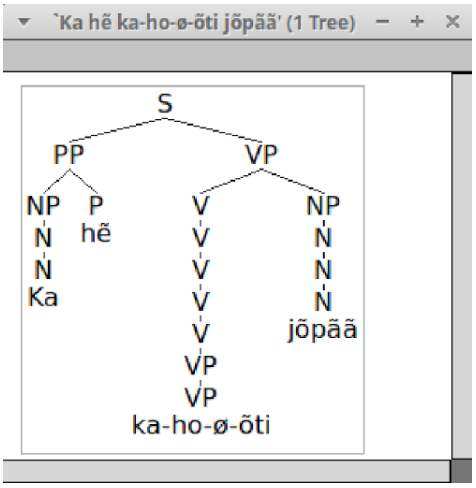


Figure 4: LKB output: Parse tree of sentence (7)

(7) Ka hẽ kahosõti jõpãã.
Ka hẽ ka-ho-ø-sõti jõpãã
2SG ERG 2SG.A-CAUS-3SG.O-sleep child
'You made the child sleep.' (Based on Bardagil 2018: 108)

The default MRS of a causative structure output by the valence-changing operation library (Curtis 2018) produces a valid semantic representation for the sentence in Figure 5.[10] The Causative position class contributes a cause

[9]The original data from (Bardagil 2018) was slightly altered to remove the possessive relation of 'your child' and instead use 'the child' because inalienable possession was not implemented in the grammar.

[10]This analysis treats cause_rel as a three-place relation. A two-place relation could

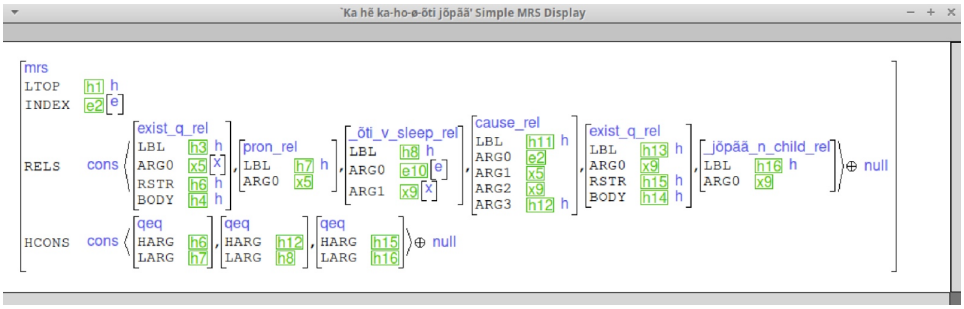


Figure 5: LKB output: MRS representation of sentence (9)

relation to the RELS list. The ARG0 of the O argument, _jõpãã_n_child_rel, is identified with the ARG1 of the verb and the ARG2 of the cause relation. The ARG0 of the pronominal A argument is identified with the ARG1 of the cause relation.

5 Validation

Over a period of 10 weeks, in the context of our coursework, we curated a test suite containing grammatical and ungrammatical sentences in the language. The final test suite includes 230 items, aimed at a broad sampling of over various grammatical features, including agreement, clausal complements, wh-questions, and valence-changing operations.

We used the [incr tsdb()] grammar profiling software (Oepen & Flickinger 1998) for measuring the grammar's coverage and overgeneration. Over 127 grammatical test suite items, our grammar had 84.3% coverage (107/127). Over 103 ungrammatical test suite items, it has 13.6% overgeneration (14/103). The average number of parses per parsed item was 1.44.[11]

It is important to note that 142 of these test suite items were examples that we constructed in order to isolate specific phenomena of interest and to include only phenomena that could be handled by the grammar during its incremental development. Each author-constructed example is based on the data and analysis from Bardagil (2018), but has not yet been vetted by speakers of the language.

There were 8 test suite items (3 grammatical and 5 ungrammatical) constructed specifically to test the valence-changing operation analysis and implementation, including examples (5) and (7) from this paper. Specifically,

also be implemented, if further empirical work with Panãra speakers shows it is more accurate.

[11] Grammatical sentences which didn't parse stemmed from a number of phenomena not yet or not fully implemented such as constraints on evidentials, predicate nouns and adjectives, sentence coordination, and clausal complements. Ungrammatical sentences which did parse did so due to under-constraints on wh-pronouns and adjectives.

the ungrammatical constructions were created to ensure that Causative morpheme does not co-occur with the *jy-* intransitive prefix, that subjects receive ergative case marking, and that the corresponding pronominal prefix has the ergative form while the resulting object pronominal has the absolutive form. Finally, the test suite confirms that underlyingly transitive verbs cannot be inflected with the causative morpheme. On these items, the grammar had 100% coverage and 0% overgeneration; our implementation was successful, with no added ambiguity.

6 Conclusion

The support for modeling morphology, including valence changing morphology, based on the notion of position classes in the Grammar Matrix customization system, correctly predicts the interaction of morphemes for causative constructions in Panãra. The implementation of chained verbal inflection position classes realized our analysis of causatives as a jump between the inflection patterns of intransitive to transitive verbs. Since the O-pn position class is incompatible with intransitive verbs and attaches before the causative morpheme, we take advantage of the orthographic and syntactic parallels between S and O agreement affixes to create a construction with two core arguments and a Causative position class that moves the verbal inflection from intransitive rules to transitive rules. Thus, intransitive verb stems which encounter the causative are able to fully inflect with two argument agreement affixes and the correct transitivity and case-marking morphology. We believe this analysis and implementation succinctly represents the causative operation in Panãra.

This implemented grammar of Panãra can serve as a tool for further testing of phenomena in the language. The analysis of verbal agreement, transitivity marking, and causatives that we implemented is not only internally consistent but also interacts correctly with the other analyses implemented in the grammar so far. Further analyses of additional phenomena of Panãra can thus be tested for consistency with these analyses through implementation.

Looking cross-linguistically, the implementation of the causative morpheme described in this paper supports the typological applicability of the morphotactics and valence-change libraries in the Grammar Matrix customization system.

References

Bardagil, Bernat. 2018. *Case and agreement in Panará*. University of Groningen. (Doctoral dissertation).

Bender, Emily M., Scott Drellishak, Antske Fokkens, Laurie Poulson & Safiyyah Saleem. 2010. Grammar Customization. *Research on Language and Computation* 8(1). 23–72. DOI: 10.1007/s11168-010-9070-1.

Bender, Emily M., Dan Flickinger & Stephan Oepen. 2002. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-linguistically Consistent Broad-Coverage Precision Grammars. In *COLING-02: Grammar Engineering and Evaluation*. https://aclanthology.org/W02-1502.

Copestake, Ann. 2002a. Definitions of typed feature structures. In Stephan Oepen, Dan Flickinger, Jun-ichi Tsujii & Hans Uszkoreit (eds.), *Collaborative language engineering*, 227–230. Stanford, CA: CSLI Publications.

Copestake, Ann. 2002b. *Implementing Typed Feature Structure Grammars*. CSLI Publications.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language and Computation* 3. 281–332. DOI: 10.1007/s11168-006-6327-9.

Curtis, Christian Michael. 2018. *A Parametric Implementation of Valence-changing Morphology in the LinGO Grammar Matrix*. University of Washington. (MA thesis).

Goodman, Michael Wayne. 2013. Generation of Machine-Readable Morphological Rules from Human-Readable Input. *University of Washinton Working Papers in Linguistics* 30. http://depts.washington.edu/uwwpl/vol30/goodman_2013.pdf.

Lapierre, Myriam. 2023. The Phonology of Panãra: A Segmental Analysis. *International Journal of American Linguistics* 89(2). 183–218. DOI: 10.1086/723642.

Müller, Stefan, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.). 2024. *Head-Driven Phrase Structure Grammar: The handbook*. 2nd edn. (Emirically Oriented Theoretical Morphology and Syntax 9). Berlin: Language Science Press. DOI: 10.5281/zenodo.13637708.

Oepen, Stephan & Dan Flickinger. 1998. Towards systematic grammar profiling: Test suite technology ten years after. *Computer Speech and Language: Special Issue on Evaluation* 12. 411–436.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Zamaraeva, Olga, Chris Curtis, Guy Emerson, Antske Fokkens, Michael Goodman, Kristen Howell, T.J. Trimble & Emily M. Bender. 2022. 20 years of the Grammar Matrix: Cross-linguistic hypothesis testing of increasingly complex interactions. *Journal of Language Modelling* 10(1). 49–137. DOI: 10.15398/jlm.v10i1.292.

An HPSG account for German numeral classifiers

Chenyuan Deng
Humboldt-Universität zu Berlin

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 88–103

Keywords: German, Classifier, HPSG

Deng, Chenyuan. 2024. An HPSG account for German numeral classifiers. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 88–103. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.6.
CC-BY 4.0

Abstract

This paper analyzes German numeral classifier phrases within HPSG, focusing on their morphosyntax and semantics. I differentiate between sortal and measure classifiers, which, despite being structurally identical, contribute differently semantically. In a German classifier phrase ([Num CL NP]), the classifier serves as the head, with the counted NP as its complement and requires a specifier to form a complete classifier phrase. This specifier can be realized by a determiner in the traditional sense but also by a numeral. Additionally, numerals are treated as underspecified for a specifier or a modifier, allowing for a flexible combination of determiner, numeral, and other modifier elements in the structure.

1 Introduction

In contrast to languages with rich classifier systems, such as Chinese, Japanese, or Thai, German would not be considered as a typical classifier language. But indeed there are elements in German that are suspected to be a numeral classifier (see Allan 1977, Lehrer 1986, Krifka 1989, Aikhenvald 2000, Lehmann 2000) such as *Stück* 'SCL'[1], compared to the Mandarin Chinese classifier *tou* 'SCL' below.

(1) a. zwei Stück Vieh
two SCL cattle
'two heads of cattle'

b. liang tou niu
two SCL cattle
'two heads of cattle'

However, research on German classifiers has been relatively marginal, especially regarding their structural analysis. Unlike Mandarin Chinese, where the absence of inflection complicates the debate on branching structure ([Num [CL N]] or [[Num CL] N], see Her & Tsai 2020 and Jiang et al. 2022), the rich variation in German w.r.t. declension offers an interesting perspective on the formal study of classifiers. This paper aims to focus on the morphosyntax and semantics of German classifiers and provides an analysis in the framework of Head-driven Phrase Structure Grammar (HPSG, Pollard & Sag 1994, Müller et al. 2021). Minimal Recursion Semantics (MRS, Copestake et al. 2005) is used for the semantic analysis.

The paper is structured as follows: in Section 2, I provide a brief overview of classifiers in German. Next, *Stück* 'SCL' and *Scheibe* 'slice' are selected as examples of sortal and measure classifiers, respectively, and will be discussed in Sections 3 and 4. These sections will progressively address the combinations of classifiers with counted

[1]The following abbreviations are used in the paper: N=noun, NP= noun phrase, Mod=modifier, Num=numeral, CL=classifier, SCL=sortal classifier, MCL=measure classifier, PL=plural, c–n=count noun, DAT=dative, SG=singular, GEN=genitive, WK=weak, PST=past, AKK=akkusative, NOM=nominative, ST=strong, NEU=neutral, FEM=feminine, MAS=masculine, Det=determiner, Spr=specifier, PASS=passive.

NPs, as well as their interactions with numerals or determiners. In the last section, I draw the conclusions of this paper.

2 Classifiers in German

In a study of German numeral classifiers, the first question that needs to be confronted is: what precisely constitutes a numeral classifier? Drawing on the key properties outlined by Lehrer (1986: 110–115), Craig (1992: 280–282), and Lehmann (2000: 249), a typical numeral classifier: a) combines a cardinal numeral (or a quantifier) and a counted nominal; b) is chosen predominantly based on semantics. By these standards, *Stück* 'SCL' in (1a) can be considered as a classifier. It enables the connection of a numeral and a mass noun.[2] The choice of *Stück* 'SCL' is semantically constrained: *Kuh* 'cow' cannot combine with *Stück* 'SCL' to express a counting sense as in (2).[3] This also demonstrates that classifiers and number-gender systems are not mutually exclusive. Therefore, classifiers in German should also be considered within the broader context of cross-linguistic classifier research.

(2) zwei Stück Kuh
two SCL cow
'two pieces of cow' (not 'two cows')

The discussion of German classifiers typically traces back to the numerative construction: A noun phrase consisting of three members (a numeral, a noun (N1) used as a unit of measurement or counting, and another noun (N2) being measured or counted) is known as a numerative construction (Krifka 1991: 401). Based on the semantic contribution of N1 they can be further divided into six subcategories (Löbel 1986, Krifka 1989, Gunkel et al. 2017).

(3) a. measuring constructions
zwei Liter_{N1} Bier_{N2}
two liter beer
'two liters of beer'

b. container constructions
zwei $\text{Flasche-n}_{\text{N1}}$ Milch_{N2}
two bottle-PL milk
'two bottles of milk'

[2] As noted by an anonymous reviewer referencing Little et al. (2022), some studies suggest that classifiers function to bridge numerals and atomic nouns, i.e., count nouns. Based on this, the reviewer questions whether classifiers truly exist in German, as *Vieh* is not a count noun. However, there is both theoretical and empirical research (Krifka 1995, Chierchia 1998, Craig 1992, Zhou et al. 2024) that generally agrees that classifiers are primarily used with mass nouns, enabling them to be counted.

Given the ongoing debate regarding the distinction between count and mass nouns in prototypical classifier languages such as Chinese, this paper does not adopt this distinction as a cross-linguistic criterion for identifying classifiers. Instead, it focuses on summarizing the nouns following *Stück* in German, emphasizing that *Stück* in a counting (rather than partitive) sense is generally followed by mass nouns.

[3] It is acceptable with a partitive reading.

c. counting constructions
zwei Scheibe-n_{N1} $Brot_{N2}$
two slice-PL bread
'two slices of bread'

d. classifier constructions
zwei $Stück_{N1}$ $Vieh_{N2}$
two SCL cattle
'two heads of cattle '

e. collective constructions
zwei $Hundertschaften_{N1}$ $Polizei_{N2}$
two group.of.hundred police
'two hundred police'

f. kind constructions
zwei Sorte-n_{N1} $Bier_{N2}$
two kind-PL beer
'two kinds of beer'

Among the six categories mentioned, *Liter* 'liter', *Flasche* 'bottle', *Scheibe* 'slice', *Hundertschaften* 'group.of.hundred', and *Sorte* 'kind' all perform additional semantic operations on N2 during the counting process, meaning that N1 specifies the form in which N2 is quantified.

In this paper, I treat N1 that do not contribute extra semantic content as sortal classifiers (see *Stück* 'SCL' in (3d)), and all other types of N1 are considered as measure classifiers (MCL, represented by *Scheibe* 'slice' in (3c)). This paper focuses on the structure of German numeral classifier phrases and will therefore center the discussion on *Stück* 'SCL' and *Scheibe* 'slice.MCL' as prototypical examples of different types of numeral classifiers in German. Other numeral classifiers share the same syntactic structure but differ in aspects such as the morphological variations of N1 and N2.

In addition to the semantic differences, the immediate morphosyntactic difference between (3d) and (3c) is that there is no morphological change[4] in a sortal classifier. *Scheibe* 'slice.MCL' retains its marking as a count noun because of the plural cardinal relation *zwei* 'two', whereas *Stück* 'SCL' has no such markings.[5] Although *Stück* has a plural form *Stücke*, this form does not apply to classifier phrases [Num CL N].

[4] As one reviewer pointed out, while *zwei Stücke Vieh* 'two pieces of cattle' is possible, *Stück* 'SCL' here is not functioning as a classifier but rather as part of a partitive construction. This paper focuses exclusively on classifiers.

[5] Please note that I do not claim the lack of plural marking to be a sufficient condition for identifying sortal classifiers, nor do I assert that all measure classifiers have plural marking. The absence of plural marking in their use as classifiers is one of the important characteristics of sortal classifiers and can be seen as an example of decategorialization (Lehmann 2000: 253). The representative measure classifier analyzed in this paper, *Scheibe* 'slice.MCL', retains plural marking; however, some measure classifiers, such as *Meter* 'meter.MCL' and *Pfund* 'pound.MCL', do not exhibit plural marking. Since this study focuses on *Stück* 'SCL' and *Scheibe* 'slice.MCL' as representatives of different types of classifiers to analyze the structure of classifier phrases, it does not delve further into the subclassification of measure classifiers.

Arguably, the nouns *Stück* 'SCL' and *Scheibe* 'slice.MCL', both of which have grammatical gender (4) and can be employed as classifiers, are located at different stages of grammaticalization, as in (5) from Lehmann (2010: 442-443). Given that their properties as nouns in German remain intact (grammatical gender), the treatment of classifiers as subtypes of *noun* is reasonable.

(4) a. das Stück Vieh
the.NEU SCL cattle
'the head of cattle'

b. die Scheibe Brot
the.FEM slice.MCL bread
'the slice of bread'

(5) generic noun → measure classifier → sortal classifier

In German, nouns are classified as count or mass nouns. This should also be taken into account when integrating classifiers into this system. Both sortal classifiers and measure classifiers can be directly associated with numerals, thus categorizing them as count nouns. Since the lexical meaning of a measure classifier is more pronounced, any measure classifier i.e. a meas(ure)-cl(assifier)-n(oun) is also a n(ouny)-n(oun). Combining the count/mass distinction and classifiers, I assume a hierarchy of nominal HEAD values in Figure 1: *Vieh* 'cattle', *Stück* 'SCL', *Scheibe* 'slice.MCL' and *Tisch* 'table' have the HEAD values of *mass-n*, *sort-cl-n* (sortal-classifier-noun), *meas-cl-n*, and *lex-n* respectively. *n-n* is further divided into *meas-cl-n* and *lex-n* and a lexical rule is assumed to change for instance the *Scheibe* 'slice' (*lex-n*) to *Scheibe* 'slice.MCL' (*meas-cl-n*).

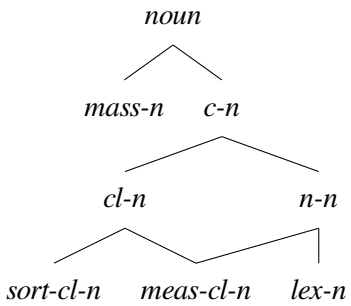


Figure 1: Hierarchy of nominal HEAD values

In general, only mass nouns need to be made 'countable' with the help of other elements, i.e. classifiers in this study, when expressing a quantitative meaning, since count nouns can be used directly in conjunction with numerals. Therefore it is not

surprising that in (3) N2 are nouns without plural inflection.[6] Only *cl-n* can enable mass nouns to appear in a numerative construction, the combination of classifiers (N1, *cl-n*) and mass nouns (N2, *mass-n*) will be discussed in the next section in detail.

3 Combining classifiers and the counted NP

After clarifying the HEAD value of a classifier, this section will discuss the structure of a classifier phrase, beginning with the combination of classifiers and the counted NP.

So far only the case where N1 and N2 are juxtaposed (Kobele & Zimmermann 2012: 265) is mentioned in the paper, meaning that N1 and N2 have the same case inflection, for instance, dative, as in (6a). But there are actually some measure classifiers that allow N2 to be combined with it in genitive case (6b) or with the aid of a preposition *von* 'of' (6c), in other words, N2 is N1's attribute.

(6) a. mit zwei Scheibe-n köstlich-em Brot
with two slice.MCL-PL.DAT delicious-SG.DAT bread
'with two slices of delicious bread'

b. zwei Scheibe-n köstlich-en Brot-es
two slice.MCL-PL delicious-SG.GEN bread-SG.GEN
'two slices of delicious bread'

c. zwei Scheibe-n von diesem köstlich-en Brot
two slice.MCL-PL of this.DAT delicious-DAT.WK bread
' two slices of this delicious bread'

Regarding sortal classifiers, if the classifier phrase functions as the complement of a preposition that requires the dative case, such as *mit* 'with' in (7a), N2 is unequivocally in the dative case. As for N1, the sortal classifier without morphological changes, can be underspecified for the CASE value, which means, it is also dative. Sortal and measure classifiers exhibit the same structural features. In contrast, N2 with prepositions like (7b) involves a partitive *Stück* 'piece' which will not be addressed in this paper.[7]

(7) a. ein Garten mit 300 Stück krank-em Vieh
one garden with 300 SCL sick-SG.DAT cattle
'a garden with 300 head of sick cattle'

[6]Indeed, the number of N2 varies according to the detailed subtypes of the classifier's have. In the case of measuring structures, counting constructions, and classifier constructions (Krifka 1989: 12), N2 must be a mass noun, but the HEAD value of N2 is underspecified in the case of container constructions and collective constructions. In this paper, I treat *Scheibe* 'slice.MCL' as a representative of most kinds of measure classifiers, taking into account the semantic differences between sortal and measure classifiers.

[7]All referenced and marked newspaper examples cited in this article are drawn from the German Reference Corpus (DeReKo).

b. ein-em gut-en Stück vom Vieh des Rittergut-es
one-DAT good-DAT.WK piece of.DET cattle DET.GEN manor-GEN
‘ a good piece of the manor’s cattle’
(Braunschweiger Zeitung, 23.03.2009)

Considering the other semantic components (Löbel 1986: 77–87) brought by (6b), (6c), and (7b), I follow Krifka (1989: 15), limiting myself to the juxtaposed structure (6a) and (7a) that is more focused on the function of quantitative information.[8]

3.1 Headedness of a classifier phrase

Since N1 and N2 are combined juxtaposed, this inevitably brings up the discussion of the headedness of a classifier phrase, that is an NP. The case of measure classifiers like *Scheibe* ‘slice.MCL’ is more straightforward, since the verb and *Scheibe* ‘slice.MCL’ (the subject) agree in number, that is plural in (8).

(8) a. Auf dem Teller der Frau lieg-en zwei kleingeschnitten-e
one the.DAT plate the.GEN woman lie-PL two chopped-NOM.WEAK
Scheibe-n Brot
slice.MCL-PL bread
‘On the woman’s plate are two slices of bread that have been cut into small pieces.’
(Braunschweiger Zeitung, 21.05.2010)

b. 1955 kam-en 0,82 Stück Vieh auf einen Einwohner
1995 come.PST-PL 0.82 SCL cattle on one.AKK inhabitant
‘In 1955, there were 0.82 head of cattle per inhabitant.’
(St. Galler Tagblatt, 24.01.1998)

As for sortal classifiers, although *Stück* ‘SCL’ does not exhibit morphological inflection to provide direct evidence of its status as the head in (8b), its role as the head is evident, as *Vieh* ‘cattle’ cannot be plural. Furthermore, since phrases like *ein- nach d- ander-* ‘one after the other’ in (9) can only refer in gender to *Stück* (neutral) and not to *Rhabarber* (masculine). Thus, it can be concluded that *Stück* ‘SCL’ is the head of the classifier phrase. (For more on this test, see Höhle 2019: 52 and Müller 2002: 49.)

(9) a. Drei Stück Rhabarber wurden eines nach dem
three SCL.NEU rhubarb.MAS PASS.PST one.NOM.NEU after the.DAT.NEU
anderen klein geschnitten.
other.DAT.NEU small chopped
‘Three pieces of rhubarb were cut into small pieces one by one.’

[8] As noted by one reviewer, (6b) represents a regular instance of a partitive measure construction. As previously mentioned, this paper does not address the partitive reading of *Stück* ‘SCL’ and maintains a reserved stance on the relationship between partitive and counting readings. For readers interested in further exploration, Zimmer (2015) and Schäfer (2018) provide a detailed discussion.

b. * Drei Stück Rhabarber wurden einer nach dem
three SCL.NEU rhubarb.MAS PASS.PST one.NOM.MAS after the.DAT.MAS
anderen klein geschnitten.
other.DAT.NEU small chopped
Int: 'Three pieces of rhubarb were cut into small pieces one by one.'

The alternative, where *Vieh* 'cattle' is the head of the NP, is not possible. If that were the case, one would have to make a mass noun open to the numerals, which is subversive to the basic logic of German grammar: only count nouns can be used directly with numerals. In our case, *cl-n* is a subtype of count nouns, so *Stück* 'SCL' can be the head of the NP and all is safe.

Thus, morphosyntactically, the classifier is always the head of the German classifier phrase (NP). But semantically there is a problem that requires attention. There are examples in the DeReKo (*Das Deutsche Referenzkorpus* 'The German reference corpus') where *Stück* 'SCL' can be preceded by adjectives modifying N2, see (10). As a sortal classifier, *Stück* does not provide an extra lexical contribution, and it's only the *Vieh* 'cattle' that gets sick or dies.[9]

(10) a. ein tot-es Stück Vieh
one dead-ST.NOM/AKK.NEU SCL cattle
'a dead head of cattle'
(Neue Zürcher Zeitung, 29.05.2004)

b. ein-em krank-en Stück Vieh
one-DAT.NEU sick-WK.DAT SCL cattle
'a sick head of cattle.'
(die tageszeitung, 02.10.2009)

If *Stück* 'SCL' is the head, then this semantic connection of modifiers preceding the classifier and N2 needs to be possible. This can be regarded as a special feature of sortal classifiers as "functional" nouns: In CONT (11a), a sortal classifier takes the IND of N2 as its argument and shares the IND of N2. If there is an adjective modifying N1, it still modifies the IND of N2.[10] This is an important difference between sortal and measure clasifiers. A measure clasifier will also take the IND of N2 as the argument of

[9]The adjectives modifying measure classifiers before the measure classifier have a different meaning, as in (i). In (i.b) it is still about a counting-*Stück*, whereas *Stück* in (i.a) is partitive, hence (i.a) will not be discussed in this paper.

(i) a. ein groß-es Stück Schokolade
one big-NEU.ST.SG SCL chocolate
'a large piece of chocolate'

b. ein süß-es Stück Schokolade
one sweet-NEU.ST.SG SCL chocolate
'a sweet piece of chocolate'

[10]A similar method of index inheritance can be found in Bender & Siegel (2005: 631) and Levine (2010: 271) about the analysis of parasitic heads.

for instance *scheibe-rel* (11b), but it still has its own distinct IND that can be modified by other adjectives, such as *kleingeschnitten* 'chopped' in (8).

(11) a. Sample CONT of a sortal classifier

$$\left[\text{CONT}\left[\begin{array}{ll}\text{IND} & \boxed{1}\\ \text{RELS} & \left\langle\left[\begin{array}{ll}\textit{stück-rel}\\ \text{ARG} & \boxed{1}\end{array}\right]\right\rangle\end{array}\right]\right]$$

b. Sample CONT of a measure classifier

$$\left[\text{CONT}\left[\begin{array}{ll}\text{IND} & \boxed{1}\\ \text{RELS} & \left\langle\left[\begin{array}{ll}\textit{scheibe-rel}\\ \text{ARG}0 & \boxed{1}\\ \text{ARG}1 & \boxed{2}\end{array}\right]\right\rangle\end{array}\right]\right]$$

After clarifying the headedness question of a classifier phrase, it is possible to give a more detailed answer to the juxtaposed combination. N1 and N2 seem to be combined by juxtaposing, but in fact, this combination should be realized through a *head-complement-phrase*, N1 is the head and N2 is the complement. An important point that distinguishes *cl-n* from *lex-n* is that classifiers need a noun to fulfill their counting function. And this noun (N2) needs to share the same CASE value as the classifier (N1), see (12).

(12) Lexical entry for *Stück* 'SCL' (preliminary version without Num and Det)

$$\left[\begin{array}{ll}\text{CAT} & \left[\begin{array}{ll}\text{HEAD} & \left[\begin{array}{ll}\text{CONC} & \left[\text{CASE}\ \boxed{3}\right]\\ \textit{sort-cl-n}\end{array}\right]\\ \text{COMPS} & \left\langle \text{NP}\left[\begin{array}{ll}\text{HEAD} & \left[\begin{array}{ll}\text{CONC} & \left[\text{CASE}\ \boxed{3}\right]\\ \textit{mass-n}\end{array}\right]\\ \text{IND} & \boxed{1}\end{array}\right]\right\rangle\end{array}\right]\\ \text{CONT} & \left[\begin{array}{ll}\text{IND} & \boxed{1}\\ \text{RELS} & \left\langle\left[\begin{array}{ll}\textit{stück}\\ \text{ARG} & \boxed{1}\end{array}\right]\right\rangle\end{array}\right]\end{array}\right]$$

As shown in Figure 1, the hierarchy of nominal HEAD values has been applied here. It is worth noting that, count and mass are typically considered semantic features and are generally not modeled under HEAD. However, it is necessary for *Stück* 'SCL' to choose a complement of type *mass-n* in this case. Without the distinction between count and mass, a nominal headed phrase such as *zwei Stück Vieh* could still be selected by another *Stück* 'SCL', as in (13), which is an undesired outcome.

(13) * zwei Stück zwei Stück Vieh
two SCL two SCL cattle

4 Adding numerals and determiners

So far it seems that when N1's COMPS is satisfied by N2, it is already a complete NP. But in fact, a classifier, as a count noun, no matter sortal or measure, cannot be directly selected by the verb unless it is preceded by a numeral or a determiner, i.e., (3) and (4). Therefore, numerals or determiners are required. But the syntactic behavior of *ein-* 'one' and other numerals differs when modifiers are involved, and next I will discuss them separately.

4.1 Case 1: when Num is other than *ein-* 'one'

German numerals such as *zwei* 'two' are traditionally treated as adjectives in the *Duden – Die Grammaik* (2022). However in analysis of other classifier languages, numerals are generally treated as a specifier or complement of the head classifier (Bender & Siegel 2005, Ng 1997). This implies that the connection of a numeral to a classifier is specific and restricted. But when a definite article or demonstrative is present, as in (14), two specifiers are required: one for the definite article and one for the numeral. Classifier phrases are not particularly common in German, and there is no supporting evidence or alternative structures in German to justify a double specifier treatment; therefore, this analysis may not be the most suitable for the German data.

(14) die zwei Scheibe-n Brot
the two slice.MCL-PL bread
'these two slices of bread'

Furthermore, German numerals other than *ein-* 'one', which means the NUM value of the numeral is always *pl*, can have a flexible position before the noun and can be exchanged in order with a modifier without affecting the truth condition of the phrase, see (15a) and (15b). If I insist on maintaining the double specifier treatment and consider the numeral as a specifier, technically (15b) would not pose a problem because the order of Spr-Mod-Spr is permissible. However, this would also predict phrases like (16), which are not possible in German.

(15) a. diese zwei trocken-en Scheibe-n Brot
these two dry-WK.PL slice.MCL-PL bread
'these two dry slices of bread'

b. diese trocken-en zwei Scheibe-n Brot
these dry-WK.PL two slice.MCL-PL bread
'these two dry slices of bread'

(16) a. * schöne das Buch
nice.WK DET book
Int: 'the nice book'

b. * schöne zwei Bücher
nice.WK two book.PL
Int: 'two nice books'

Therefore, the flexible order of classifiers w.r.t. modifiers actually indicates a modifier status of classifiers in German. Together with the fact that a numeral may not appear within a classifier phrase (4), repeated as (17), I treat numerals other than *ein-* 'one' as a modifier, i.e., an undeclinable adjective.

(17) a. das Stück Vieh
the.NEU SCL cattle
'the head of cattle'

b. die Scheibe Brot
the.FEM slice.MCL bread
'the slice of bread'

Without a definite article or demonstrative, this kind of Mod-flexibility of numerals disappears: numerals can only be placed on the leftmost side of the classifier phrase, comparing (18a) and (18b). In other words, in the absence of a definite determiner, the numeral seems to take over the function of a specifier in the classifier phrase. Just as with every count noun, the classifier phrase is not complete without this specifier, namely the numeral.[11]

(18) a. zwei klein-e Scheibe-n Brot
two small-ST.PL slice.MCL-PL bread
'two small slices of bread'

b. * klein-e zwei Scheibe-n Brot
small-ST.PL two slice.MCL-PL bread
Int: 'two small slices of bread'

The analysis on numerals other than *ein-* 'one' should fulfill both cases (15) and (18): a numeral is a modifier if there is a specifier within a classifier phrase, and the numeral will be the specifier if no other specifier is present. Therefore I assume that the HEAD value of a number is an underspecified type of *num* as in Figure 2, which has two subtypes, *num-det* (numeral-determiner) and *num-adj* (numeral-adjective).

num-det is at the same time a subtype of *det.* In German the specifier of a classifier can be: *dies-* 'this', *d-* 'the', *mein-* 'my', *zwei* 'two', *ein-* 'one', *jede-* 'every' as in (19). A Hierarchy of HEAD values for determiners is proposed in Figure 3.

[11] The difference is that in the case of a sortal classifier, there is no bare plural, a specifier is always required. But *Scheibe* 'slice.MCL' allows a bare plural and there would be an optional specifier.

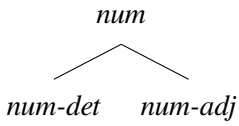


Figure 2: Hierarchy of HEAD values for numerals

(19) a. *dem-def-cl-d*
dieses Stück Vieh
'this head of cattle'

b. *art-def-cl-d*
das Stück Vieh
'the head of cattle'

c. *poss-def-cl-d*
unsere 8 Stück Vieh
'our 8 heads of cattle'

d. *num-det*
zwei Stück Vieh
'two heads of cattle'

e. *art-ind-cl-d*
ein Stück Vieh
'a head of cattle'

f. *jed-ind-cl-d*.
jedes Stück Vieh
'every head of cattle'

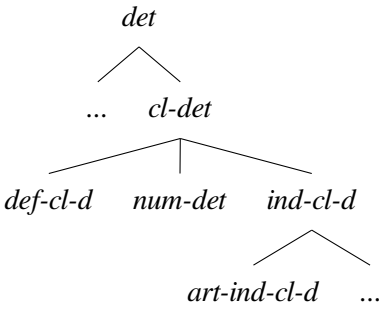


Figure 3: Hierarchy of HEAD values for determiners

Now it is necessary to include the information of a specifier in the entries of a classifier, i.e., a word with HEAD value *cl-n* not only needs to select a *mass-n* as its complement, but must also select a determiner of type *cl-det* to be its specifier. The CONC value of the specifier needs to be shared with its head.

All numerals with [HEAD *num-adj*], whose NUM value is *pl*, cannot co-occur with singular *Stück* 'SCL' or *Scheibe* 'slice.MCL'. Due to the internal agreement of NP, the combination of *ind-cl-d* and *num-adj* in cases such as (20) is ruled out. [NUM *sg*] only appears when the numeral is *ein-* 'one', that is, a *num-det*. Two numerals, no matter whether the combination of *num-det* and *num-adj* or iteration of *num-adj*, are not possible for semantic reasons: there should be only one *card-rel* per *index*.

(20) * ein / jedes zwei Stück Vieh
one every two SCL cattle
Int: 'a / every two head of cattle'

This underspecified treatment of *num* is further supported by empirical evidence. Both (21a) and (21b) are German expressions for *a theme of this year*, with *dies-* in (21a) being declined as a determiner, and (21b) being declined as an adjective analog to (21c). An underspecified HEAD value of *dies-* to be *det* or *adj* is expected or at least two entries are necessary.

(21) a. ein Thema dies-es Jahr-es
a theme this.DET-GEN.ST year-GEN
'a theme of this year'

b. ein Thema dies-en Jahr-es
a theme this.ADJ-GEN.ST year-GEN
'a theme of this year'

c. ein Thema letzt-en Jahr-es
a theme last-GEN.ST year-GEN
'a topic of last year'

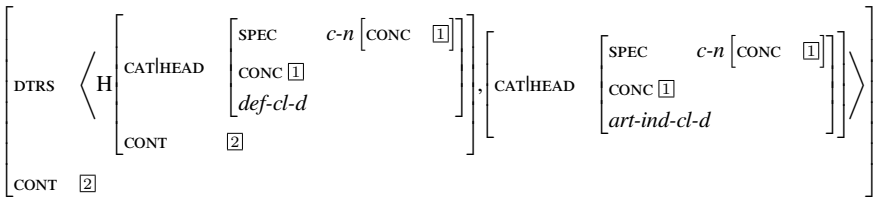
4.2 Case 2: iff Num=1

When the numeral is 1, i.e. *ein-* 'one', it has only the properties of a determiner. Even if there is already a definite determiner, the numeral *ein-* 'one' cannot change places with an adjective and must be fixed in the second position after the definite determiner, see (22). In this case, there exist two determiners tightly tied together and nothing can be inserted between them. Thus I assume a compound structure *compl-det* combining a *def-cl-d* and *art-ind-cl-d*.

(22) a. diese ein-e klein-e Scheibe Brot
that one.WK.SG small-WK.SG slice.MCL bread
'this one small slice of bread'

b. * diese kleine eine Scheibe Brot
that small-WK.SG one.WK.SG slice.MCL bread
'this one small slice of bread'

(23) *compl-det* ⇒



This complex determiner has *def-cl-d* as its head and can be selected by a *c-n* (not only by a *cl-n*), which correctly predicts (24) in German.

(24) das eine Buch
that one.WK.SG buch
'the one book'

5 Conclusions

This paper provides an HPSG analysis for German numeral classifiers. Based on the semantic differences and morphological behavior, numeral classifiers in German can be categorized into sortal and measure classifiers. Both types of classifiers take N2, the counted NP as their complement, and require a specifier to form a complete classifier phrase. This specifier can be realized by a determiner in the traditional sense but also by a numeral. In plural cases, the numeral can function as a modifier if an additional determiner is present. An underspecified HEAD value of numerals is proposed allowing for the combination of Det-Mod-Num. When the numeral is *ein-* 'one', a *comp-det* is introduced to ensure that nothing can be inserted between these two determiners.

Acknowledgements

I would like to thank Marc Felfe for his judgment on German examples and thank Antonio Machicao y Priemer, Stefan Müller, Giuseppe Varaschin, the two anonymous reviewers, as well as Francis Bond, Berthold Crysmann, Jakob Maché, and other participants in the 31st International Conference on Head-Driven Phrase Structure Grammar for comments on earlier versions of this paper.

References

Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

Allan, Keith. 1977. Classifiers. *Language* 53(2). 285–311. DOI: 10.1353/lan.1977.0043.

Bender, Emily M. & Melanie Siegel. 2005. Implementing the syntax of Japanese numeral classifiers. In Keh-Yih Su, Jun'ichi Tsujii, Jong-Hyeok Lee & Oi Yee Kwong (eds.), *Natural language processing – IJCNLP 2004*, 626–635. Berlin: Springer. DOI: 10.1007/978-3-540-30211-7_66.

Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural Language Semantics* 6. 339–405. DOI: 10.1023/A:1008324218506.

Copestake, Ann, Daniel P. Flickinger, Carl J. Pollard & Ivan A. Sag. 2005. Minimal Recursion Semantics: An introduction. *Research on Language and Computation* 3(4). 281–332. DOI: 10.1007/s11168-006-6327-9.

Craig, Colette. 1992. Classifiers in a functional perspective. In Michael Fortescue, Peter Harder & Lars Kristoffersen (eds.), *Layered structure and reference in a functional perspective* (Pragmatics and Beyond New Series 23), 277–301. Amsterdam: John Benjamins Publishing. DOI: 10.1075/pbns.23.12cra.

Gunkel, Lutz, Adriano Murelli, Susan Schlotthauer, Bernd Wiese & Gisela Zifonun. 2017. *Grammatik des Deutschen im europäischen Vergleich: das Nominal*. Berlin: Walter de Gruyter GmbH & Co KG. DOI: 10.1515/9783110341461.

Her, One-Soon & Hui-Chin Tsai. 2020. Left is right, right is not: On the constituency of the classifier phrase in Chinese. *Language and Linguistics* 21(1). 1–32. DOI: 10.1075/lali.00046.her.

Höhle, Tilman N. 2019. Topologische felder. In Stefan Müller, Marga Reis & Frank Richter (eds.), *Beiträge zur deutschen Grammatik. Gesammelte Schriften von Tilman N. Höhle. Zweite durchgesehene Auflage* (Classics in Linguistics 5), 7–89. Berlin: Language Science Press. DOI: 10.5281/zenodo.2588347.

Jiang, Li Julie, Peter Jenks & Jing Jin. 2022. The syntax of classifiers in Mandarin Chinese. In Chu-Ren Huang, Yen-Hwei Lin, I-Hsuan Chen & Yu-Yin Hsu (eds.), *The Cambridge Handbook of Chinese linguistics* (Cambridge Handbooks in Language and Linguistics), 515–549. Cambridge: Cambridge University Press. DOI: 10.1017/9781108329019.028.

Kobele, Gregory M & Malte Zimmermann. 2012. Quantification in German. In Edward L. Keenan & Denis Paperno (eds.), *Handbook of quantifiers in natural language*, 227–283. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-007-2681-9_5.

Krifka, Manfred. 1989. *Nominalreferenz und Zeitkonstitution. Zur Semantik von Massentermen, Pluraltermen und Aspektklassen*. München: Fink.

Krifka, Manfred. 1991. Massennomina. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik. Ein internationales Handbuch der zeitgenössischen Forschung*, 399–417. Berlin: De Gruyter. DOI: 10.1515/9783110126969.6.399.

Krifka, Manfred. 1995. Common nouns: A contrastive analysis of Chinese and English. In Gregory N. Carlson & Francis Jeffry Pelletier (eds.), *The generic book*, 398–411. Chicago: University of Chicago Press.

Lehmann, Christian. 2000. On the German numeral classifier system. In Chris Schaner-Wolles, John R. Rennison & Friedrich Neubarth (eds.), *Naturally! Linguistic studies in honour of Wolfgang Ulrich Dressler presented on the occasion of his 60th birthday*, 249–253. Torino: Rosenberg & Sellier.

Lehmann, Christian. 2010. On the function of numeral classifiers. In Franck Floricic (ed.), *Essais de typologie et de linguistique générale. Mélanges offerts à Denis Creissels*, 435–445. Lyon: École Normale Supérieure.

Lehrer, Adrienne. 1986. English classifier constructions. *Lingua* 68(2-3). 109–148. DOI: 10.1016/0024-3841(86)90001-X.

Levine, Robert D. 2010. The ass camouflage construction: Masks as parasitic heads. *Language* 86(2). 265–301. DOI: 10.1353/lan.0.0212.

Little, Carol-Rose, Mary Moroney & Justin Royer. 2022. Classifiers can be for numerals or nouns: Two strategies for numeral modification. *Glossa: a journal of general linguistics* 7(1). DOI: 10.16995/glossa.8437.

Löbel, Elisabeth. 1986. *Apposition und Komposition in der Quantifizierung. Syntaktische, semantische und morphologische Aspekte quantifizierender Nomina im Deutschen*. Berlin: De Gruyter. DOI: 10.1515/9783111352565.

Müller, Stefan. 2002. *Complex predicates: verbal complexes, resultative constructions, and particle verbs in German* (Studies in Constraint-Based Lexicalism 13). Stanford: Center for the Study of Language & Information. https://hpsg.hu-berlin.de/~stefan/Pub/complex.html.

Müller, Stefan, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.). 2021. *Head Driven Phrase Structure Grammar: The handbook* (Empirically Oriented Theoretical Morphology and Syntax 9). Berlin: Language Science Press. DOI: 10.5281/zenodo.5543318.

Ng, Say Kiat. 1997. *A double-specifier account of Chinese NPs using Head-driven Phrase Structure Grammar*. University of Edinburgh. (MSc Thesis).

Pollard, Carl J. & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar* (Studies in Contemporary Linguistics). Chicago, IL: University of Chicago Press.

Schäfer, Roland. 2018. Abstractions and exemplars: The measure noun phrase alternation in German. *Cognitive Linguistics* 29(4). 729–771. DOI: 10.1515/cog-2017-0050.

Wöllstein, Angelika & Dudenredaktion. 2022. *Duden – Die Grammatik: Struktur und Verwendung der deutschen Sprache. Sätze – Wortgruppen – Wörter*. Berlin: Dudenverlag.

Zhou, Wei, Guangyan Zhang & Yujie Chen. 2024. Chinese nouns are mass nouns: An information-theoretic computational proof. *Lingua* 311. 103815. DOI: 10.1016/j.lingua.2024.103815.

Zimmer, Christian. 2015. Bei einem Glas guten Wein(es): Der Abbau des partitiven Genitivs und seine Reflexe im Gegenwartsdeutschen. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 137(1). 1–41. DOI: 10.1515/bgsl-2015-0001.

Examining Delayed Complements in Norwegian within an Incremental Left-Branching Grammar Framework

Petter Haugereid
Western Norway University of Applied Sciences

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 104–117

Keywords: Norwegian, HPSG, Delayed Complements, Incremental Parsing

Haugereid, Petter. 2024. Examining Delayed Complements in Norwegian within an Incremental Left-Branching Grammar Framework. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 104–117. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.7.
CC-BY 4.0

Abstract

In this paper, I will present an analysis of complement clauses in Norwegian that are licensed by the five adverbs *så*/*såpass* 'so' and *slik*/*sånn*/*sådan* 'such'. It will be assumed that the licensed complement clause, although it is licensed by the adverb, is not a complement within the constituent with the adverb, but rather a complement of the clause. This opens for a uniform analysis of complement clauses licensed by adverbs, irrespective of their position with regard to the licensing adverb. The analysis will be conducted within the framework of an HPSG-inspired incremental typed feature structure grammar of Norwegian.

1 Introduction

A little studied, however not completely infrequent, phenomenon is that of *delayed complement clauses* in examples like (1) from Huddleston & Pullum (2002: 967).

(1) So many people enrolled for the course that we had to move to a larger room.

So is here a degree adverb, modifying a degree determinative *many*, and it requires a complement clause *that we had to move to a larger room*. This complement clause is according to Huddleston & Pullum (2002) always at the end of the clause.

The construction is related to the more common construction where the complement clause appears adjacent to the phrase with *so*, referred to hereafter as the *so-phrase*. This is exemplified for Norwegian in (2a) where the so-phrase *så sen* 'so late' is directly followed by the complement clause *at jeg smiler* 'that I smile'. This construction is semantically equal to the corresponding delayed complement construction demonstrated in (2b). In Norwegian it is always possible to front the so-phrase, and the complement clause then is left behind.

(2) a. Han er så sen at jeg smiler.
he is so late that I smile
'He is so late that I smile.'

b. Så sen er han at jeg smiler.
so late is he that I smile
'He is so late that I smile.'

Probably the most famous sentence with a delayed complement clause is the first sentence of John 3:16 in the Bible, as shown in (3). In this sentence, the so-phrase *så høyt* 'so highly' is thematized, making the complement clause a delayed complement. The degree adverb here modifies an adverb *høyt* 'highly'. Note that in the English translation, the so-phrase consists only of the adverb *so*.

[†]I would like to thank three anonymous reviewers and the audience at the HPSG 2024 conference in Olomouc, Czech Republic, for very useful comments and suggestions.

(3) For så høyt har Gud elsket verden at han ga sin Sønn, den
for so highly has God loved world-DEF that he gave REFL son, the
enbårne [...]
one and only [...]
'For God so loved the world that he gave his one and only Son [...]'

There are also other (degree) adverbs that require complement clauses; *såpass* 'so', *slik* 'such', *sånn* 'such', and *sådan* 'such'. While *så* and *såpass* function as degree adverbs modifying adjectives, adverbs, and prepositions, *slik* and *sånn* modify determiners (example (4a)) or function alone, as adverbs (example (4b)).

(4) a. Med slik en kraft traff den at jeg falt.
with such a force struck it that I fell
'It struck with such a force that I fell.'

b. For slik har vi elsket naturen at vi er nødt til å drepe
for so have we loved nature-DEF that we are obliged to to kill
den.
it
'For we have so loved the nature that we have to kill it.'

There are similar comparative constructions where a subordinate clause or an infinitival clause at the end of a sentence is dependent on a comparative element earlier in the sentence, as shown in (5).[1]

(5) a. More people enrolled for the course than we had expected.

b. Too many people enrolled for the course to fit in the room.

In (5a) the comparative governor *more* licenses the comparative complement *than we had expected* at the end of the clause. In (5b) the comparative governor *too* licenses the comparative complement *to fit in the room*. The difference from the delayed complement clauses like (1), apart from the licensing element, is that while the complement clause in delayed complement clauses is a regular that-clause, the complement in (5a) is a subordinate clause introduced by *than*, and the complement in (5b) is an infinitival clause. It shows that the licensing element has a syntactic requirement for the comparative complement.

However, whereas the complement clause required by *so* must appear at the end of the clause, as illustrated by (6a), the comparative complements licensed by *more* and *too* may appear within the sentence, as shown in (6b) and (6c).

(6) a. * So many people that we had to move to a larger room, enrolled for the course.

b. More people than we had expected enrolled for the course.

c. Too many people to fit in the room enrolled for the course.

[1]Thanks to Dan Flickinger for pointing this out to me.

In this paper, the focus will be on the delayed complement construction illustrated in (1)–(3). However, it will also be outlined how the analysis can be extended to comparative constructions, as illustrated in (5).

2 Corpus Searches

A search for the words *så*/*såpass* 'so' and *slik*/*sånn*/*sådan* 'such' followed by the complementizer *at* 'that' within a window of the following 10 words in the 100 million word Leksikografisk bokmåskorpus (Fjeld et al. 2020) yielded the number of matches shown in Table 1.[2] A manual inspection of the first 50 matches in each search revealed that a significant number of the complement clauses were licensed by the adverb.[3] The total number of complement clauses licensed by the five adverbs is estimated to be about 40,000. The total number of complement clauses with the complementizer *at* in the corpus is 1,025,355. This implies that about 4% of the *at* complement clauses are licensed by an adverb. Among these, about 1,000 (1 of 40) is a delayed complement construction.

	Matches	Manual inspection	Estimate
så . . . at	59,671	29/50	34,609
såpass . . . at	1,346	46/50	1,238
slik . . . at	9,723	19/50	3,694
sånn . . . at	1,260	22/50	554
sådan . . . at	65	10/65	10
Total			40,105

Table 1: Estimated number of complement clauses licensed by adverbs in Leksikografisk bokmåskorpus

3 HPSG Analysis

The ERG (Flickinger 2000) provides an analysis for sentences like (2a) where the complement clause is adjacent to the so-phrase. The complement clause is then treated as a complement of *so*, as shown in Figure 1. The MRS (Copestake et al. 2005) of the sentence is given in Figure 2. It shows how the predicate of *so* (*_so_x_comp*) takes *late* as its first argument (11) and the subordinate clause as its second argument (12).

However, in cases where the so-phrase is nonadjacent to the licensed CP, as in (1), the ERG lacks an analysis where the CP is an argument of *so*. The delayed complement construction poses a challenge for regular HPSG grammars given that the element that selects for the complement clause, *so*, occurs at the top of the tree,

[2]The corpus is not syntactically annotated, so a syntactic search is not possible.

[3]All the 65 matches with *sådan* were manually inspected.

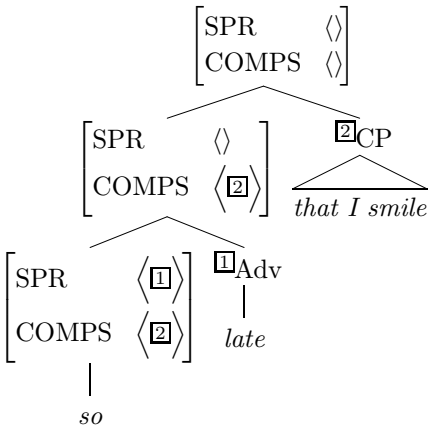


Figure 1: Analysis of *so late that I smile* by the ERG

$$
\begin{bmatrix}
\textit{mrs} & \\
\text{TOP} & \boxed{0}\,h \\
\text{INDEX} & \boxed{2}\,e \\
\text{RELS} & \left\langle
\begin{array}{l}
\begin{bmatrix} \textit{pron} & \\ \text{LBL} & \boxed{4}\,h \\ \text{ARG0} & \boxed{3}\,x \end{bmatrix},
\begin{bmatrix} \textit{pronoun\_q} & \\ \text{LBL} & \boxed{5}\,h \\ \text{ARG0} & \boxed{3}\,x \\ \text{RSTR} & \boxed{6}\,h \\ \text{BODY} & \boxed{7}\,h \end{bmatrix},
\begin{bmatrix} \textit{loc\_nonsp} & \\ \text{LBL} & \boxed{1}\,h \\ \text{ARG0} & \boxed{2}\,e \\ \text{ARG1} & \boxed{3}\,x \\ \text{ARG2} & \boxed{8}\,x \end{bmatrix},
\begin{bmatrix} \textit{\_so\_x\_comp} & \\ \text{LBL} & \boxed{9}\,h \\ \text{ARG0} & \boxed{10}\,e \\ \text{ARG1} & \boxed{11}\,e \\ \text{ARG2} & \boxed{12}\,h \end{bmatrix}, \\
\begin{bmatrix} \textit{time\_n} & \\ \text{LBL} & \boxed{9}\,h \\ \text{ARG0} & \boxed{8}\,x \end{bmatrix},
\begin{bmatrix} \textit{def\_implicit\_q} & \\ \text{LBL} & \boxed{13}\,h \\ \text{ARG0} & \boxed{8}\,x \\ \text{RSTR} & \boxed{14}\,h \\ \text{BODY} & \boxed{15}\,h \end{bmatrix},
\begin{bmatrix} \textit{\_late\_p} & \\ \text{LBL} & \boxed{9}\,h \\ \text{ARG0} & \boxed{11}\,e \\ \text{ARG1} & \boxed{8}\,x \end{bmatrix}, \\
\begin{bmatrix} \textit{pron} & \\ \text{LBL} & \boxed{16}\,h \\ \text{ARG0} & \boxed{17}\,x \end{bmatrix},
\begin{bmatrix} \textit{pronoun\_q} & \\ \text{LBL} & \boxed{18}\,h \\ \text{ARG0} & \boxed{17}\,x \\ \text{RSTR} & \boxed{19}\,h \\ \text{BODY} & \boxed{20}\,h \end{bmatrix},
\begin{bmatrix} \textit{\_laugh\_v\_at} & \\ \text{LBL} & \boxed{12}\,h \\ \text{ARG0} & \boxed{21}\,e \\ \text{ARG1} & \boxed{17}\,x \end{bmatrix}
\end{array}
\right\rangle \\
\text{HCONS} & \left\langle
\begin{bmatrix} \textit{qeq} & \\ \text{HARG} & \boxed{0}\,h \\ \text{LARG} & \boxed{1}\,h \end{bmatrix},
\begin{bmatrix} \textit{qeq} & \\ \text{HARG} & \boxed{6}\,h \\ \text{LARG} & \boxed{4}\,h \end{bmatrix},
\begin{bmatrix} \textit{qeq} & \\ \text{HARG} & \boxed{14}\,h \\ \text{LARG} & \boxed{9}\,h \end{bmatrix},
\begin{bmatrix} \textit{qeq} & \\ \text{HARG} & \boxed{19}\,h \\ \text{LARG} & \boxed{16}\,h \end{bmatrix}
\right\rangle
\end{bmatrix}
$$

Figure 2: MRS of the sentence *He was so late that I laughed.*

as part of a subject or a filler, and even though one would allow for a complement clause to be realized after the the subject or filler, once the phrase containing *so* has been realized, there is no way to access the valence requirements of *so*.

If we assume that the canonical position of the complement clause in a delayed complement construction is adjacent to the so-phrase, as implied in the analysis in Figure 1, the delayed complement construction could be considered a case of extraposition. However, since the complement clause consistently appears at the end of the matrix clause, there is no evidence supporting such an analysis. A version of the delayed complement construction where the complement clause is not at the end, would be ungrammatical, as illustrated in (6a).

4 Analysis

In this section, the delayed complement construction and the more regular construction, with the complement clause adjoined to the so-phrase, will be given a uniform analysis.

4.1 Incremental analysis

The analysis is conducted within the framework of an HPSG-inspired incremental typed feature structure grammar for Norwegian (Haugereid 2009), implemented using the LKB system (Copestake 2002) as part of the Delph-In effort.[4] This approach assumes a distinction between a parse tree and a constituent tree (Haugereid & Morey 2012), where utterances are parsed incrementally in a bottom-up fashion from left to right, resulting in a completely left-branching tree structure. The grammar functions similarly to a shift-reduce parser, utilizing a STACK feature to monitor matrix constituents during the parsing of embedded constituents. Upon completing the parse, the constituent structure of the clause can be inferred from examining the STACK feature of each node in the parse tree. As will be shown, this incremental approach is well-suited for the analysis of delayed complement constructions.

Figure 3 provides the parse tree for example (2b), repeated below as (7). In this figure, the stacked constituent (V) is put on a list while the subordinate clause is parsed. At the top of the tree, the V constituent is popped from the list. A more detailed analysis of the same sentence is provided later, in Section 4.6.

(7) Så sen er han at jeg smiler.
so late is he that I smile
'He is so late that I smile.'

4.2 Licensed Complement

The central assumption of the analysis presented in this paper, drawn from Huddleston & Pullum (2002: 967), is that the complement clause consistently appears at the end of the clause and that it is a complement of clause structure, rather than the licensing adverbs (in Norwegian, *så*/*såpass* 'so' and *slik*/*sånn*/*sådan* 'such').

[4] https://github.com/delph-in/docs/wiki

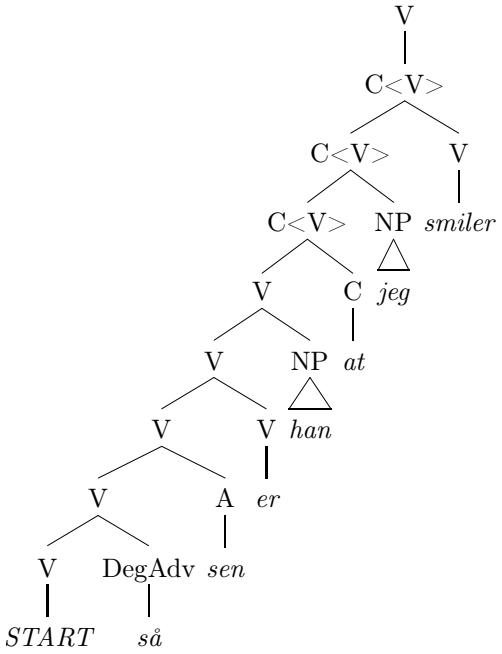


Figure 3: Incremental parse tree

This is achieved by allowing the feature licensing the complement clause ascend the tree from the point where the licensing adverb is realized until it triggers a rule, initiating the parsing of a complement clause. The analysis encompasses lexical entries for the licensing degree adverbs, a rule for the licensing adverbs, a feature LC (Licensed Complement), and a rule for the licensed complement.

4.3 Degree Adverbs

The lexical entry for the degree adverb *så* 'so' is given in (8). It modifies an adjective, adverb or preposition.

(8)
$$\begin{bmatrix} \textit{degadv-word} & \\ \text{STEM} & \left\langle \text{“så”} \right\rangle \\ \text{HEAD} & \begin{bmatrix} \textit{degadv} \\ \text{MOD} \left\langle \left[\text{HEAD } \textit{adj-adv-prep} \right] \right\rangle \end{bmatrix} \\ \text{KEYREL} & \left[\text{PRED } \textit{så\_deg} \right] \end{bmatrix}$$

The predicate of *så*, *så_deg*, is an underspecified type with two possible subtypes, *så_deg_rel* and *så_deg-cp_rel*, as illustrated in Figure 4. The regular degree adverb type *så_deg_rel* inherits from the type *comp–*, which means that it is not compatible with a complement argument, while the type *så_deg-cp_rel* inherits from the type

comp+, which means that it requires a CP complement. This underspecification is unique to the five CP-licensing adverbs.

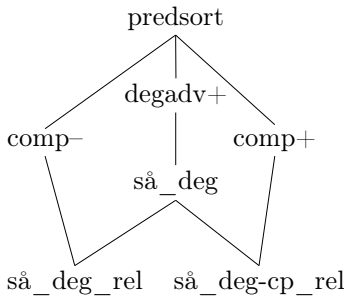


Figure 4: Type hierarchy of predicate types for the degree adverb *så 'so'*

4.4 Degree Adverbs Requiring CP Complements

The rule responsible for attaching degree adverbs that require a CP complement is presented in Figure 5. Given the left branching structures in this approach, coupled with the leftward attachment of degree adverbs, these adverbs are parsed before the modified word. To accommodate this, the element on the MOD list of the degree adverbs is unified with the feature PREMOD in the mother node. The subsequent rule that attaches the modified word unifies the PREMOD feature of its first daughter with the LOCAL value of its second daughter. Consequently, the degree adverb has the LOCAL features of the word it modifies on its MOD list. This is demonstrated later, in Figure 7.

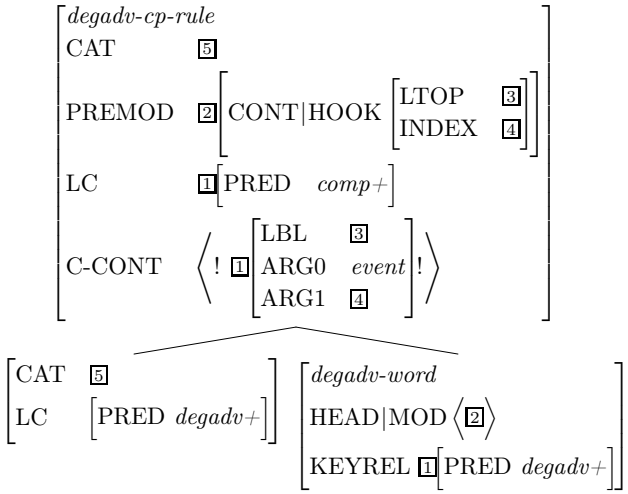


Figure 5: Rule for attaching degree adverb that requires a complement clause

The degree adverb rule unifies the KEYREL of the degree modifier with the feature LC (Licensed Complement) in the mother. This relation is specified to have the PRED value *comp+*. The corresponding value in the first daughter is underspecified, allowing for more than one CP-licensing adverb.

4.5 Delayed Complement Rule

The final part of the analysis involves a rule that initiates the parsing of a CP required by an adverb, given in Figure 6.

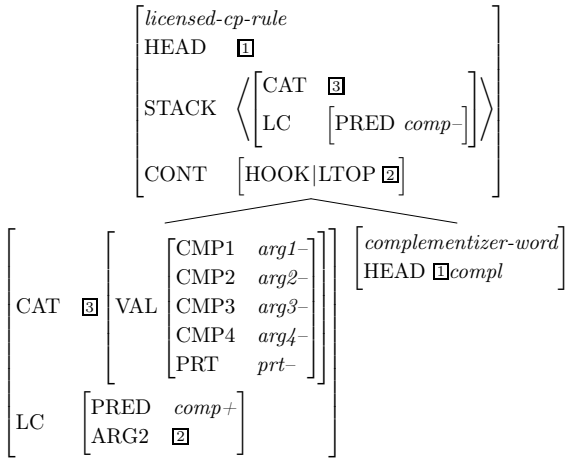


Figure 6: Rule for attaching complementizer initiating CP licensed by degree modifier

The first daughter of the rule is a clause where all the arguments and particles are realized,[5] and which licenses a complement clause (the LC|PRED value is *comp+*). The second daughter is a complementizer. In the mother node, the CAT features of the initial daughter are placed on a STACK.[6] Additionally, the ARG2 of the LC relation in the first daughter is unified with the LTOP of the complement clause (the mother).

4.6 Analysis of Sentence with Delayed Complement

The analysis of sentence (2b)/(7) with a delayed complement is illustrated in Figure 7. It demonstrates the incremental parsing of the sentence, detailing how the degree adverb *så* 'so' licenses the delayed complement clause through the feature LC. The figure also depicts how the relation of the degree adverb is linked to the modified

[5]The negative values of CMP1, CMP2, CMP3, CMP4, and PRT indicate that all the dependents of the main verb are realized.

[6]The STACK feature allows for parsing of embedded structures, see Haugereid & Morey (2012).

adjective and the complement clause, and that the relation is added to RELS via C-CONT|RELS.

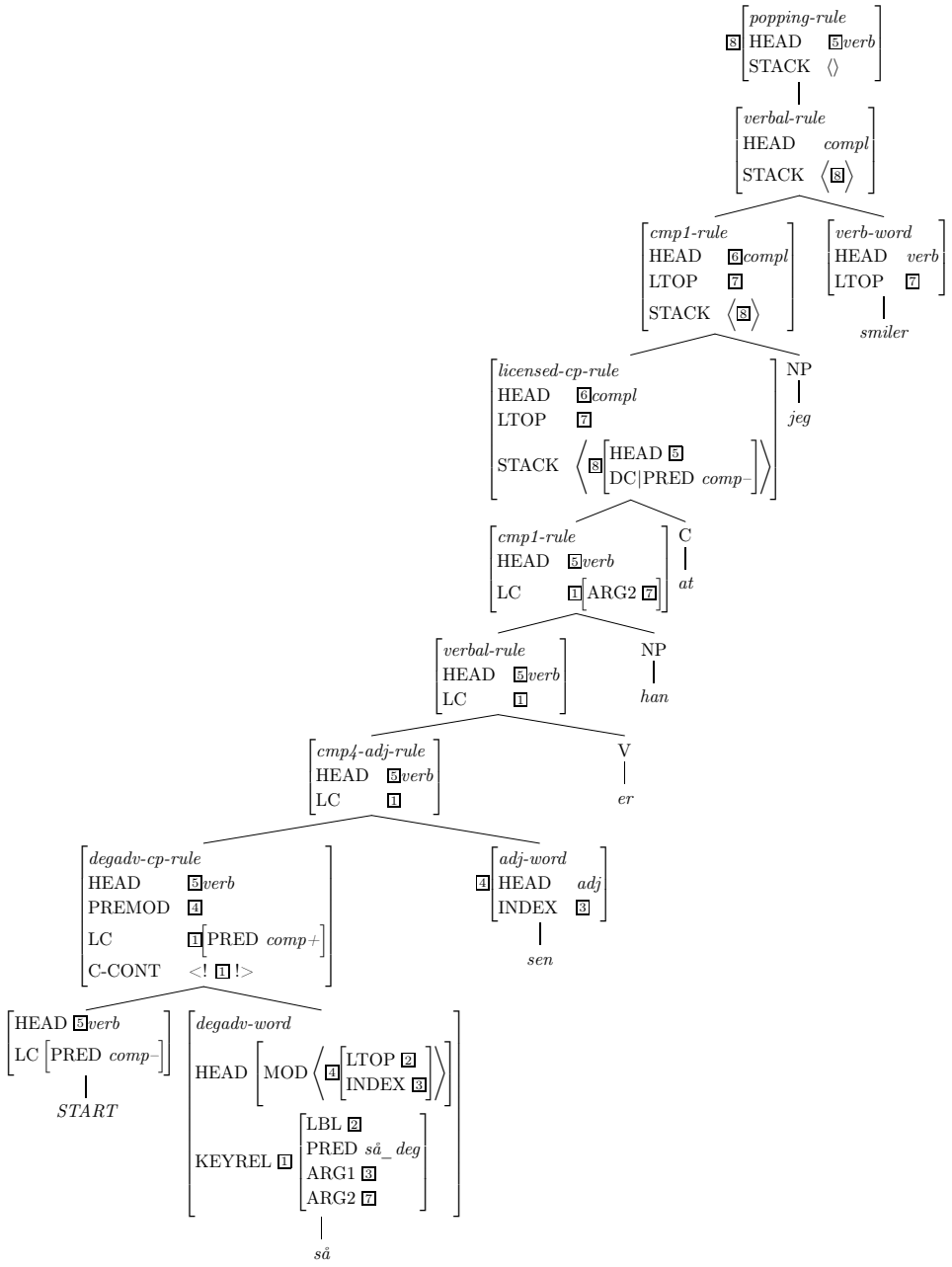


Figure 7: Incremental analysis of sentence with delayed complement

The MRS resulting from the analysis in Figure 7 is given in Figure 8. It illustrates how the relation of the degree adverb *_så_deg-cp_rel* has two arguments.

The first (5) is the index of of the modified adjective *_sen_a_rel*, and the second (6) is the index of the the delayed complement clause.

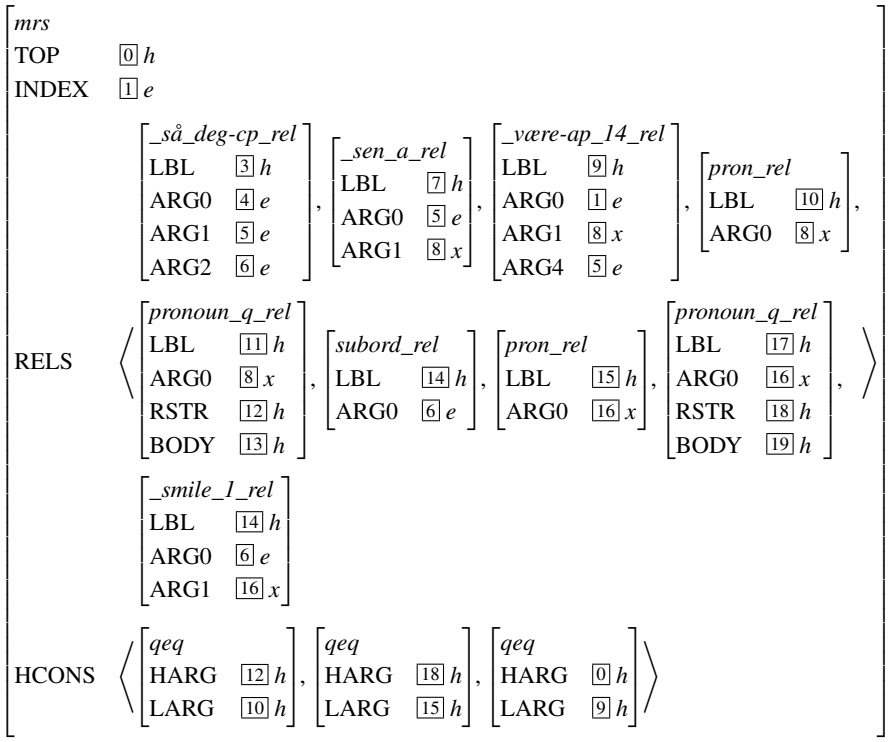


Figure 8: MRS of the sentence *Så sen er han at jeg smiler* 'He is so late that I laugh.'

4.7 Comparative Structures

When it comes to the comparative structures exemplified in (5), repeated below as (9), they can be analyzed in a similar fashion to the delayed complement constructions.

(9) a. More people enrolled for the course than we had expected.

b. Too many people enrolled for the course to fit in the room.

Just like *so*, the comparative governors *too* and *more* would have a relation with an underspecified PRED value, as illustrated in the simplified type hierarchy in Figure 9.

There would be separate rules for attaching *too* and *more* as comparative governors requiring a complement clause. Additionally, there would be separate rules for initiating an infinitival clause if the LC feature has a *vp+* requirement, and a than-clause if the LC feature has a *than-cp+* requirement.

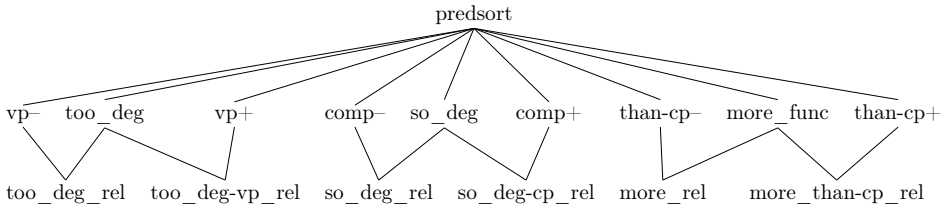


Figure 9: Simplified type hierarchy of predicate types for *too*, *so*, and *more*

The analysis of the comparative structures differs from the analysis of delayed complements in that the licensed clause does not need to be sentence-final. The rules that trigger the parsing of the licensed clause can also apply within the sentence, accommodating examples like (6b) and (6c).

5 A Challenge for Regular HPSG Grammars

The analysis presented in Section 4 can be adapted to regular HPSG grammars. One would then have to assume that the feature LC ascends to the sentence level and there combines the sentence with the complement clause.

However, the position of the licensing element in a so-phrase is not always the same. For example, in (10a) the licensing element *so* comes after the determiner (*a*), and in (10b) the licensing element *such* comes before the determiner.

(10) a. a so good sound

b. such a good sound

This means that the NP would get the licensing feature LC from the right daughter if the licensing element follows the determiner, as exemplified in Figure 10, while it would get the LC feature from the left daughter if the licensing element precedes the determiner (see Figure 11). This would require two different NP rules.

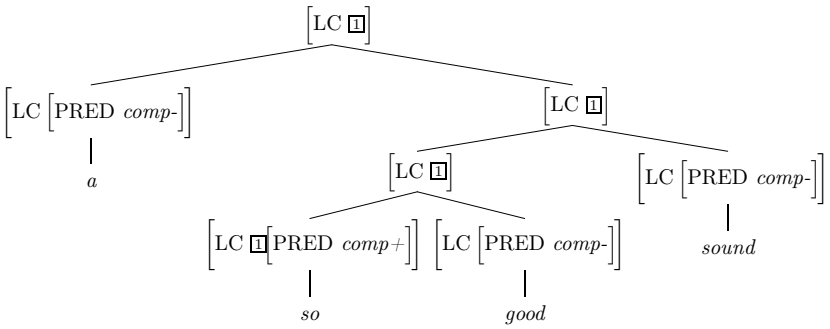


Figure 10: Getting the LC feature from the right daughter of an NP

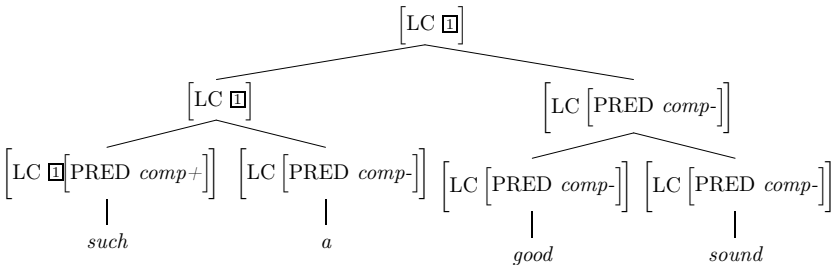


Figure 11: Getting the LC feature from the left daughter of an NP

With the incremental approach, however, the LC feature is just passed up from first daughter to mother once the degree adverb has been realized, and there is no need for separate NP rules.

6 Conclusion

An analysis of delayed complement constructions is provided, wherein the complement clause appearing at the end of the sentence is assumed to be a complement of the clause licensed by an adverb like *så* 'so', rather than by the adverb itself.

By adopting this assumption, a consistent analysis can be applied regardless of the proximity between the phrase with *så* 'so' and the complement clause. This approach allows for flexibility, accommodating scenarios where the phrase with the licensing adverb is either adjacent to or distant from the complement clause, while maintaining a uniform analysis throughout.

The left-branching structures assumed in this approach facilitate an analysis where the licensing adverb's requirement is registered and passed up from the daughter node to the mother node, regardless of whether the adverb appears at the beginning of the sentence or not. This method can be adapted to a standard HPSG grammar. However, this adaptation poses challenges, as it necessitates considering that the licensing condition (LC) feature could originate from either the first or the second daughter of an NP.

References

Copestake, Ann. 2002. *Implementing typed feature structure grammars* (CSLI Lecture Notes 110). Stanford, CA: CSLI Publications.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag. 2005. Minimal Recursion Semantics: An introduction. *Research on Language and Computation* 3(2–3). 281–332. DOI: 10.1007/s11168-006-6327-9.

Fjeld, Ruth Vatvedt, Anders Nøklestad & Kristin Hagen. 2020. Leksikografisk bokmålskorpus (LBK) – bakgrunn og bruk. *Oslo Studies in Language* 11(1). 47–59.

Flickinger, Daniel P. 2000. On building a more efficient grammar by exploiting types. *Natural Language Engineering* 6(1). 15–28.

Haugereid, Petter. 2009. *Phrasal subconstructions: a constructionalist grammar design, exemplified with norwegian and english*. Norwegian University of Science & Technology. (Doctoral dissertation). http://urn.kb.se/resolve?urn=urn:nbn:no:ntnu:diva-5755.

Haugereid, Petter & Mathieu Morey. 2012. A left-branching grammar design for incremental parsing. In Stefan Müller (ed.), *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar, Chungnam National University Daejeon*, 181–194. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2012.11.

Huddleston, Rodney & Geoffrey K. Pullum (eds.). 2002. *The Cambridge grammar of the English language*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/9781316423530.

Hybrid Agreement in Bosnian/Croatian/Serbian

János Litzinger

Physikalisch-Technische Bundesanstalt

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 118–131

Keywords: hybrid agreement, Bosnian/Croatian/Serbian, HPSG

Litzinger, János. 2024. Hybrid Agreement in Bosnian/Croatian/Serbian. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 118–131. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.8.


CC-BY 4.0

Abstract

This paper examines the hybrid agreement patterns in Bosnian/ Croatian/Serbian (BCS). Building on the previous work of the analysis demonstrates that by adopting a default unification mechanism and extending the existing type hierarchy, a feasible analysis of BCS hybrid agreement can be accomplished. The paper also explores a functor analysis, which delivers successful results with minor changes to the type hierarchy and lexical entries. The findings also refute the claim that the phenomenon of hybrid agreement in BCS gives evidence for the DP-hypothesis, thereby rendering the argument unable to favour either the NP or the DP hypothesis.

1 Introduction

The debate whether the noun or the determiner is the head of a nominal phrase has been ongoing since the 1980s with the emergence of the so-called DP-hypothesis. Prior to its emergence, the standard analysis held that the noun constitutes the head of a nominal phrase. The DP-hypothesis was first proposed by Szabolcsi (1983), Fukui (1986) and Abney (1987) in a reaction to Chomsky's reworking of the clause in order to maintain the parallelism between the structure of the clause and the nominal domain. As Minimalism in syntax (Chomsky 1993) became increasingly popular, the DP-hypothesis gained textbook status e.g. Adger (2003).

Nonetheless the debate is far from being settled in favour of the DP-hypothesis. Salzmann (2018) and Salzmann (2020) revisit the NP vs. DP debate and discuss the arguments seemingly supporting each hypothesis. To refresh the debate Salzmann (2020) works out a sharp definition of headedness and introduces a puzzle regarding hybrid agreement in Bosnian/Croation /Serbian (henceforward BCS).[1]

In BCS, the noun class II is of grammatical gender feminine but refers to male entities. While only agreeing in semantic gender for singular number, grammatical gender is triggered when the noun is being used in plural number. To complicate things even more some speakers allow for a mixed agreement inside the same sentence or noun phrase, hence feminine and masculine gender.

[†]This paper is the result of a master's thesis written at Humboldt-University of Berlin in 2024. I thank my supervisors Stefan Müller and Antonio Machicao y Priemer, and the attendees of the HPSG 2024 conference (Olomouc, Czechia) for their comments. I also thank the anonymous reviewers for their valuable feedback.

[1]The language naming is based on alphabetical order to avoid any value-ranking (Alexander 2006: 426).

(1) (Puškar 2018: 278)

a. star-**i**/*star-**a** vladik-**a** me je juče
old-M.SG/old-F.SG bishop-SG me is yesterday
posetio-∅/*posetil-**a**
visit.PTCP-M.SG/F.SG
'the old bishop visited me yesterday'

b. star-**e** vladik-**e** su me juče posetil-**e**/posetil-**i**
old-F.PL bishop-PL are me yesterday visit.PTCP-F.PL/M.PL
'the old bishops visited me yesterday'

c. star-**i** vladik-**e** su me juče posetil-**i**/*posetil-**e**
old-M.PL bishop-PL are me yesterday visit.PTCP-M.PL/F.PL
'the old bishops visited me yesterday'

In (1a) the agreeing adjective *stari* 'old' and participle verb *posetio* 'visit' show agreement with the semantic gender of the noun *vladik* 'bishop', hence masculine gender. If the noun is used in plural number as in (1b), the agreeing elements bear feminine gender. Some speakers of BCS allow for semantic agreement for plural nouns as shown by the sentence in (1c) shows, some speakers even allow for a mixed agreement pattern as in (1b). Note, that once semantic agreement is used, succeeding agreeing elements can not bear grammatical gender. This behaviour is made explicit by the sentence in (1c), where the adjective bears masculine gender and the succeeding participle verb is only not permitted to bear feminine gender. This behaviour in hybrid agreement is in line with Agreement Hierarchy by Corbett (1979), which describes a decreasing probability of grammatical agreement along the agreement hierarchy. It ranges from the attributive domain via the predicative domain and the relative pronoun and end in the personal pronoun. This means that the chance of semantic agreement rises (with no intervening decrease) from the attributive domain to the personal pronoun (Corbett 2006: 207).

2 Discussion

The phenomenon of hybrid agreement was already addressed by Wechsler & Zlatić (2003) who assume two gender attributes, namely CONCORD|GENDER and INDEX|GENDER. The former being the grammatical and the latter the semantic property of the noun. Furthermore, they work out that while adjectives and determiners agree with the CONCORD gender the participle verb in BCS agrees with the INDEX gender. To derive the behaviour of class II nouns like *vladika* 'bishop', they employ a default unification system proposed by Lascarides & Copestake (1999). It unifies default values (represented on the right side of the forward slash), if there is no conflicting hard value, or other soft value specified by a type lower in the hierarchy (Wechsler &

Zlatić 2003: 42). That means that subtypes have priority when unifying over supertypes (Wechsler & Zlatić 2003: 66). Wechsler & Zlatić use the mechanism of default unification to constrain words of the type *noun-word*$_{ci}$ (concord-index) which is a subtype of *noun-word*$_{si}$ (semantics-index). In the type hierarchy they are structurally above the types for the declension classes.

(2) Default unification constraints (Wechsler & Zlatić 2003: 66)

a. *noun-word*$_{si}$:

$$\begin{bmatrix} \text{INDEX} & \left[\text{GENDER} \quad \textit{gender}/\boxed{1}\right] \\ \text{RESTR} & /\left[\text{SEX} \quad \boxed{1}\textit{sex}\right] \end{bmatrix}$$

b. *noun-word*$_{ci}$:

$$\begin{bmatrix} \text{CONCORD} & \begin{bmatrix} \text{GENDER} & /\boxed{3} \\ \text{NUMBER} & /\boxed{4} \end{bmatrix} \\ \text{INDEX} & \begin{bmatrix} \text{GENDER} & /\boxed{3} \\ \text{NUMBER} & /\boxed{4} \end{bmatrix} \end{bmatrix}$$

Wechsler & Zlatić posit that the type for the declension class II is further specified such that it employs a type *noun-II*$_{\emptyset}$ and *noun-II*$_{f}$, whereas the former is constrained for singular number and the latter for plural number and feminine concord gender (Wechsler & Zlatić 2003: 43). Bringing together the default unification and the type hierarchy has the effect that while singular nouns of type *noun-II*$_{\emptyset}$ are not constrained for concord gender the default unification of (2b) and (2a) apply. Assuming a male bishop the lexical sign for the noun *vladika* will show masculine gender values for the attributes SEX, INDEX and CONCORD as illustrated by (3). Thus, the sentence in (1a) can be derived.

(3)

$$\begin{bmatrix} \text{PHONOLOGY} & \begin{bmatrix} \text{STEM} & \textit{vladik} \\ \text{DECL} & \textit{II} \end{bmatrix} \\ \text{SYNSEM} & \begin{bmatrix} \text{CAT} \mid \text{HEAD} \mid \text{CONCORD} \mid \text{GENDER} \ \boxed{1} \\ \text{CONTENT} & \begin{bmatrix} \text{INDEX} & \textit{i}\left[\text{GENDER} \quad \boxed{1}\right] \\ \text{RESTRICTIONS} & \begin{bmatrix} \text{PRED} & \left\{\textit{bishop(i)}\right\} \\ \text{SEX} & \boxed{1}\textit{masc} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

On the other hand, if the noun is used in plural number with the type *noun-II*$_{f}$, CONCORD|GENDER is set to *feminine*. Through the default unification constraint on type *noun-word*$_{ci}$ INDEX|GENDER is shared with CONCORD|GENDER. The default unification on *noun-word*$_{si}$ can not apply since the subtype has the priority here. The resulting lexical entry for *vladike* 'bishops' now bears feminine CONCORD and INDEX gender being able to produce the sentence in (1b). Wechsler & Zlatić note that some speakers of BCS allow masculine gender agreement also for nouns with plural number, thus they assume that for these speakers the type *noun-II*$_{\emptyset}$ is not constrained for singular number (Wechsler & Zlatić 2003: 71). But as the example in

(1b) shows, also mixed agreement patterns are possible. This pattern can not be derived with the system sketched by Wechsler & Zlatić, since the default unification constraint on *noun-word*$_{ci}$ ensures identical values for CONCORD and INDEX gender.

Salzmann (2020) introduces an example from BCS with even more complexity. Here, the switch from grammatical gender to semantic gender is made between the attributive adjective and the demonstrative *oni* 'those'.

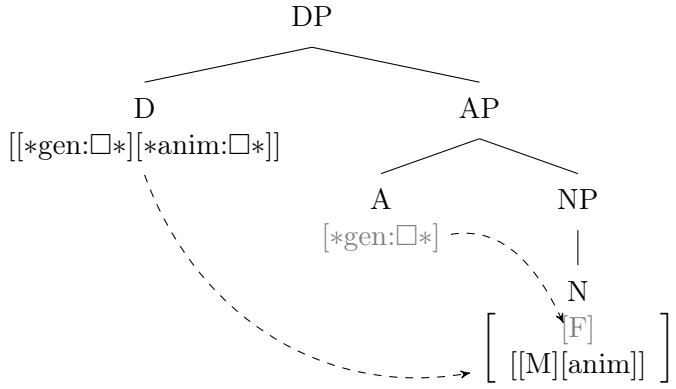
(4) BCS (Salzmann 2020: 34)
Oni star-e vladike su se posvadjal-i/*posvadjal-e na
Those-**M**.PL old-**F**.PL bishops are REFL argued-**M**.PL/argued-F.PL on
ulici.
street
'Those old bishops argued on the street'

For Salzmann this is evidence, that the head of the nominal phrase is D. He builds his argumentation upon the work of Puškar (2017) and Puškar (2018), who employs relativized probing to derive the agreement patterns of BCS. In short, the difference between semantic and grammatical gender lies in the complexity of their probes. The feature for semantic gender has an additional node for animacy. Salzmann defines that complex probes can only be valued by complex features. Furthermore, they can pass simple features, although simple probes cannot pass complex features. Moreover, simple probes cannot be valued by complex features. This has the effect that once a head's probe is valued by a complex feature, simple probes cannot look past it (Salzmann 2020: 35).

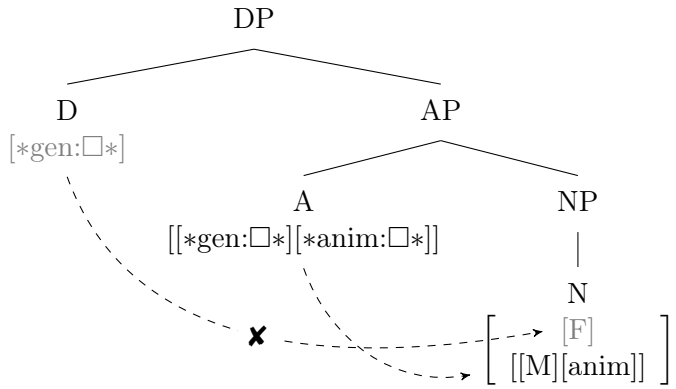
In (5) the adjective is merged first and therefore also probes first. The adjective in (5a) probes for a simple gender feature and is valued by the simple feature on the noun. After that the determiner with a complex gender probe is merged. It probes passed the adjective and is valued by the complex feature on the noun.

(5) (Salzmann 2020: 36)

a.



b.



On the other hand, if the adjective probes for a complex feature, like in (5b), and the determiner then tries to probe for a simple feature, derivation fails. The simple probe from the determiner cannot be valued by the features on the adjective nor can it probe past the adjective to be valued by the simple feature on the noun. In cases where the probes of the adjective and determiner correspond in terms of complexity, the derivation will always be successful. If another noun phrase, such as V, were to enter the derivation, the same principles relating to the complexity of probes would apply. Consequently, this system can derive the concordance patterns (1) and (4).

Salzmann states that under the NP-hypothesis this analysis would not function, as the features on N would be projected and thus accessible for probes from D and V, even in scenarios where grammatical agreement is disregarded (Salzmann 2020: 38). As will be shown in the next section an NP-analysis is possible, building on the work by Wechsler & Zlatić (2003). Furthermore, an alternative approach following Van Eynde (2020) will be assessed showing that default unification is not strictly necessary to derive the mixed agreement patterns of BCS.

3 Proposal

To account for the mixed agreement patterns of (1b,1c) and (4), the type hierarchy for class II nouns is augmented as shown in Figure 3. The type *noun-II*$_\emptyset$ is still constrained for singular number and the default unification constraints apply as described above and illustrated by (3). The second subtype is *noun-II*$_{pl}$, which is only constrained for plural number. Its subtypes are *noun-II*$_f$ and *noun-II*$_m$, whereas the former works the same as with Wechsler & Zlatić (2003) and the latter is underspecified for CONCORD|GENDER with the value *sex*. Furthermore, it is constrained for INDEX|GENDER *masculine*.

The type *noun-II*$_f$ will result in a lexical sign with INDEX|GENDER *feminine* producing sentences with an all feminine pattern. The type *noun-II*$_m$ has both CONCORD and INDEX gender specified and thus the default unification on type *noun-word*$_{ci}$ can not apply. Recall that in the YADU system the subtype

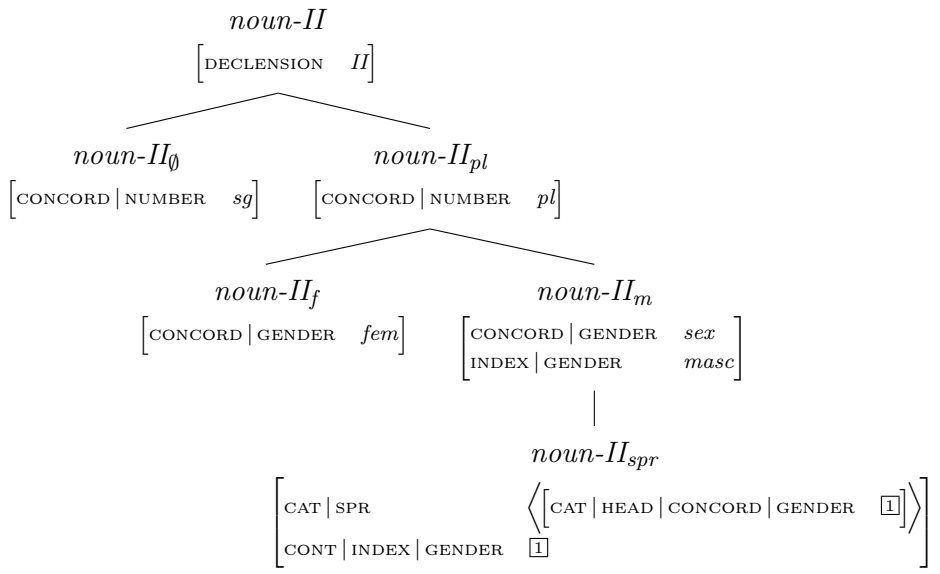


Figure 1: Revised version of the class II type hierarchy

takes priority over the supertype. Since CONCORD|GENDER is underspecified, the noun can combine with either feminine or masculine adjectives, ensuring the participle verb to bear masculine gender, since the type is constrained for INDEX|GENDER *masculine*.

With the subtype *noun-II*$_{spr}$ of type *noun-II*$_m$ the determiner comes into play. It is defined such that the CONCORD|GENDER of the specifier is bound to the INDEX|GENDER of the noun. Additionally, a default unification constraint is introduced on the type *word-noun*. By default it binds the CONCORD|GENDER value of the specifier to the CONCORD|GENDER value of the noun. This has the effect that nouns of all types except *noun-II*$_{spr}$ trigger concord agreement with the specifier.

(6) *noun-word*:

$$\begin{bmatrix} \text{SPR} & \left\langle \left[\text{CAT} \mid \text{HEAD} \mid \text{CONCORD} \mid \text{GENDER} \quad /\,\boxed{1} \right] \right\rangle \\ \text{CONCORD} \mid \text{GENDER} & /\,\boxed{1} \end{bmatrix}$$

To summarize class II nouns of type *noun-II*$_\emptyset$ produce an all masculine gender pattern in singular number as in (1a). This is achieved through the default unification constraint on *noun-word*$_{ci}$ (2b) which binds the CONCORD|GENDER to INDEX|GENDER and INDEX|GENDER through the constraint on *noun-words*$_{si}$ (2a) to the value of SEX. Since the gender of the specifier

is not defined the constraint on *noun-word* applies that binds the specifiers GENDER value by default to CONCORD|GENDER resulting in the lexical sign in (7). Recall that *noun-II*$_\emptyset$ is restricted to singular number ensuring that singular nouns of class II only trigger semantic agreement as shown in (4).

(7)
$$\left[\begin{array}{l} \textit{noun-II}_\emptyset \\ \\ \text{SYNSEM} \end{array}\right.\left[\begin{array}{l} \text{CAT}\left[\begin{array}{l} \text{CONCORD} \mid \text{GENDER } \boxed{1} \\ \text{SPR} \left\langle \left[\text{CAT} \mid \text{HEAD} \mid \text{CONCORD} \mid \text{GENDER } \boxed{1}\right] \right\rangle \end{array}\right] \\ \text{CONT}\left[\begin{array}{ll} \text{INDEX} \mid \text{GENDER} & \boxed{1} \\ \text{REST} \mid \text{GENDER} & \boxed{1}\ \textit{sex} \end{array}\right] \end{array}\right]\left.\right]$$

All feminine gender patterns as in (1b) can be derived with *noun-II*$_f$ nouns as represented by the lexical sign in (8). The type *noun-II*$_f$ is constraint for CONCORD|GENDER *feminine* and since the INDEX|GENDER is bound to CONCORD|GENDER through constraint on (2b) the participle verb bears feminine gender. The specifier bears feminine gender since again the constraint on *noun-word* (6) applies binding its GENDER value to the one of CONCORD.

(8)
$$\left[\begin{array}{l} \textit{noun-II}_f \\ \\ \text{SYNSEM} \end{array}\right.\left[\begin{array}{l} \text{CAT}\left[\begin{array}{l} \text{CONCORD} \mid \text{GENDER } \boxed{1}\ \textit{fem} \\ \text{SPR} \left\langle \left[\text{CAT} \mid \text{HEAD} \mid \text{CONCORD} \mid \text{GENDER } \boxed{1}\right] \right\rangle \end{array}\right] \\ \text{CONT} \mid \text{INDEX} \mid \text{GENDER } \boxed{1} \end{array}\right]\left.\right]$$

Mixed gender patterns as in (1b) and all masculine patterns in plural number such as in (1c) can be analysed with the type *noun-II*$_m$ and *noun-II*$_{spr}$. Both types are constraint for INDEX|GENDER *masculine* and underspecified for CONCORD|GENDER *sex* overwritting the default unification constraint on *noun-word*$_{ci}$ 2b, since in the YADU system subtypes take priority over their supertype. The constraint on *noun-word* binds the specifiers gender by default to CONCORD|GENDER. For sentences with a determiner and adjective such as (4) this has the result that both the determiner and adjective share the same gender as reflected by the lexical sign in (9). Furthermore, the underspecification of CONCORD|GENDER allows for feminine and masculine gender on both the determiner and adjective. Since the value of INDEX|GENDER is *masculine* ungrammatical sentences as indicated by the stared forms in (1c) can not be produces.

(9)
$$\begin{bmatrix} \textit{noun-II}_m & \\ \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{CONCORD} \mid \text{GENDER } \boxed{1}\ \textit{sex} \\ \text{SPR} \left\langle \left[\text{CAT} \mid \text{HEAD} \mid \text{CONCORD} \mid \text{GENDER } \boxed{1} \right] \right\rangle \end{bmatrix} \\ \text{CONT} & \left[\text{INDEX} \mid \text{GENDER } \textit{masc} \right] \end{bmatrix} \end{bmatrix}$$

To derive sentences with a mismatch between the gender of the adjective and determiner, as shown in (4), type *noun-II*$_{spr}$ come into play. The lexical sign is shown in (10). The type *noun-II*$_{spr}$ is constraint such that the gender of the specifier is bound by INDEX|GENDER with the effect that the default unification on *noun-word* (6) does not apply. Thus, the determiner would bear masculine gender while the adjective is free of bearing masculine or feminine gender due to the underspecification of CONCORD|GENDER.

(10)
$$\begin{bmatrix} \textit{noun-II}_{spr} & \\ \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{CONCORD} \mid \text{GENDER } \textit{sex} \\ \text{SPR} \left\langle \left[\text{CAT} \mid \text{HEAD} \mid \text{CONCORD} \mid \text{GENDER } \boxed{1} \right] \right\rangle \end{bmatrix} \\ \text{CONT} & \left[\text{INDEX} \mid \text{GENDER } \boxed{1}\ \textit{masc} \right] \end{bmatrix} \end{bmatrix}$$

This point is a possible weakness of this analysis since both *noun-II*$_m$ and *noun-II*$_{spr}$ can produce all masculine gender patterns in the plural. If type *noun-II*$_m$ is modified by a masculine adjective, the nouns CONCORD|GENDER resolves to *masculine* and through the default unification constraint on *noun-word* the specifier gender is also masculine. The type *noun-II*$_{spr}$ on the other hand defines specifier's *gender* as masculine. Hence, both types have a identical feature structure only differing in their structure sharing. This posits a challenge, since this type hierarchy then produces two solutions for those sentences. To work around this problem one could assume that the final linguistic representation need to have types that are maximal specific. That way semantic agreement patterns would be only derived by *noun-II*$_{spr}$.

3.1 Functor analysis

An alternative way to circumvent the problem of ambiguous solutions is to follow the functor analysis of the determiner. In contrast to the approach above based on Wechsler & Zlatić (2003), where determiners satisfy the object in the specifier list of the noun, determiners as functors attach to nouns like adjectives. They have a type *noun-word* as their value of the attribute SELECT and if satisfied resulting in a head-functor-phrase as shown in (11).

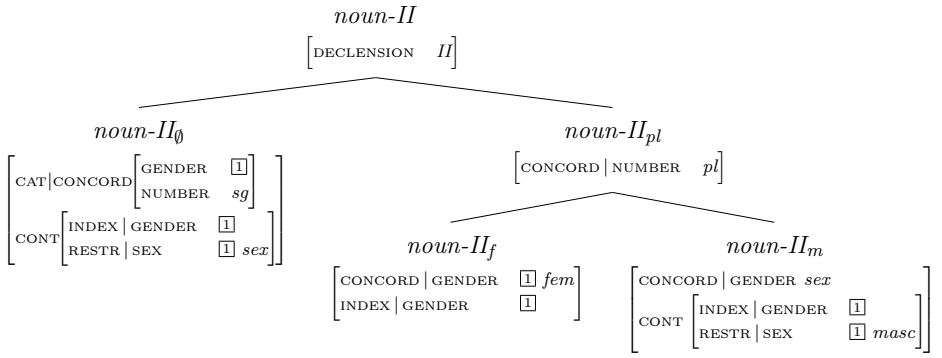


Figure 2: Revised version of the class II type hierarchy

(11) (Van Eynde 2020: 10)
head-functor-phrase ⇒ $\begin{bmatrix} \text{DAUGHTERS} & \left\langle \left[\text{SYNSEM}\,|\,\text{CAT}\,|\,\text{HEAD}\,|\,\text{SELECT}\ \boxed{1}\right], \text{X} \right\rangle \\ \text{HEAD-DTR}|\text{SYNSEM} & \boxed{1}\ \textit{synsem} \end{bmatrix}$

To derive the agreement patterns of BCS the type hierarchy of the type *noun-word* is revised such that it is striped of the default unification constraints as shown in Figure 2. Moreover, the subtypes of type *noun-II* are reduced to the three subtypes *noun-II*$_\emptyset$, *noun-II*$_f$ and *noun-II*$_m$. The type *noun-II*$_\emptyset$ is constrained for singular number as in its previous version. Additionally, it is defined such that the value of CONCORD and INDEX GENDER is shared with the value of the SEX attribute. The resulting lexical sign is the same as the one employing default unification in (2a). This ensures that only semantic agreement can apply for singular nouns.

Grammatical agreement, hence an all feminine gender pattern is achieved with the type *noun-II*$_f$ mimicking the default unification constraint in (2b), binding the value of INDEX|GENDER to the one of CONCORD|GENDER. The remaining type *noun-II*$_m$ is underspecified for CONCORD|GENDER *sex*. By sharing the value of the attribute SEX with INDEX|GENDER and additionally constraining it for *masculine* SEX, it is ensured that the participle verb agrees in masculine gender. With these mechanism semantic agreement patterns like the one in (1c). In order to derive patterns with a mismatch in gender between the adjective and determiner, as illustrated by the sentence in (4), further assumptions about determiners need to be made. (12) gives a simplified lexical sign for a masculine determiner. The object in the SELECT list is underspecified for CONCORD|GENDER, whereas INDEX|GENDER is bound by the CONCORD|GENDER of the determiner, which has the value *masculine*.

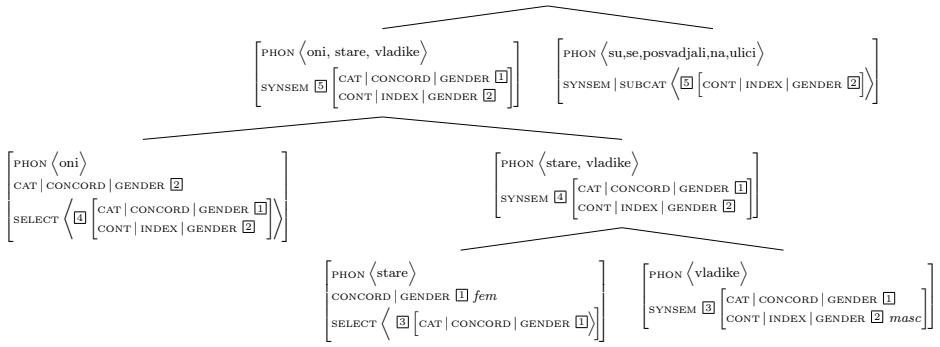


Figure 3: Derivation tree for mixed agreement pattern under the functor analysis for the sentence *Oni stare vladike su se posvadjali na ulici* 'Those old bishops argued on the street' (4).

(12)
$$\begin{bmatrix} \text{det} & \\ \text{CONCORD} \mid \text{GENDER} & \boxed{1}\ \textit{masc} \\ \text{SELECT} & \left\langle \begin{bmatrix} \text{CAT} \mid \text{CONCORD} \mid \text{GENDER} & \textit{sex} \\ \text{CONT} \mid \text{INDEX} \mid \text{GENDER} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

Underspecifying the concord gender of the selected item the determiner is able to attach to both feminine and masculine noun phrases. By sharing the same value between CONCORD|GENDER and the selected item's INDEX|GENDER , which resolves to masculine, a sentence with a verb bearing feminine gender is excluded. Making this assumption exclusively for masculine determiners in BCS is only possible because mixed agreement patterns in BCS can only be observed for male referring entities (Puškar 2018: 282).

The tree in Figure 3 shows a derivation for the sentence in (4). First, the the feminine adjective *stare* selects the noun *vladike* of noun type *noun-II*$_m$ which is underspecified for CONCORD|GENDER *sex*. Through structure sharing with the adjective's CONCORD|GENDER the noun's CONCORD|GENDER resolves *fem*, the subtype of *sex* in the gender type hierarchy. The resulting phrase saturates the object in the determiner's SELECT list, since it bears the INDEX|GENDER *masculine*. Lastly, the verb cluster has a object of type *noun* in its SUBCAT list. As it bears masculine gender the SUBCAT list can only be satisfied by an object, whose INDEX|GENDER is *masculine*.

The main advantage of the functor analysis is, that it does not rely on the usage of defaults, making this analysis easier to implement using a grammar development framework like TRALE (Penn et al. 2003). On the contrary, it mimics a DP-analysis by reverting the selector-selectee relationship of nouns and determiner/adjectives. On the other hand noun still projects its features and thus the NP analysis could still hold for the functor analysis.

The strength of the analysis that is based on Wechsler & Zlatić 2003 in comparison to the functor analysis is in fact the use of the default unification, since it could represent the actual usage of hybrid agreement by speakers of BCS. The other noun classes in BCS don't allow for a mismatch in gender and thus bear the same gender of INDEX and CONCORD, which is mapped by the default unification of those features. Furthermore, the default unification that binds the INDEX|GENDER to the SEX attribute is also applied for nouns of the other noun classes.

Both analysis have in common that they are not that easily transferable to other cases of mixed agreement such as found in Hebrew, Russian, Finnish and Chichewa. Those languages allow for a mismatch between to adjectives in the nominal domain (Landau 2016: 1004–1008). In BCS both adjectives need to bear the same gender Puškar (2017: 102).

Finally, it should be noted that the evidence for the hybrid agreement pattern heavily relies on the data provided by Puškar (2018) and Salzmann (2020). This raises the question on how the phenomenon of hybrid agreement is distributed throughout the speakers of BCS. A corpus study would certainly help to clarify the sparse data situation but is beyond the scope of this paper.

4 Conclusion

The analysis of hybrid agreement in BCS initially followed the approach of Wechsler & Zlatić (2003). They distinguish between grammatical and semantic gender, which are present in the feature structure under CONCORD and INDEX respectively. The analysis demonstrated that by adopting the default unification mechanism of Wechsler & Zlatić and extending their type hierarchy, a feasible analysis of BCS hybrid agreement could be accomplished with the noun serving as the head of the nominal phrase. The analysis was embedded within the framework of HPSG whereas the agreement mechanism of HPSG remains untouched. Exploring a functor analysis based on Van Eynde (2020) delivers also successful results. It was shown that minor changes to the type hierarchy and lexical entries could derive the agreement patterns of BCS. This approach is more efficient in implementation complexity and addresses the weaknesses of the other approach, such as ambiguous solutions with defaults. Most importantly the analyses refute Salzmann's claim that the phenomenon of hybrid agreement in BCS gives evidence for the DP-hypothesis. Therefore, Salzmann's argument cannot be used to favour either the NP or the DP hypothesis, and loses its epistemic value.

References

Abney, Steven Paul. 1987. *The English noun phrase in its sentential aspects.* Cambridge, MA: MIT. (Doctoral dissertation). `http://www.ai.mit.edu/projects/dm/theses/abney87.pdf` (23 January, 2024).

Adger, David. 2003. *Core syntax: A Minimalist approach* (Oxford Core Linguistics 1). Oxford: Oxford University Press.

Alexander, Ronelle. 2006. *Bosnian, croatian, serbian, a grammar with sociolinguistic commentary*. Madison, WI: Univiversity of Wisconsin Press.

Chomsky, Noam. 1993. A Minimalist Program for linguistic theory. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger*. MIT Press.

Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15(2). 203–224. https://www.jstor.org/stable/4175494 (23 January, 2024).

Corbett, Greville G. 2006. *Agreement* (Cambridge Textbooks in Linguistics). Cambridge, UK: Cambridge University Press.

Fukui, Naoki. 1986. *A theory of category projection and its application*. Cambridge, MA: MIT. (Doctoral dissertation). http://www.ai.mit.edu/projects/dm/theses/fukui86.pdf (23 January, 2024).

Landau, Idan. 2016. DP-internal semantic agreement: A configurational analysis. *Natural Language & Linguistic Theory* 34(3). 975–1020. DOI: 10.1007/s11049-015-9319-3.

Lascarides, Alex & Ann Copestake. 1999. Default representation in constraint-based frameworks. *Computational Linguistics* 25(1). 55–105. https://aclanthology.org/J99-1002 (23 January, 2024).

Penn, Gerald, Detmar Meurers, Kordula De Kuthy, Mohammad Haji-Abdolhosseini, Vanessa Metcalf, Stefan Müller & Holger Wunsch. 2003. Trale milca environment v. 2.5.0 user's manual (draft). https://jakobson.korpus.cz/~rosen/public/trale-manual.pdf (23 January, 2024).

Puškar, Zorica. 2017. *Hybrid agreement: Modelling variation, hierarchy effects and ϕ-feature mismatches*. Leipzig: University of Leipzig. (Doctoral dissertation). https://ling.auf.net/lingbuzz/003795 (23 January, 2024).

Puškar, Zorica. 2018. Interactions of gender and number agreement: Evidence from Bosnian/Croatian/Serbian. *Syntax* 21(3). 275–318. DOI: 10.1111/synt.12154.

Salzmann, Martin. 2018. Revisiting the NP vs. DP debate. https://lingbuzz.net/lingbuzz/003823 (23 January, 2024).

Salzmann, Martin. 2020. The NP vs. DP debate. Why previous arguments are inconclusive and what a good argument could look like. Evidence from agreement with hybrid nouns. *Glossa: a journal of general linguistics* 5(1). 83. 1–46. DOI: 10.5334/gjgl.1123.

Szabolcsi, Anna. 1983. The possessor that ran away from home. *The Linguistic Review* 3(1). 89–102. DOI: 10.1515/tlir.1983.3.1.89.

Van Eynde, Frank. 2020. Agreement, disagreement and the NP vs. DP debate. *Glossa: A Journal of General Linguistics* 5(1). 1–23. DOI: 10.5334/gjgl.1119.

Wechsler, Stephen & Larisa Zlatić. 2003. *The many faces of agreement* (Stanford Monographs in Linguistics). Stanford, CA: CSLI Publications.

A new account of the case alternation in Korean complex predicates

Jieun Oh
Humboldt-Universität zu Berlin

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 132–151

Keywords: complex predicates, auxiliary verb -siph, case alternation, information structure markers, -i/ka, Korean, HPSG

Oh, Jieun. 2024. A new account of the case alternation in Korean complex predicates. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 132–151. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.9.
CC-BY 4.0

Abstract

This paper investigates the phenomenon referred to as 'case alternation' in the complex predicate with the auxiliary verb *siph-* 'want' in Korean. It provides an account for the analysis within the framework of Head-Driven Phrase Structure Grammar (HPSG). It begins by reviewing previous analyses that case alternation in the construction of the complex predicate *-ko siph-* 'want to' can be accounted for by a dual inheritance property specified in the lexicon. This study, by contrast, proposes a new account of case alternation in the complex predicate *-ko siph-* 'want to'. It introduces a new classification of the particle *-i/ka*, arguing that it functions not only as a subject case marker but also as an information structure marker. Furthermore, it argues that the grammatical case of the second argument marked with the particle *-i/ka* has not changed. Instead, the particle *-i/ka* enables the argument to be focused. Through a detailed analysis of the complex predicate *-ko siph-* 'want to', this research proposes a lexical entry for the particle *-i/ka* as an information structure marker, elucidating its role in the complex predicate construction *-ko siph-* 'want to'. The findings have implications for our understanding of case marking and the argument structure of the Korean complex predicate *-ko siph-* 'want to'.

1 Introduction

The interesting phenomenon of case alternation can be observed in Korean complex predicates. Typically, the arguments of complex predicates receive the case marking assigned by the embedded verb. For example, the verb *mek-* 'eat' selects two NPs that are realized as a nominative and accusative NP, as illustrated in (1a) and the verb *toy-* 'become' selects two NPs that are both marked with the particle *-i/ka*, as shown in (2a). Similarly, when an auxiliary verb (e.g. *po-* 'try') is combined with these verbs, the case marking remains consistent, as shown in (1b) and (2b). This pattern of case assignment applies to complex predicate constructions with auxiliary verbs in Korean. The examples provided show the typical case marking system in complex predicate constructions with auxiliary verbs.

(1) a. Hyenwu-ka sakwa-lul mek-ess-ta.
Hyenwu-NOM apple-ACC eat-PST-DECL
'Hyenwu ate an apple.'

b. Hyenwu-ka sakwa-lul mek-e po-ass-ta.
Hyenwu-NOM apple-ACC eat-CONN try-PST-DECL
'Hyenwu tried to eat an apple.'

(2) a. Hyenwu-ka sensayng-nim-i toy-ess-ta.
Hyenwu-NOM teacher-HON-NOM become-PST-DECL
'Hyenwu became a teacher.'

b. Hyenwu-ka sensayng-nim-i toy-e po-ass-ta.
Hyenwu-NOM teacher-HON-NOM become-CONN try-PST-DECL
'Hyenwu tried to become a teacher.'

However, the auxiliary verb *siph-* 'want' (cf. Sohn 1999; Kim 2016; Lee 2016b; Song 2020, a.o.) exhibits idiosyncratic properties in terms of case assignment, deviating from the typical patterns observed with other auxiliary verbs. In the complex predicate construction with the auxiliary verb *siph-* 'want', the second argument can be realized as either a nominative or accusative NP, as illustrated in (3a). This demonstrates that the second argument can be marked not only with the accusative marker *-ul/lul* but also with the nominative marker *-i/ka*. On the other hand, when the auxiliary verb *siph-* 'want' combines with verbs like *toy-* 'become', the second argument must be realized as an NP marked with the marker *-i/ka*, as shown in (3b).

(3) a. Hyenwu-ka {sakwa-lul / sakwa-ka} mek-ko siph-ta.
Hyenwu-NOM apple-ACC apple-NOM eat-CONN want-DECL
'Hyenwu wants to eat an apple.'

b. Hyenwu-ka {*sensayng-nim-ul / sensayng-nim-i}
Hyenwu-NOM teacher-HON-ACC teacher-HON-NOM
toy-ko siph-ta.
become-CONN want-DECL
'Hyenwu wants to become a teacher.'

There has been a lot of research on the phenomenon commonly referred to as case alternation in complex predicate constructions (cf. Chang & Cho 1991; Kim & Maling 1998; Um 2003; Chae 2015, a.o.). However, controversy persists over whether the grammatical case of the argument actually changes in constructions involving the auxiliary verb *siph-* 'want'. This paper proposes that there is no evidence to support a change in the grammatical case of the second argument in the construction with the auxiliary verb *siph-* 'want', by examining the function of the Korean particle *-i/ka*. Instead, it is argued that there is potential for emphasizing the second argument of the complex predicate *-ko siph-* 'want to'. Furthermore, it contributes to this discussion by demonstrating that the particle *-i/ka* can also function as an information structure marker, specifically indicating focus.

Based on this observation, this paper provides a new account of case alternation in the Korean complex predicate within the framework of Head-driven Phrase Structure Grammar (HPSG, Pollard & Sag 1994, Müller et al. 2021).

This paper is structured as follows: In Section 2, I provide a brief review of previous analyses of the phenomenon. This is followed by an exploration of the function of the particle *-i/ka* in Korean, based on double nominative constructions. Section 4 examines the case alternation phenomenon in the

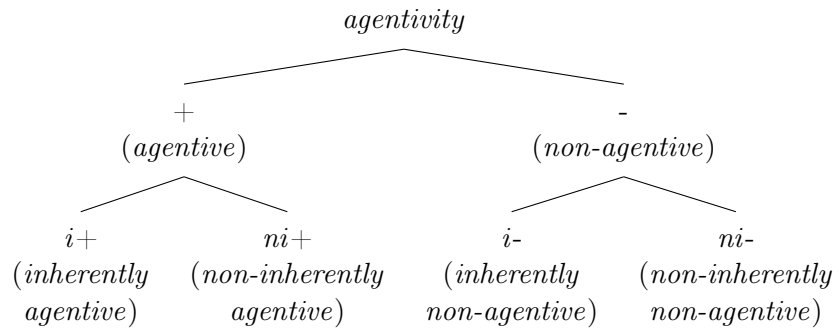


Figure 1: Classification of agentive types

complex predicate *-ko siph-* 'want to', specifically investigating whether the grammatical case changes from accusative to nominative. Formal analyses of this phenomenon in the Korean complex predicate *-ko siph-* 'want to' are presented in Section 5. The final section summarizes the conclusions of the study.

2 Previous analyses

It has been argued that predicates exhibit varying degrees of agentivity, typically contingent upon the presence or absence of agent subjects (cf. Yoo 2002: 1026; Kim 2016: 76–77). For the determination of structural case values, predicates possess [Agentive +/-] values (henceforth, [AG ±]), which are broadly based on whether they have agentive subjects (cf. Kim 1990; Bratt 1997). It has also been proposed to classify the AG value in the type hierarchy, as shown in Figure 1 (cf. Yoon 2012: 1026).[1]

While the agentive value of non-auxiliary verbs can be determined lexically by considering their argument structure and content value, auxiliary verbs demonstrate transparency regarding their agentive value. They inherit the value of their embedded verbs. Specifically, it has been suggested that the auxiliary verb *siph-* 'want' has two lexical entries, as seen in (4) (cf. Yoon 2012: 1029).

When the auxiliary verb *siph-* 'want' does not express an agentive relation and combines with a verb, its agentive value is inherited from the embedded verb, as illustrated in (4a). Additionally, the auxiliary verb *siph-* 'want' can inherently exhibit a non-agentive value when combined with an inherently agentive verb, as shown in (4b).

[1]In Figure 1 the values such as *i+*, and *ni+* are used as shorthand for the full value names in the parentheses, i.e., *inherently agentive*, *non-inherently agentive*.

(4) *siph-* 'want'

a. $\left[\text{AG } ni\ \alpha,\ \text{GOV}\langle \text{V[AG } \alpha]\rangle\right]$

b. $\left[\text{AG } i\text{–},\ \text{GOV}\langle \text{V[AG } i\text{+}]\rangle\right]$

It has been argued that the sentence in (5) illustrates only one possible constituent structure. However, the two potential AG values of the complex predicate with the auxiliary verb *siph-* 'want' result in different case values. Specifically, when the auxiliary verb *siph-* 'want' does not exhibit agentive values, its agentive value is passed on from the embedded verb (e.g. *mek-* 'eat' in (5)), and the second argument is realized as an accusative NP (e.g. *sakwa-lul* 'apple-ACC' in (5)), according to the AVM in (4a). On the other hand, when the auxiliary verb *siph-* 'want' inherently has a non-agentive value, the second argument can be realized as a nominative NP (e.g. *sakwa-ka* 'apple-NOM' in (5)). This is because the auxiliary verb *siph-* 'want' can also take on a non-agentive value, as demonstrated by the AVM in (4b).

(5) na-nun {sakwa-lul / sakwa-ka} mek-e po-ko siph-ta.
I-TOP apple-ACC apple-NOM eat-CONN try-CONN want-DECL
'I want to try to eat an apple.'

nanun	sakwalul/sakwaka	meke	poko	siphta.
	ACC	[AG *i*+]	[AG *i*+]	[AG *ni*+]
	NOM	[AG *i*+]	[AG *i*+]	[AG *i*-]

(from Yoo 2002: 1031)

In contrast, when *siph-* 'want' combines with a non-agentive verb, as shown in (6), the entire complex predicate is simply [AG *ni*-], because (4a) does not apply.

(6) nay-ka {*tayphyo-lul / tayphyo-ka} toy-ko siph-ta.
I-NOM chef-ACC chef-NOM become-CONN want-DECL
'I want to become a chef.'

nay-ka	tayphyo-ka	toy-ko	siph-ta.
	NOM	[AG *i*-]	[AG *ni*-]

(from Yoo 2002: 1031)

However, I propose that there is no evidence to support that the grammatical case of the second argument in the construction with the auxiliary verb *siph-* 'want' was changed. Additionally, it does not strictly necessitate the dual lexical entry of *siph-* 'want'. Instead, I argue that the potential in the complex predicate *-ko siph-* 'want to' exists to emphasize the second argument through an information structure marker. For this assumption,

I contribute by demonstrating that the particle *-i/ka* can also serve as an information structure marker, particularly indicating focus.

3 The particle *-i/ka* in Korean

The Korean particle *-i/ka* is widely recognized as a subject case marker, signifying that the nominal phrase with *-i/ka* serves as the subject of a sentence as seen in (7). According to this explanation, it can be used to identify double-subject constructions in Korean. Ko (2001: 12–16) has suggested these constructions depend on the semantic property of the verb, specifically on the factor of agentivity. It refers to the degree of control or volition exerted by the subject of a verb in an action. Testing for the property of agentivity involves verifying whether an event can be appropriately modified by the adverb 'intentionally' as shown in the examples (8a) and (8b) (cf. Verhoeven 2010: 224–227). It has been argued that double-subject constructions can be formed with verbs that do not have the property of agentivity. Accordingly, the double-subject constructions can only be formed with adjectives as seen in the example (8a), and with non-agentive verbs as seen in (8b) (from Ko 2001: 13).

(7) Cwunhuy-ka ilccik hakkyo-ey ka-n-ta.
Cwunhuy-NOM early school-LOC go-PRS-DECL
'Cwunhuy goes to school early.'

(8) a. ku salam-i son-i (*uytocekulo) kkway khu-ta.
the person-NOM hand-NOM intentionally pretty big-DECL
'The person's hands are (*intentionally) pretty big.'

b. namwu-ka saylo iph-i (*uytocekulo) tot-ass-ta.
tree-NOM newly leaf-NOM intentionally sprout-PST-DECL
'New leaves (*intentionally) sprouted on the tree.'

On the other hand, it has been claimed that although double nominative NPs may appear in a sentence, only one constituent can function as the subject (cf. Song 2009: 454–468; Kim et al. 2007: 25–29). These constructions will therefore be referred to as double nominative constructions[2]. In sentences (8a) and (8b), the first NP is not an obligatory argument. They remain grammatically well-formed even when the initial NP is omitted, as you can see in (9a) and (10a). Additionally, the particle *-i/ka* attached to the first NP (e.g. *ku salam-i* 'the person-NOM' in (8a)) can be replaced by the genitive particle *-uy*, as shown in (9b), and the first NP (e.g. *namwu-ka* 'tree-NOM' in (8b)) can also be replaced with the PP, as seen in (10b).

[2]In this paper, it is crucial to maintain a strict differentiation between 'double nominative constructions', where two NPs marked with *-i/ka* are present in a sentence, and 'double-subject constructions' defined as sentences containing two subjects.

(9) a. son-i kkway khu-ta.
hand-NOM pretty big-DECL
'The hands are pretty big.'

b. ku salam-uy son-i kkway khu-ta.
the person-GEN hand-NOM pretty big-DECL
'The person's hands are pretty big.'

(10) a. saylo iph-i tot-ass-ta.
newly leaf-NOM sprout-PST-DECL
'New leaves sprouted.'

b. namwu-ey saylo iph-i tot-ass-ta.
tree-PREP newly leaf-NOM sprout-PST-DECL
'New leaves sprouted on the tree.'

Based on this observation, it is essential to consider the functions of the first NP marked with *-i/ka* in a sentence, specifically whether it serves as the subject of a sentence. Various methods have been suggested to test for subjecthood, including the agreement with honorific inflected form *-(u)si*[3] and plural morpheme *-tul*, the scope of adverbs, the substitutability of a Korean subject case marker for the honorific form *-kkeyse*, and the acceptability of relative clauses (cf. Hong 1994: 100–115; Rhee 1999: 401–413; Park 2004: 107–110; Park & Kim 2022: 1504–1507, a.o.). To examine whether the first NP assumes the role of the subject in the sentence, it will be shown with some tests in this paper.

Firstly, the agreement of the inflected form *-(u)si* is assessed. It is claimed that the first nominative NP does not fulfill the subject function, as evidenced by the sentence (11a). Namely, since the first nominative NP (e.g. *sensayng-nim-i* 'teacher-HON-NOM') does not function as the subject of the sentence, it is not acceptable for it to agree with the predicate inflected with the honorific form *-(u)si* (e.g. *chincelha-si-ta* 'kind-HON-DECL'). If the first nominative NP (e.g. *sensayng-nim-i* 'teacher-HON-NOM') is the subject of the sentence, it should agree with the honorific-inflected predicate (e.g. *chincelha-si-ta* 'kind-HON-DECL'). However, this agreement is not observed, which suggests that the first nominative NP may not function as the subject in (11a). Otherwise, since the subject in the sentence (11b) is the second nominative NP (e.g. *sensayng-nim-i* 'teacher-HON-NOM'), it can be agreed with the predicate with the inflected form *-(u)si* (e.g. *yeyppu-si-ta* 'pretty-HON-DECL').

(11) a. sensayng-nim-i haksayng-i chincelha-ta / *chincelha-si-ta.
teacher-HON-NOM student-NOM kind-DECL kind-HON-DECL
'The teacher's student is kind.'

[3]The Korean honorific system requires that when the subject is in the honorific form (usually with the marker *-nim*), the predicate also be inflected with the honorific form *-(u)si*. (cf. Kim 2016: 318)

b. chinkwu-ka sensayng-nim-i yeyppu-ta / yeyppu-si-ta.
friend-NOM teacher-HON-NOM pretty-DECL pretty-HON-DECL
'The friend's teacher is pretty.'

Secondly, the particle *-i/ka* of the first nominative NP (e.g. *sensayng-nim-i* 'teacher-HON-NOM') cannot be substituted by the subject case marker for the honorific form *-kkeyse*, as shown in (12). Based on this evidence, it is claimed that only the second nominative NP in a sentence functions as a subject, even though there are two nominative NPs in the sentence.[4]

(12) a. sensayng-nim-i haksayng-tul-i chincelha-ta.
teacher-HON-NOM student-PL-NOM kind-DECL
'The teacher's students are kind.'

b. * sensayng-nim-kkeyse haksayng-tul-i chincelha-ta.
teacher-HON-HON.NOM student-PL-NOM kind-DECL

Hence, a question arises regarding the role of the first NP in a sentence. I assume that the first NP marked with *-i/ka* does not serve as the subject of a sentence and the particle *-i/ka* is used as an information structure marker (cf. Park 2004: 113–114; Kim et al. 2007: 27–35; Kim 2014: 13–14; Kim 2015: 45–50, a.o.). This research suggests that the Korean particle system is initially classified into case markers and information structure markers, with the former being further subdivided into the structural and lexical case (cf. Kim 1990; Hong 1992; Lee 2006: 86–87). As represented in Figure 2, I propose that the particle *-i/ka* functions as a case marker and an information structure marker. An information structure marker is defined as a particle that adds information structure properties—such as focus or topic—to the NP.

[4]In sentences where the semantic relation between the first nominative NP and the second NP is 'object-property' or 'whole-part', some subjecthood tests are met only when the referent of the first NP is identical to the referent of the possessor of the second NP, as seen in (i) and (ii). However, when the referent of the first NP differs from the referent of the second NP's possessor, the subjecthood tests are not satisfied (cf. Lee 2018: 286–290). According to Lee (2018), the first nominative NPs in state-property adjective sentences are not arguments of the adjectives but syntactic topics of the sentences.

(i) sensayng-nim-i khi-ka khu-ta / khu-si-ta.
teacher-HON-NOM height-NOM tall-DECL tall-HON-DECL
'The teacher is tall.'

(ii) sensayng-nim-kkeyse khi-ka khu-ta.
teacher-HON-HON.NOM height-NOM tall-DECL

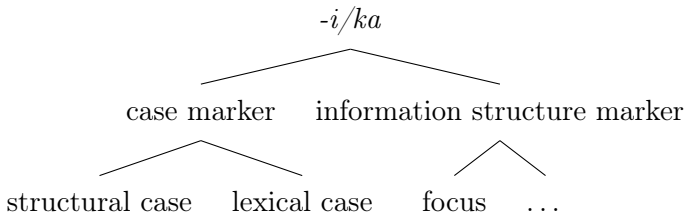


Figure 2: Different functions of the particle *-i/ka* in Korean

4 Reevaluating case alternation in the complex predicate *-ko siph-* 'want to'

This section investigates the phenomenon of case alternation in the complex predicate construction with the auxiliary verb *siph-* 'want' (cf. Kim & Maling 1998; Jung 2011). When the auxiliary verb *siph-* 'want' combines with verbs, the particle[5] *-i/ka* can be attached to the second argument of the complex predicate *-ko siph-* 'want to', as already seen in (3a). Some researchers argue that the grammatical case of the second argument changes from accusative to nominative. This section will test whether the second argument marked with the particle *-i/ka* functions as the subject of a sentence.

I now proceed with testing the subjecthood of the second argument (cf. Lee 2016a: 281–297; Park & Kim 2022: 1504–1508). To begin with, as seen in (13b), when examining the agreement between the predicate inflected in the honorific form *–(u)si* and the immediately preceding NP marked with *-i/ka*, it becomes evident that subjecthood is not confirmed. This is to say that the example (13b) demonstrates that the predicate (e.g. *siph-usi-ta* 'want-HON-DECL') cannot agree with the second NP (e.g. *halmeni-ka* 'grandmother-ka'). Otherwise, the first NP (e.g. *halmeni-ka* 'grandmother-NOM' in (13a)) can agree with the predicate inflected in the honorific form *–(u)si*, This implies that the second NP of the complex predicate *-ko siph-* 'want to' does not function as the subject of the sentence.

(13) a. halmeni-ka soncwu-ka po-ko siph-usi-ta.
grandmother-NOM grandchild-NOM see-CONN want-HON-DECL
'The grandmother wants to see a grandchild.'

b. *soncwu-ka halmeni-ka po-ko siph-usi-ta.
grandchild-NOM grandmother-NOM see-CONN want-HON-DECL
(Lit.) 'The grandchild wants to see a grandmother.'

Regarding the test of the plural morpheme *-tul*, the morpheme can recur in a sentence, when the subject is plural. Through this examination, it also

[5]In this paper, the term "particle" is introduced as a supertype to delineate between the case marker and information structure marker categories.

becomes evident that the second NP (e.g. *cokha-tul* 'niece-PL' in (14b)) does not function as the subject, as it fails to correlate with the embedded verbal element marked with the plural morpheme *-tul*. Conversely, the initial NP marked for plurality (e.g. *samchon-tul* 'uncle-PL' in (14a)) appropriately coincides with the verbal element affixed with the plural morpheme *-tul*.

(14) a. samchon-tul-i cokha-ka po-ko / po-ko-tul siph-ta.
uncle-PL-NOM niece-NOM see-CONN see-CONN-PL want-DECL
'The uncles want to see a niece.'

b. samchon-i cokha-tul-i po-ko / *po-ko-tul siph-ta.
uncle-NOM niece-PL-NOM see-CONN see-CONN-PL want-DECL
'An uncle wants to see nieces.'

Furthermore, as for the substitutability of the Korean subject case marker for the honorification *-kkeyse*, if the subject were *apeci*[6] 'father' in the sentence (15), the honorific marker *-kkeyse* would be applicable for subject honorification. The example sentence (15) illustrates that the second NP marked with *-i/ka* cannot be substituted with the subject honorific marker *-kkeyse*. This shows the second NP (e.g. *apeci* 'father' in (15)) does not serve as the subject.[7]

(15) Hyenwu-ka apeci-ka / *apeci-kkeyse po-ko siph-ta.
Hyenwu-NOM father-NOM father-HON.NOM see-CONN want-DECL
'Hyenwu wants to see his father.'

The examples (13) - (15) for the subjecthood test indicate that the second argument marked with *-i/ka* does not exhibit subjecthood in a sentence (cf. Lee 2016a: 281–297; Park & Kim 2022: 1504–1508). As indicated by the subjecthood test, I argue that the grammatical case of the second argument remains unchanged. Instead, I propose that the particle *-i/ka* attached to the second argument of the complex predicate *-ko siph-* 'want to' functions as an information structure marker, thereby emphasizing the constituent.

[6]The NP *apeci* 'father' can be honored with the subject honorific marker *-kkeyse*

[7]When the constituent serves as the subject of a sentence, the particle *-i/ka* attached to the first NP *apeci-ka* 'father-NOM' in the example (i.a) can be substituted with the subject honorific form *-kkeyse*, as seen in (i.b).

(i) a. apeci-ka atul-i po-ko siph-ta.
father-NOM son-I see-CONN want-DECL
'The father wants to see his son.'

b. apeci-kkeyse atul-i po-ko siph-ta.
father-HON.NOM son-I see-CONN want-DECL

5 Analysis for constructions with the complex predicate *-ko siph-* 'want to'

This section[8] analyzes the previously mentioned phenomenon in the complex predicate construction with the auxiliary verb *siph-* 'want' within an HPSG framework. Firstly, I propose the lexical entry of the auxiliary verb *siph-* 'want' as seen in (16) (cf. Müller 2002: 86; Müller 2013: 243). The auxiliary verb *siph-* 'want' combines with the dependent verbal element with the connective marker *-ko*[9]. I assume that the auxiliary verb *siph-* 'want' functions as the head of the complex predicate. The lexical entry (16) specifies that the auxiliary verb functioning as the head combines with a complement that has a [LEX +] value. The arguments of the embedded verb are attracted to be the arguments of the complex predicate (cf. Hinrichs & Nakazawa 1989; Hinrichs & Nakazawa 1994). The combination of the auxiliary verb *siph-* 'want' and its complement verbal element (e.g. *mek-ko* 'eat-CONN') is represented in Figure 3.

(16) Lexical entry of the auxiliary verb *siph-* 'want':

$$\left[\begin{array}{l}\text{PHON} \langle\ \textit{siph}\ \rangle \\ \\ \text{SYNSEM|LOC|ARG-ST}\ \boxed{1} \oplus \boxed{2} \oplus \boxed{3} \left\langle \text{V}\left[\begin{array}{r}\text{VFORM}\ \ \textit{-ko}\ ,\ \text{LEX}+, \\ \text{SUBJ}\ \boxed{1},\ \text{COMPS}\ \boxed{2}\ \end{array}\right] \right\rangle\end{array}\right]$$

As already indicated in (3a)—repeated here as (17)— the particle *-i/ka* can be attached to the second argument of the complex predicate *-ko siph-* 'want to'. According to the subjecthood test in section 4, it was confirmed that the particle *-i/ka* attached to the second argument does not function as the subject case marker. Instead, I propose that it may serve as an information structure marker. In this paper, the information structure introduce as part of the CONTEXT value. This is because, although the proposition conveyed by both sentences in (17a) and (17b) remains the same, the speaker's intention is additionally included by means of the information structure marker *-i/ka* attached to the second argument (e.g. *sakwa-ka* 'apple-KA' in (17b)).[10] I assume that using the information structure marker *-i/ka* in the complex predicate construction with *siph* 'want' creates a contrastive expression. This

[8]Previously, the particle *-i/ka* was glossed as NOM. However, from now on, when it is used as an information structure marker in the complex predicate *-ko siph-* 'want to', it should be glossed as -I/KA.

[9]*-ko* is a connective ending that links the preceding verb to another verb. In other words, the auxiliary verb *siph-* 'want' only selects the embedded verbal element with this connective marker *-ko*. Therefore, it could be said that the dependent verbal element *mek-ko* 'eat-CONN' is an inflected form of the verb *mek-* 'eat', allowing it to combine with the auxiliary verb *siph-* 'want'.

[10]Additionally, the first argument, which functions as the subject, can also be topicalized using the information structure marker *-un/nun*. Building upon this, the following sentence (i) can be constructed.

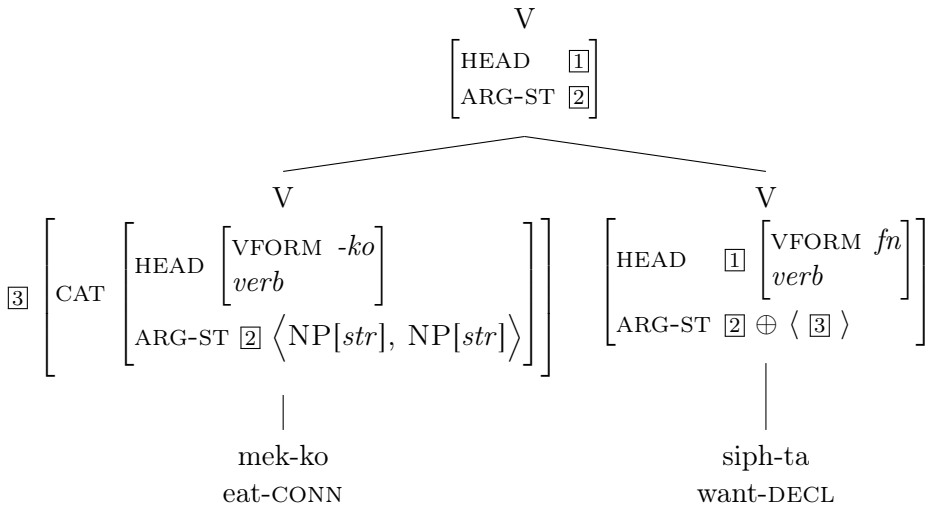


Figure 3: Analysis of *mek-ko siph-ta* 'want to eat'

is illustrated in the following example: *It is an apple that Hyenwu wants to eat, not a pear.* So, I propose that the information structure marker *-i/ka*, by its inherent characteristics, serves a semantic function of focusing on the constituent. However, when the NP marked with the marker *-i/ka* is realized, it creates a contextual meaning, particularly in the complex predicate construction with *siph-* 'want'.

The AVM for the information structure marker *-i/ka* is suggested as shown in (18). The MKG (MarKinG) value is indicated as *fc*[11] and is not co-referenced with any other element[12]. Additionally, the ICONS (Individual CONStraints) element specifies the relation of the complement. The ICONS-KEY feature is used to impose a more specific constraint to an information structure element that has already been enhanced within the ICONS list. This implies that, given the particle *-i/ka*'s ability to function as markers for various information structures such as topic or focus, it becomes necessary to

(i) Hyenwu-nun sakwa-ka mek-ko siph-ta.
Hyenwu-TOP apple-FOC eat-CONN want-DECL
'HYENWU wants to eat an APPLE.'

[11]MKG features are exclusively concerned with markings of information structure. The types of MKG are underspecified with regard to *fc* (focus), *non-fc* (non-focus), *tp* (topic), and *non-tp* (non-topic). (Song 2017: 121–124)

[12]The information structure marking is encoded through a morphosyntactic feature MKG, within SYNSEM|CAT. This feature imposes lexical and syntactic constraints on forms that convey information structure meanings. MKG features specifically address information structure markings. While the MKG value represents the morphosyntactic marking, it does not always align with the semantic value(cf. Song 2017: 121–125).

constrain its meaning accordingly (cf. Song 2017: 118).[13]

(17) a. Hyenwu-ka sakwa-lul mek-ko siph-ta.
Hyenwu-NOM apple-ACC eat-CONN want-DECL
'Hyenwu wants to eat an apple.'

b. Hyenwu-ka sakwa-ka mek-ko siph-ta.
Hyenwu-NOM apple-KA eat-CONN want-DECL

(18) *-i/ka marker* ⇒

$$\begin{bmatrix} \text{PHON} & \langle\ i/ka\ \rangle \\ \text{ICONS-KEY} & \boxed{2} \\ \text{MKG} & fc \\ \text{COMPS} & \left\langle \left[\text{INDEX } \boxed{1}\right] \right\rangle \\ \text{ICONS} & \left\langle ! \ \boxed{2} \begin{bmatrix} focus \\ \text{TARGET } \boxed{1} \end{bmatrix} \ ! \right\rangle \end{bmatrix}$$

As previously mentioned in (3b)—repeated here as (19)— it should be noted that when the auxiliary verb *siph-* 'want' is combined with a non-agentive verbal complement like *toy-* 'become', the second argument is not grammatically allowed to bear the accusative case marker *-ul/lul*. I suggest that since the verb requires two arguments with structural and lexical case (cf. Müller 2002: 12–16; Müller 2013: 221–225), it is not acceptable for the argument with lexical case to be changed to another marker, such as the accusative marker *-ul/lul*. As indicated in the lexical entries (20) and (21), since the verb *mek-* 'eat' selects two arguments with a structural case, the second argument with a structural case can be focused using the information structure marker *-i/ka*. Conversely, when the argument of the verb *toy-* 'become' has a lexical case, it cannot be exchanged for another marker different from the nominative marker *-i/ka* regardless of the syntactic structure.[14]

[13]As seen in the works of Song (2017) and Song & Bender (2012), the ! symbol is used to mark the boundaries of a *diff-list* representation for RELS, HCONS, and ICONS. This notation helps to clearly delineate the different parts of the list, making the relationships between grammatical information and semantic constraints more explicit. Using the symbol !, we can efficiently handle these constraints, ensuring that the boundaries between different components are easily identifiable. So, this paper accepts the use of the *diff-list* representation and the ! symbol, as proposed by Song (2017) and Song & Bender (2012).

[14]As already mentioned in this paper, the second argument marked with the nominative lexical case marker *-i/ka* of the complex predicate *toy-ko siph-ta* 'become-CONN want-DECL' cannot be attached to the accusative marker *-ul/lul*. However, this might not apply to auxiliary particles. In Korean, particles that add specific meanings are often referred to as auxiliary particles. These particles can be also attached to the second argument of verbs, such as *toy-* 'become', as illustrated in the example sentence (i). In the sentence (i), the use of the particle *-to* adds the meaning of 'also'.

(i) Hyenwu-ka sensayngnim-to toy-ko siph-ess-ta.
Hyenwu-NOM teacher-TO become-CONN want-PST-DECL
'Hyenwu also wanted to become a teacher.'

(19) a. Hyenwu-ka sensayng-nim-i toy-ko siph-ta.
Hyenwu-NOM teacher-HON-NOM become-CONN want-DECL
'Hyenwu wants to become a teacher.'

b. Hyenwu-ka *sensayng-nim-ul toy-ko siph-ta.
Hyenwu-NOM teacher-HON-ACC become-CONN want-DECL

(20) *mek-* 'eat':

$$\begin{bmatrix} \text{PHON} & \langle\ mek\ \rangle \\ \text{SYNSEM|LOC|ARG-ST} & \left\langle \text{NP}[str],\ \text{NP}[str] \right\rangle \end{bmatrix}$$

(21) *toy-* 'become':

$$\begin{bmatrix} \text{PHON} & \langle\ toy\ \rangle \\ \text{SYNSEM|LOC|ARG-ST} & \left\langle \text{NP}[str],\ \text{NP}[lex] \right\rangle \end{bmatrix}$$

To summarize what has been said so far, when the argument of a verb can be assigned to structural cases, the second argument can be marked with the information structure marker *-i/ka* to indicate focus, particularly when combined with the auxiliary verb *siph-* 'want'. Based on this assumption, I propose that the lexical rule for the auxiliary verb *siph-* 'want' allows the second argument—namely, the accusative object— to be emphasized with the information structure marker *-i/ka* when it combines with a verb that has agentivity, as shown in (22).[15] The structure of the sentence (17b) is given in Figure 4.

(22) The lexical rule for the auxiliary verb *siph-* 'want' with agentive verbs:

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & verb \\ \text{ARG-ST} & \left\langle \boxed{1}\begin{bmatrix} \text{LOC|CAT|HEAD} & \begin{bmatrix} noun \\ \text{CASE}\ str \end{bmatrix} \end{bmatrix}, \boxed{2}\begin{bmatrix} \text{LOC|CAT|HEAD} & \begin{bmatrix} noun \\ \text{CASE}\ str \end{bmatrix} \end{bmatrix} \right\rangle \\ & \oplus \boxed{3}\left\langle \text{V}\left[\begin{array}{r} \text{VFORM}\ \textit{-ko},\ \text{LEX}+,\ \text{SUBJ}\ \boxed{1}, \\ \text{COMPS}\ \boxed{2},\ \text{INDEX}\ \textit{agentive}\ \end{array}\right] \right\rangle \end{bmatrix} \end{bmatrix} \mapsto$$

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & verb \\ \text{ARG-ST} & \left\langle \boxed{1}\begin{bmatrix} \text{LOC|CAT|HEAD} & \begin{bmatrix} noun \\ \text{CASE}\ str \end{bmatrix} \end{bmatrix}, \boxed{2}\begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD}\ noun \\ \text{CASE}\ non\text{-}str \end{bmatrix} \\ \text{CTXT} & \begin{bmatrix} infostr \\ \text{FOCUS}\ + \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\ & \oplus \boxed{3}\left\langle \text{V}\left[\begin{array}{r} \text{VFORM}\ \textit{-ko},\ \text{LEX}+,\ \text{SUBJ}\ \boxed{1}, \\ \text{COMPS}\ \boxed{2},\ \text{INDEX}\ \textit{agentive}\ \end{array}\right] \right\rangle \end{bmatrix} \end{bmatrix}$$

The structures resulting from the combination of (16) and (20) are illustrated in (23) and (24). The entry (23) shows that the complex predicate

[15] In this paper, the value of INDEX, *agentivity* is used as an abbreviation to represent the index that serves as the argument of the agentive verb, reflecting its role in the verb's argument structure rather than an inherent property of the referent.

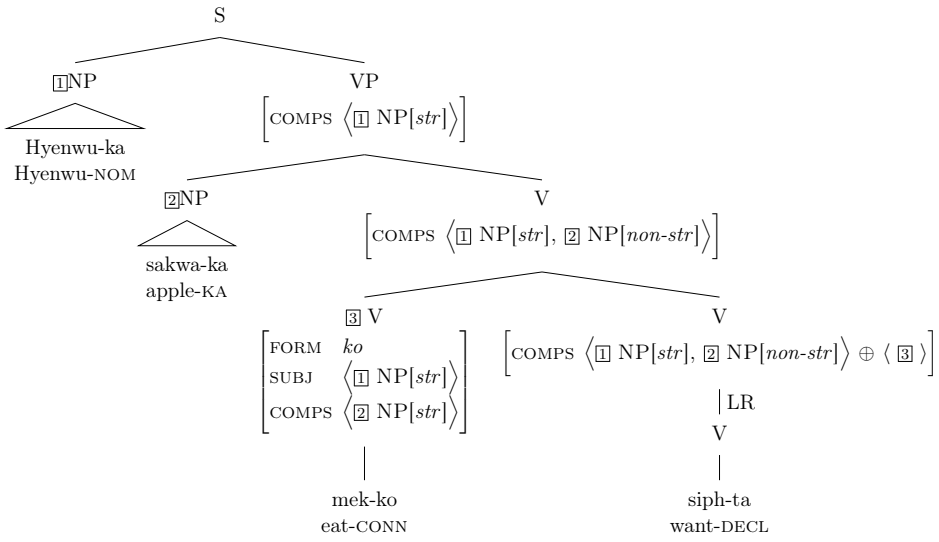


Figure 4: The structure of the complex predicate *-ko siph-* 'want to' with a focused second argument

mekko siphta 'want to eat' entails two arguments with structural cases, the second of which is marked with the accusative case marker *-ul/lul*. According to the lexical rule (22), when the auxiliary verb *siph-* 'want' combines with agentive verbs that select two arguments with structural cases, a non-structural case, namely the information structure marker *-i/ka*, can be attached to the second argument of the complex predicate *–ko siph-* 'want to' (e.g. *sakwa* 'apple' in (24)). When the second argument is focused using the information structure marker *-i/ka*, the sentence additionally acquires contextual value as information structure (cf. Paggio 2009: 105).

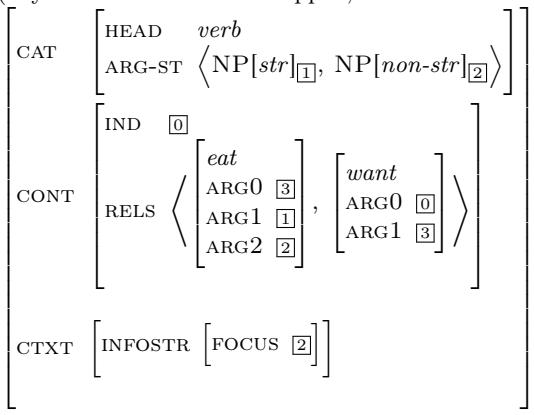
(23) *Hyenwu-ka sakwa-lul mek-ko siph-ta*
('Hyenwu wants to eat an apple'):

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{ARG-ST} & \left\langle \text{NP}[\textit{str}]_{\boxed{1}},\ \text{NP}[\textit{str}]_{\boxed{2}} \right\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{IND} & \boxed{0} \\ \text{RELS} & \left\langle \begin{bmatrix} \textit{eat} \\ \text{ARG0} & \boxed{3} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \end{bmatrix}, \begin{bmatrix} \textit{want} \\ \text{ARG0} & \boxed{0} \\ \text{ARG1} & \boxed{3} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

I assume that the two structures, namely the second argument marked with the information structure marker *-i/ka* and the accusative marker *-eul/reul*, deliver the same proposition: "He wants to eat an apple." However, by using the information structure marker *-i/ka*, the speaker indicates their intention

to focus on the complement. To account for this point, the information is incorporated into the CONTEXT value, as illustrated in (24).[16]

(24) *Hyenwu-ka sakwa-ka mek-ko siph-ta*
('Hyenwu wants to eat an apple', with the focused element):



[16]In response to a reviewer's query about how the analysis addresses instances where an auxiliary verb like *ha-* 'do' is followed by a verb such as *siph-* 'want'. I provide the following comment. It is proposed that the verb *ha-* 'do' is polysemous, necessitating a detailed classification of its various functions, as multiple types can be discerned in constructions involving the verb *ha-* 'do'. For instance, when the verb *ha-* 'do' is combined with a verbal noun (e.g. *kongpwu* 'study'), both markers *-ul/lul* and *-i/ka* can be attached to the second argument of the complex predicates *-ko siph-* 'want to', as seen in (i). However, when the verb *ha-* 'do' is combined with a stative verb (e.g. *mwusep-* 'fearful'), it is unacceptable for the accusative case marker *-ul/lul* attached to the second argument to change to the marker *-i/ka*, as shown in (ii). This structure, involving a stative verb, is perceived to lack the property of agentivity, as evidenced by its limited compatibility with modifying adverbs such as 'intentionally', as demonstrated in (iii.b), which differs from the example sentence (iii.a). Additional research will be conducted through testing with native Korean speakers to identify whether these stative verbs, when combined with the verb *ha-* 'do', have the semantic property of agentivity.

(i) Hyenwu-ka {yenge-lul / yenge-ka} kongpwu-ha-ko siph-ess-ta.
Hyenwu-NOM English-ACC English-KA study-DO-CONN want-PST-DECL
'Hyenwu wanted to study English.'

(ii) Hyenwu-ka {kangaci-lul / *kangaci-ka} mwuse-we ha-ko
Hyenwu-NOM puppy-ACC puppy-KA fearful-CONN DO-CONN
siph-ess-ta.
want-PST-DECL
'Hyenwu wanted to be afraid of the puppy.'

(iii) a. Hyenwu-ka yenge-lul yilpwule kongpwu-hay-ss-ta.
Hyenwu-NOM English-ACC intentionally study-DO-PST-DECL
'Hyenwu intentionally studied English.'

b. ?Hyenwu-ka kangaci-lul ilpwule mwuse-we hay-ss-ta.
Hyenwu-NOM puppy-ACC intentionally fearful-CONN DO-PST-DECL
(Lit.) 'Hyenwu intentionally was afraid of the puppy.'

6 Conclusion

This research argued that the phenomenon in which the second argument of the complex predicate *-ko siph-* 'want to' is marked with the particle *-i/ka* should not be regarded as case alternation. Instead, I have argued in this paper that the second argument of the complex predicate *-ko siph-* 'want to' can be focused by means of the information structure marker *-i/ka*, suggesting that this construction may involve a focused NP.

To examine whether the second accusative argument of the complex predicate *-ko siph-* 'want to' changes to the nominative argument, I initially investigated the function of the particle *-i/ka* based on double nominative constructions. By conducting subjecthood tests—such as agreement with the honorific inflected form *-(u)si*, agreement with the plural morpheme *-tul*, and the substitutability of the Korean subject case marker for the honorific form *-kkeyse*— on the second argument, it was confirmed that the particle *-i/ka* does not always function as a subject case marker. Based on the test, I argued that the particle *-i/ka* can function as a case marker and an information structure marker, adding information structure properties to the NP. Moreover, I concluded that the particle *-i/ka*, when attached to the second argument in the complex predicate construction *-ko siph-* 'want to', does not change its grammatical case. Rather, it imbues an additional pragmatic meaning, particularly regarding information structure. This conclusion was further supported by the subjecthood test, which confirms that the second argument marked with the particle *-i/ka* does not function as the subject of a sentence.

I have integrated these findings into an HPSG fragment of Korean. It was claimed that when the auxiliary verb *siph-* 'want' combines with an agentive verb, the second argument with a structural case can be realized through the accusative case marker and can also be focused using the information structure marker *-i/ka*. I also hypothesized that the information structure marker *-i/ka* attached to the argument leads to additional interpretations in the context. Therefore, it was suggested that the value acquired from the information structure marker *-i/ka* is added to the CONTEXT feature, as the proposition of the sentence remains unchanged. To further elucidate the phenomenon in which it is impermissible for the marker attached to the second argument to change—particularly when the auxiliary verb *siph-* 'want' combines with a verb that selects two NP marked with *-i/ka* (e.g. *toy-* 'become')— the grammatical case was classified into the structural and lexical case.

Acknowledgements

I would like to express my sincere gratitude to Stefan Müller, Antonio Machicao y Priemer, Giuseppe Varaschin and Paola Fritz-Huechante, as well as to the anonymous reviewers, for their invaluable guidance, constructive feedback, and continuous support throughout this research. I am also grateful to the participants of the 31st International Conference on Head-Driven Phrase Structure Grammar for their helpful comments on earlier versions of this paper.

References

Bratt, Elizabeth Owen. 1997. *Argument composition and the lexicon: lexical and periphrastic causatives in Korean*. stanford university.

Chae, Hee-Rahk. 2015. A comprehensive syntactic analysis of Korean [... v ... v] expressions i: auxiliary predicate constructions and case alternation. *Korean Journal of Linguistics,* 40(3). 543–578.

Chang, Jongwhan & Dongin Cho. 1991. Head movement and case theory. *Harvard Studies in Korean Linguistics IV*. 217–227.

Hinrichs, Erhard & Tsuneko Nakazawa. 1989. Flipped out: Aux in german. In *Papers from the 25th regional meeting of the Chicago Linguistic Society*, vol. 25, 193–202. Chicago, Illinois: CLS.

Hinrichs, Erhard & Tsuneko Nakazawa. 1994. Linearizing AUXs in German verbal complexes. In John Nerbonne, Klaus Netter & Carl J. Pollard (eds.), *German in Head-Driven Phrase Structure Grammar* (CSLI Lecture Notes 46), 11–37. Stanford University: CSLI Publications.

Hong, Ki-Sun. 1992. *Argument selection and case marking in Korean*. Ph.D. Dissertation. Stanford, CA: Stanford University.

Hong, Ki-Sun. 1994. Subjecthood tests in Korean. *Language Research*. 99–136.

Jung, Hyun-Kyoung. 2011. Nominative/accusative case alternation in the Korean 'siph-ta' construction. *Coyote Papers*. `http://hdl.handle.net/10150/140749`.

Kim, Ilkyu. 2014. On the meaning of Korean -i/ka. *Language and Linguistics* 63. 1–26.

Kim, Ilkyu. 2015. How can Korean -i/ka express exhaustivity? *Language and Information* 19(2). 37–53.

Kim, Jong-Bok. 2016. *The syntactic structures of Korean: A construction-based perspective*, vol. 1. Cambridge University Press.

Kim, Jong-Bok, Peter Sells & Jae-Hyung Yang. 2007. Parsing two types of multiple nominative constructions: A constructional approach. *Language and Information* 11(1). 25–38.

Kim, Soowon & Joan Maling. 1998. Case assignment in the siphta construction and its implications for case on adverbials. *Cornell East Asia Series* 98. 133–168.

Kim, Young-Joo. 1990. *The syntax and semantics of Korean case: the interaction between lexical and syntactic levels of representation.* Ph.D. Dissertation. Cambridge, MA: Harvard University.

Ko, Kwangju. 2001. Why does double subject phenomenon appear? *Korean Language Research* 9. 1–26.

Lee, Eun-Sub. 2016a. On the verifying the subjectivity of 'np+i/ka' preceding the main verb in '-ko shipta' construction. *Language and Literature* 134. 21–51.

Lee, Hanjung. 2006. Parallel optimization in case systems: evidence from case ellipsis in Korean. *Journal of East Asian Linguistics* 15(1). 69–96. DOI: 10.1007/s10831-005-4905-8.

Lee, Kwan-gyu. 2016b. *Korean Grammar (Revised Edition).* Seoul, South Korea: Yeokrak Publishing.

Lee, Ho-Seung. 2018. A syntactic topic of state-property adjective sentences and the subjecthood tests. *Language Information Society* 35. 273–301.

Müller, Stefan. 2002. *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German* (Studies in Constraint-Based Lexicalism 13). Stanford, CA: CSLI Publications. http://hpsg.fu-berlin.de/~stefan/Pub/complex.html.

Müller, Stefan. 2013. *Head-Driven Phrase Structure Grammar: Eine Einführung.* 3rd edn. (Stauffenburg Einführungen 17). Tübingen: Stauffenburg Verlag. https://hpsg.hu-berlin.de/~stefan/Pub/hpsg-lehrbuch.html.

Müller, Stefan, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.). 2021. *Head-Driven Phrase Structure Grammar: The handbook.* Berlin. DOI: 10.5281/zenodo.5543318.

Paggio, Patrizia. 2009. The information structure of Danish grammar constructions. *Nordic Journal of Linguistics* 32(1). 137–164. DOI: 10.1017/S0332586509002066.

Park, Chongwon & Jong-Bok Kim. 2022. Nominative objects in Korean. *Linguistics* 60(5). 1487–1537.

Park, Hogwan. 2004. The type and syntactic structure of a dual/subjective construction in Korean. *The Korean Language and Literature.* 99–118.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar* (Studies in Contemporary Linguistics). Chicago, London: University of Chicago Press.

Rhee, Seongha. 1999. On the multiple nominative constructions in Korean. In YoungWha Kim, IlKon Kim & JeongWoon Park (eds.), 198–430. Seoul: Hankuk Publisher.

Sohn, Ho-min. 1999. *The Korean Language.* Cambridge, UK: Cambridge University Press.

Song, Changseon. 2009. The critical remarks upon the double subject constructions in Korean. *The Journal of Korean Language and Literature Education* 45. 449–474.

Song, Changseon. 2020. *Korean Grammar*. Seoul, South Korea: Parkijeong.
Song, Sanghoun. 2017. *Modeling information structure in a cross-linguistic perspective*. Berlin: Language Science Press.
Song, Sanghoun & Emily M Bender. 2012. Individual constraints for information structure. In *Proceedings of the 19th international conference on head-driven phrase structure grammar*, 330–348.
Um, Jung-ho. 2003. Case alternation in "-go siphta" construction. *Journal of Korean Linguistics* 41. 169–196.
Verhoeven, Elisabeth. 2010. Agentivity and stativity in experiencer verbs: Implications for a typology of verb classes. *Linguistic typology* 14(2-3). 213–251.
Yoo, Eun-Jung. 2002. Auxiliary verbs and structural case assignment in Korean. *Korean Linguistic* 38(4). 1009–1036.
Yoon, Hangjin. 2012. Different types of kes constructions in Korean. *The Korean Generative Grammar* 22(3). 557–577.

Horn clauses and strict NPIs under negated matrix clauses

Manfred Sailer
Goethe-University Frankfurt a.M.

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 152–172

Keywords: Negative Polarity Items, Neg Raising, Negative Inversion, Horn Clauses, enriched semantic representation, collocation, semantics-pragmatics interface

Sailer, Manfred. 2024. Horn clauses and strict NPIs under negated matrix clauses. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 152–172. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.10.
CC-BY 4.0

Abstract

The present paper makes four main contributions: First, it argues for a surface-scope oriented approach to phenomena that have been considered strong arguments against surface scope of negation: the licensing of embedded strict Negative Polarity Items (NPIs) and embedded Negative Inversion (i.e. Horn clauses). Second, this surface-scope analysis is expressed within existing HPSG analyses of NPI licensing and Negative Inversion. Third, the strict/non-strict distinction of NPIs is reduced to universal vs. existential quantification over the licensing requirement within a semantic representation. Fourth, Negative Inversion is analyzed as a constructional NPI. The existence of constructional NPIs should not be surprising, but no such example has been previously discussed in HPSG to my knowledge.

1 Introduction

The paper investigates under which conditions strict Negative Polarity Items (such as *lift a finger*) and subject-auxiliary inversion can be licensed in an embedded clause through a negated matrix predicate. The most commonly known context for such non-local licensing is *Negation Raising* (Neg Raising), i.e., cases in which a negation that appears in the matrix clause is apparently interpreted in the embedded clause, as indicated in (1).

(1) I don't think [Chris won]. ⇒ I think [Chris didn't win].

This apparent low interpretation has been observed to correlate with the possible occurrence of strict Negative Polarity Items as in (2a), and of embedded "Negative" Inversion, so-called *Horn clauses*, see (2b).

(2) a. I *(don't) think [Chris will lift a finger].
b. I *(don't) think [that ever before has Chris been in Olomouc].

However, Horn (2014) and Hoeksema (2017) show that neither of these two phenomena is restricted to Neg Raising. I will take their empirical observations as the basis to develop a more fine-grained characterization and analysis of Negative Polarity Item licensing in general, and in HPSG in particular.

Before going into the more detailed discussion, I need to introduce some terminology. I distinguish three licensing aspects for Negative Polarity Items (NPIs): Strength, at-issueness, and locality. Each of these aspects has been used in the literature before, however, their independence is usually not discussed. I will introduce a typographic marking for each dimensions at first mention.

[†]This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — project number 509468465 — CRC 1629 "Negation in Language and Beyond", projects A04 (*Resolving the Neg Raising Paradox*) and B05 (*Negative Polarity Items in Non-negative Contexts*). I would like to thank my project colleagues for comments and discussion: Nicolas Lamoure, Zahra Mirrazi, Frank Richter, and Hedde Zeijlstra. I presented parts of this paper at the *MECORE Closing Workshop*, Constance, June 2024. I am grateful to the reviewers and the audiences of that event and of *HPSG 2024* for their comments. All errors are mine.

Strength: strong vs. weak NPIs (van der Wouden 1997) Strong NPIs require a verbal negation or a negative indefinite as licenser, but are not licensed by a "weak" licenser like *not every* or *few*, see (3).

(3) a. Strong NPI: No one/*Not everyone lifted a finger to help Alex.
b. Weak NPI: No one/Not everyone has ever helped Alex.

At-issueness: ⌜regular⌝ vs. ⌞lexical⌟ NPIs (Sedivy 1990) Lexical NPIs can be licensed pragmatically, as in (4a) to reject the claim that Cynthia never lifts a finger. Regular NPIs require an overt (i.e. at-issue) licenser: ⌞*at all*⌟ is not licensed in (4b) even in a context where it is claimed that Bert doesn't care about the homeless.

(4) a. Cynthia DOES ⌞lift a finger⌟ when there is work to be done.
b. * Bert DOES care about the homeless ⌜at all⌝. (Sedivy 1990: 98)

Locality: strict vs. non-strict NPIs (Hoeksema 2017) Non-strict NPIs can occur embedded under negated factive predicates, but strict NPIs cannot, (5).[1]

(5) a. he didn't know [that the building had ever been used as a dry cleaner . . .] (English Trends)
b. * he didn't know [that the building was **all that** old].

From here on, I will typographically indicate all three NPI-licensing aspects. For example, ⌜*ever*⌝ is a weak, ⌜regular⌝, non-strict NPI. In its NPI-use, ⌜***need***⌝ is weak, ⌜regular⌝, and **strict**. The NPI ⌞***lift a finger***⌟ is strong, ⌞lexical⌟, and **strict**. Furthermore, I will put overt licensers in **bold face**.

I will argue for the relevance of distinguishing these licensing aspects in Section 2. I do not claim, however, that they constitutes an exhaustive classification of NPIs. For example, van der Wouden (1997) shows that there are more subcategories with respect to strength than just two. Also, I do not consider the NPI licensing in questions, which represent yet another aspect. Section 3 presents a theory of NPI licensing based on the generalizations from Section 2. Section 4 provides an HPSG encoding, and Section 5 is a short conclusion.

2 Generalizations about NPI types

In this section I will propose the following three generalizations that seem to be not always prominent in the literature.

G-1 Licenser strength and locality requirement are independent of one another

G-2 Embedded licensing of strict NPIs does not require Neg-Raising

[1] The corpus *English Trends* is available via `www.sketchengine.eu` (Kilgarriff et al. 2014).

G-3 Horn clauses behave exactly like strict NPIs

I will discuss data from the literature that support these generalizations.

2.1 Ad G-1: Independence of strength and locality

In many papers on NPI licensing in embedded clauses, all examples of strict NPIs are also strong NPIs. Usually, we find ⌞***lift a finger***⌟ and ⌜***until***⌝. But modal auxiliary ⌜***need***⌝ is a strict NPI as well, though it is weak, as it is licensed by *few*, (6a). However, as a strict NPI, it is not licensed by negated factive *know*, (6b).

(6) a. Of course, **not every** criticism ⌜**need**⌝ be taken at face value
b. Boris Johnson does **not** think/*know that rules ⌜**need**⌝ apply to him
(grammatical versions from English Trends)

I list the classifications of some NPIs in Table 1. This table shows that there are weak strict NPIs. However, the table only shows five out of eight possible combinations of the three licensing aspect. The missing combinations are: strong non-strict lexical; weak non-strict regular, and strong strict lexical. Or, expressed differently: All strong NPIs and all lexical NPIs I looked at are strict. It is an empirical question if these non-listed combinations exist.[2] NPIs such as ⌜***need***⌝ and ⌞***all that***⌟ are weak but strict, showing that not all strict NPIs are strong. Strict NPIs can be strong or weak, and lexical or regular. However, all non-strict NPIs might be weak.

	strength	at-issueness	locality	
ever, *any*	weak	⌜regular⌝	non-strict	(Sedivy 1990)
NPI *need*, *at all*	weak	⌜regular⌝	**strict**	
all that Adj.	weak	⌞lexical⌟	**strict**	(Horn 2014)
until, *either*	strong	⌜regular⌝	**strict**	
lift a finger	strong	⌞lexical⌟	**strict**	(Sedivy 1990)

Table 1: NPIs, classified by strength, at-issueness, and locality

2.2 Ad G-2: Strict NPIs under negated matrix predicates

The contrast in (5) has led to the simplified assumption that Neg Raising is the only constellation in which strict NPIs can be licensed by a negation in the matrix clause. Horn (2014) lists instances of Horn clauses and strict NPIs under negated non-Neg-Raising predicates, such as non-factive uses of *know* in (7).

[2] A comprehensive list of English NPIs with licensing contexts, comparable to Hoeksema (2024) for Dutch, is a research desideratum. A good collection of English NPIs is provided in von Bergen & von Bergen (1993), but they do not discuss licensing profiles in detail.

(7) Strict NPI: I do**n't** know [that this is ⌞**all that**⌟ complicated].
≠ I know that it is **not** ⌞**all that**⌟ complicated.

Note that in (7), the negation is not interpreted in the embedded clause, i.e., there is no *Neg-Raising inference*. Instead the version with matrix negation expresses a lower degree of certitude than the one with negation in the embedded clause. This can be seen over and over in occurrences of strict NPIs under negated non-factive matrix predicates, as in (8)–(10).

(8) I do**n't** know [that it ⌜**need**⌝ create any serious difficulties].
≠ I know [that it ⌜**need**⌝ **not** create any serious difficulties].

(9) But that does**n't** mean [that she ⌜**need**⌝ correspond to contemporary notions of what a feminist should be] (English Trends)
≠ that means [that she ⌜**need**⌝ **not** correspond . . .]

(10) I'm **not** sure [he's done ⌞**a damn thing**⌟ to correct it] (Hoeksema 2017)
≠ I'm sure [he has**n't** done ⌞**a damn thing**⌟ to correct it] . . .

Hoeksema (2017) explores the properties of negated matrix predicates that allow for embedded strict NPIs further. He observes that strict NPIs are blocked with factive predicates and, more generally, whenever the truth of the embedded clause can be inferred. This excludes strict NPIs under negated factive *know*, but also under matix expressions such as *I wouldn't have thought*, both of which imply the truth of their complement clause.

While Horn (2014) and Hoeksema (2017) show that Neg Raising is not a necessary condition for long-distance licensing of strict NPIs, the availability of a Neg-Raising inference is not a sufficient condition either. Zeijlstra (2017) argues that *be of the opinion* has a Neg-Raising inference, see (11). We find non-strict NPIs in the embedded clause, as in (12a), but no strict NPIs, see (12b).

(11) I am **not** of the opinion [that you are right].
= I am of the opinion [that you are **not** right]. (Zeijlstra 2017)

(12) I am **not** of the opinion . . .
a. Non-strict: [that it would ⌜ever⌝ be used . . .] (English Trends)
b. Strict: *[that Carolyn will ⌞**breathe a word**⌟ about it.] (Zeijlstra 2017)

Other predicates of this type are *it is not the case that* and *it is not true that*, which were already mentioned in Horn (1978) as involving a Neg-Raising inference but as not licensing strict NPIs, see (13).

(13) * It is **not** true/the case [that he'll get there ⌜**until**⌝ Sunday]. (Horn 1978: 207)

So far, we saw that negated matrix predicates allow for strict NPIs unless they are veridical or of the *be of the opinion/be true* type. In either case, non-strict NPIs may occur. Hoeksema (2017) mentions a class of matrix predicates which preclude both strict and non-strict NPIs: speech report predicates.

In his corpus research, Hoeksema (2017) only finds data of NPIs under verbs of saying when these are used in a non-speech report way, as in (14).

(14) a. Strict: I would**n't** say that it was ⌜**at all**⌝ likely.
b. Non-strict: I'm **no**t saying [there is ⌜anything⌝ the matter with him] (Hoeksema 2017)

But Hoeksema (2017) did not find NPIs in speech report uses as in (15)

(15) *I'm **no**t reporting/ *Alex did**n't** say ... [that there is ⌜anything⌝ the matter with him].

We can conclude that there is no correlation between the availability of a Neg-Raising inference and the occurrence of stict NPIs: we find strict NPIs with non-Neg-Raising predicates that are non-veridicial, but not with predicates with Neg-Raising inference such as *be of the opinion that* and *be the case that*.

2.3 Ad G-3: The NPI status of Negative Inversion and Horn clauses

Negative Inversion (NI) occurs primarily with a negative fronted constituent, but also with *only* and other known licensers of NPIs (Büring 2004), see (16b), including weak licensers such as *not every*, (16c).

(16) a. NI: **Not a single** word did he utter unnecessarily. (English Trends)
Weak NPI: **not a single** bullet ⌜**need**⌝ be fired (English Trends)
b. NI: **Only** two of them did he find useful. (Büring 2004)
Weak NPI: **Only** one application form ⌜**need**⌝ be filled out. (www)[3]
c. NI: **Not every** time did they hit a winning note, (English Trends)
Weak NPI: **Not every** lionfish ⌜**need**⌝ be removed. (www)[4]

It is difficult to assess NI with respect to the at-issueness of licensing, as the contexts used in Sedivy (1990), such as stressed auxiliary in (4), cannot be applied to NI. For simplicity, I assume that NI requires an at-issue licenser.

When there is a matrix negation, NI is classified as a Horn clause (HC). Hoeksema (2017) and Horn (2014) show that strict NPIs and HCs can occur not just in Neg-Raising constellations but also with other non-veridical matrix predicates, as with non-fractive *know* in (17b). However, neither strict NPIs nor Horn clauses occur with factive predicates and speech reports, see (17c). Horn clauses are also excluded under *be of the opinion*, (17d). Thus, NI and HC together (NI-HC hereafter) behave like a weak, strict, possibly regular NPI.

[3] `https://tinyurl.com/only-need`, accessed 1.9.2024
[4] `https://tinyurl.com/not-every-need`, accessed 1.9.2024

(17) a. I do**n't** think [that ever before have the media played such a major role in a kidnapping.]. (Horn 2014)

b. I do**n't** know [that ever before had all three boys slept simultaneously]. (Horn 2014)

c. * I did**n't** realize/report [that ever before had all three boys slept simultaneously].

d. * I am **not** of the opinion [that ever before have the media played such a major role in a kidnapping].

Nonetheless, there is an obvious difference between NI-HC and (ordinary) weak strict NPIs: We don't find NI when the NPI-licenser follows the inverted auxiliary, which is the prime licensing context for strict NPIs:

(18) a. NI-HC: *After a party could I **not**/**never** sleep.
vs. **Never** could I sleep after a party.

b. Strict NPI: I do**n't**/**never** ⌞**give a damn**⌟ about sleep after a party.

I characterize the NPI-hood of NI-HC with the constraint in (19):

(19) The NI-HC/NPI Constraint:
The scope of the fronted constituent in NI-HC is a strict weak NPI.

In other words: Whatever is in the scope of the fronted constituent must be (at the same time) licensed like a weak strict regular NPI.

It has been observed in the literature that the fronted constituent in NI must have wide scope within its clause. For example, a fronted negative constituent cannot express constituent negation, but marks the entire clause as negated. This does not mean that the negation must have widest scope, as modals can take scope over it (Francis 2017), see (20). The correct generalization seems to be that the fronted constituent takes scope over other quantifiers in the clause, and over the main lexical verb, but does not necessarily take widest scope.

(20) [Context: *You are teaching a class. The university is concerned that too many students have been failing in recent years, so they tell all instructors to limit the number of Fs they give out.*]
To (very) few students must you give an F. (**MUST-deontic** > **FEW**)
(Francis 2017: 216)

This captures the data: **First**, if the fronted constituent is an NPI licenser itself, the NI-HC/NPI-Constraint holds trivially, as the fronted constituent is constructionally required to take scope over (major part) of the rest of the clause. **Second**, we exclude an NPI within the fronted constituent in unembedded "negative" inversion – whether or not the clause contains a negation, see (21): Since the fronted constituent in NI takes (relatively) wide scope, a fronted NPI cannot be in the scope of a licenser later in the clause.

(21) *Ever have I (not) been to Olomouc before.

Third, HCs may only have elements in the fronted position that do not intervene with the licensing. This excludes a universal quantifier or a positive polarity item, see (22).

(22) I don't think that [in a single year]/*[every year]/*[in some years] has Alex finished a paper.

Fourth, definites are excluded in the fronted position in HCs, (23). This is surprising as definites do not block NPI licensing, contrary to universals, (24).

(23) I don't think that *[this year] has Alex finished a paper.

(24) Alex did **not** give the/*every apple to ⌜any⌝ of the kids.

The ungrammaticality of fronted definites in HCs follows from the requirement that the fronted constituent must take scope over the rest of the clause. Definites are not scopal in the relevant sense – which is why they usually do not count as interveners in NPI licensing in the first place.

The NI-HC/NPI-Constraint captures the distribution of ⌜**NI-HC**⌝. Classifying it as a strict NPI accounts for its occurrence under negated matrix predicates.

3 Semantic analysis

With the refined emprical observations on the three licensing aspects of NPIs, I can now turn to the semantic analysis. I will mainly rely on previous work for the analysis of strength and at-issueness, in particular Sailer (2021, 2022). The main contribution of this paper lies in the account of locality.

3.1 Strength

Zwarts (1981, 1986) provides a basic characterization of licenser strength based on semantic entailment properties. His categories are derived from de Morgan's laws ($\neg(\phi \wedge \psi) \equiv \neg\phi \vee \neg\psi$, and $\neg(\phi \vee \psi) \equiv \neg\phi \wedge \neg\psi$). Zwart's categories have been shown to be empirically useful for classifying NPIs (van der Wouden 1997), and have been further extended, for example in Giannakidou (1998). Sailer & Richter (2002) propose a representational reformulation of Zwart's categories, given in Table 2. They decompose downward-entailing expressions into combinations of a negation and another operator.

I express the weak/strong distinction as follows: Strong NPIs must be in the scope of negation with possibly additional intermediate existential quantifiers. Weak NPIs must be in the scope of negation, but there can be other intermediate quantifiers. This captures the contrast in (3): The strong NPI ⌞***lift a finger***⌟ is licensed by *no one* (decomposed into $\neg\exists x(\ldots\textbf{NPI}\ldots)$), but not by *not everyone* (i.e. $\neg\forall x(\ldots\textbf{NPI}\ldots)$). The weak NPI ⌜*ever*⌝ is fine in both licensing environments.

strength	example	entailment	representation
super-strong	*not*	antimorphic	$\neg(\ldots\textbf{NPI}\ldots)$
strong	*nobody*	anti-additive	$\neg\exists x(\ldots\textbf{NPI}\ldots)$
weak	*few*	downward entailing	$\neg\textbf{Many}x(\ldots\textbf{NPI}\ldots)$

Table 2: Representational encoding of entailment categories

3.2 At-issueness

The at-issueness aspect of NPI licensing requires an inclusion of various types of pragmatic inferences. Sailer (2021) presents an HPSG-compatible model of the semantics-pragmatics interface, which is heavily based on Levinson (2000), but includes Conventional Implicatures. In Figure 1, I repeat his model, indicating where regular and lexical NPIs are licensed. On the right side I also mention where the different types of semantic representation occur inside the HPSG encoding that I will use in Section 4.

In this model, there is a *primary truth-conditional content* that comprises ordinary, combinatorial semantics, including linking, scope and anaphora resolution, and accommodation of presuppositions. The *conventional content*, then, includes conventional implicatures and other use-conditional content (Potts 2007a). As proposed in Levinson (2000), the *utterance content* is arrived at by further enrichment through generalized conversational implicatures (including scalar implicatures). Particularized conversational implicatures à la Grice (1975) are not considered part of the linguistic representation in this model.

Sailer (2021, 2022) uses this model to capture the distributions of what I refer to as regular and lexical NPIs in the present paper. He proposes that a stressed auxiliary as in (4) contributes a use-conditional semantics of rejecting a salient proposition. This salient proposition may contain a negation. Regular NPIs such as ⌜*ever*⌝ need to be licensed within the primary content, i.e., they require an overt licenser. Lexical NPIs such as ⌞***lift a finger***⌟ are fine if licensed in the utterance content. Consequently, they can be licensed by a negation that is only part of the use-conditionally enriched semantic representation. I sketch the relevant semantic representations in (25) to illustrate how this accounts for the contrast in (4).

(25) a. Regular NPI: *Bert DOES care about the homeless ⌜at all⌝.
(reject: 'Bert does**n't** care about the homeless ⌜at all⌝.')

b. Lexical NPI: Cynthia DOES ⌞**lifted a finger**⌟ when …
(reject: 'C. does**n't** ⌞**lift a finger**⌟ under any circumstances.')

Primary content: …**NPI**…
Utterance content: …**NPI**…∧**reject**(**speaker**,…¬(…**NPI**…)…)

The part of the utterance content that is derived through enrichment (i.e., Conventional Implicatures, use-conditional meaning, or generalized conversa-

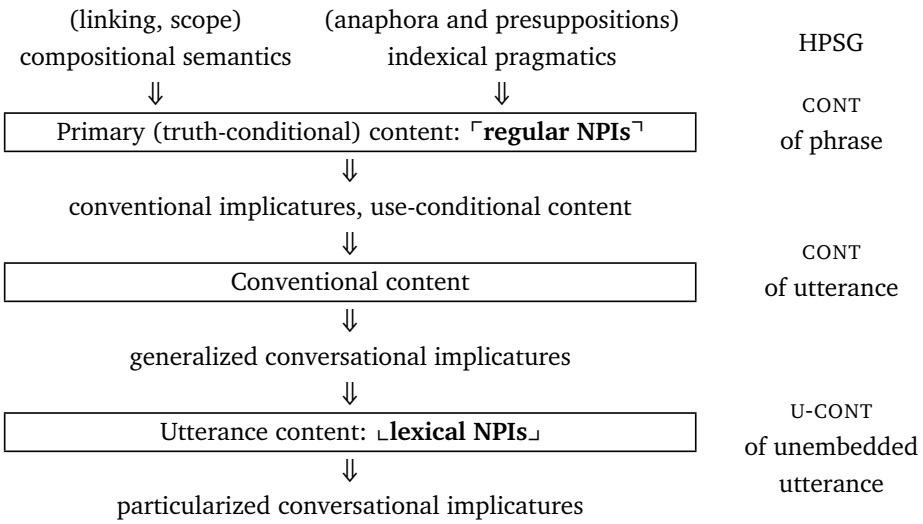


Figure 1: Semantics-pragmatics interface and NPI licensing in HPSG

tional implicatures) is underlined. Neither NPI is licensed in the primary content of the relevant example, but in both cases, an occurrence of the NPI is in the scope of negation in the utterance content. The regular NPI in (25a) is not acceptable as it is unlicensed in the primary content. The utterance content is the relevant licensing domain for lexical NPIs, therefore (25b) is fine.

3.3 Locality

We can now turn to the strict/non-strict distinction. As noted above, in the relevant examples of embedded NPI licensing beyond Neg Raising, the negation is interpreted in the matrix clause, not in the embedded clause. This is also the case in (17b). Horn (2014) explains the occurrence of HCs in non-Neg-Raising contexts as follows: In examples like (17b), non-factive *know* has a relevant weaker alternative – for example *think* – that triggers a Neg-Raising inference. I.e., the example is fine because *I don't think that ever before* . . . is.

I will propose a simpler theory, in which strict NPIs and HCs are directly licensed by a matrix negation in the complement of a Neg-Raising or a non-veridical predicate. This is particularly plausible as, semantically, these contexts pattern with other licensing cases.

Universally quantified noun phrases block NPI licensing, even for non-strict NPIs like ⌜*any*⌝, as in (24) above. However, modal and opaque predicates do not block NPI licensing, even those that are considered universal quantifiers over a set of possible worlds $\mathscr{W}$. In (26), the strict NPI ⌞***lift a finger***⌟ is licensed by the negation even with a universal modal taking intermediate scope.

(26) He won't have to ⌞**lift a finger**⌟ to win that primary. (English Trends)
$\neg\forall w \in \mathscr{W}(\ldots \textbf{NPI} \ldots)$

Just like modals, Neg-Raising predicates and other matrix predicates are typically analyzed as quantifiers over possible worlds. Consequently, it is to be expected that they do not constitute interveners for NPI licensing either.

We can apply this basic insight to Neg Raising *think* and non-factive *know*. We treat both predicates as universal quantifiers over the worlds compatible with what the subject believes or is certain about. As shown in (27), the strict NPI ⌜***all that***⌝ is in the scope of a negation with just the quantification over some relevant set of worlds taking scope between the negation and the NPI.

(27) I don't think/know [that this is ⌜**all that**⌝ complicated].
$\neg\forall w \in \mathscr{W}_{\textbf{speaker}}.(\ldots \textbf{NPI} \ldots))$

While this captures the matrix scope reading for non-Neg-Raising predicates, it also derives just a matrix-scope reading for Neg Raising. To get the low interpretation, we can apply strengthening or similar strategies, as suggested in semantic and pragmatic accounts of Neg Raising, such as Gajewski (2007) or Romoli (2013), to name just two. The same analytic strategy has been proposed in Zeijlstra (2017). In other words, the current proposal is fully compatible with semantic and pragmatic accounts of Neg Raising, but attributes the NPI licensing to an independent property that Neg-Raising predicates share with other matrix predicates. This analysis captures the licensing of all discussed NPI types under matrix negation. However, we now have shifted the analytic burden from why strict NPIs are licensed in Neg Raising to why they are not licensed under other types of matrix predicates.

Blocking case 1: Veridical inference blocks strict NPIs and HCs If we allow a matrix negation to license embedded strict NPIs, we seem to have lost control over the restricted distribution of these NPIs. Strict NPIs and HCs are not licensed if the veridicality of the complement clause is inferred or assumed. Kastner (2015) suggests that the complement of factive predicates is a definite noun phrase (and behaves semantically as one). This looks promising at first, as NPIs are not licensed inside a definite noun phrase. However, the analogy breaks down because non-strict NPIs like ⌜*ever*⌝ are excluded from licensing inside a definite noun phrase, see (28a), but can occur in the complement of negated factive predicates, see (28b).

(28) a. They didn't write [a/*the book [that could ⌜ever⌝ be published]].
b. They didn't realize [that the book could ⌜ever⌝ be published].

Montero & Romero (2023) explore whether mood choice in the complement clause of negated matrix predicates in Spanish influences factivity. They derive veridical inferences as scalar implicatures triggered by exhaustification, following Romoli (2015): As a factive predicate, **know** has the scalar alternatives **know**(x, p) and p – with p being the weaker alternative. Under negation,

$\neg$**know**(x,p) is exhaustified to $\neg$**know**$(x,p) \wedge p$, i.e., when negating a strong element on the scale, the weaker scalar alternative is still assumed to be true.

In the analysis of lexical NPIs in Section 3.2 we exploited non-at-issue inferences to license lexical NPIs. Homer (2008) has shown that non-at-issue semantics can also block NPI licensing.[5] If the factive/veridical inference is a scalar implicature, it will be part of the utterance content in the model in Figure 1. This gives us a natural way to express licensing constraints that are sensitive to the factivity/veridicality of the context. For strict NPIs we need to assume that pragmatic enrichment must not introduce a constellation that excludes the NPI – as would be the case for a strict NPI inside p in $\neg$**know**(x,p) after exhaustification to $\neg$**know**$(x,p) \wedge p$.

Blocking case 2: *be true/be of the opinion* As discussed in Section 3.3, predicates like *be the case*, *be true*, *be of the opinion* do not allow for licensing of embedded strict NPIs – even though they may trigger a Neg-Raising inference.

Typical negated uses of *be of the opinion* are as in (29). Here, the question of whether everyone should be tested is presented as being of conversational interest, and the speaker expresses a negative opinion on it.

(29) In a recent conversation, she asked me, "Why isn't everyone just tested before they get pregnant?" Good question.
To be clear, I am not of the opinion that everyone should "just get tested."
(English Trends)

This suggests that the proposition expressed in the embedded clause is contextually given in some sense. I will tentatively assume that it is part of Portner's *common propositional space*, i.e. a member of the set of propositions that are of interest for the current conversation.[6]

This leads to the semantic representations in (30), where I use the constant **CPS** for the set of propositions in the common propositional space. The primary content contains a negation. The utterance content has two more elements: (i) a Neg Raising inference, and (ii) the background assumption that the proposition expressed in the embedded clause is in the common propositional space.

(30) Kim is not of the opinion ...

a. [that Alex will ⌜ever⌝ talk about it].

$\neg\forall w.w \in \mathscr{W}_{\textbf{alex}}.(\ldots \text{NPI} \ldots)$ (primary content)

$\underline{\wedge\forall w.w \in \mathscr{W}_{\textbf{alex}}.\neg(\ldots \text{NPI} \ldots)}$ (Neg-Raising inference)

$\underline{\wedge\lambda w.(\ldots \text{NPI} \ldots) \in \textbf{CPS}}$ (CPS inference)

[5] Though Homer (2008) discusses blocking of regular NPIs through presuppositions, which the model in Figure 1 treats as part of the primary content.

[6] This is inspired by Montero et al. (2024): They argue that propositions in the common propositional space can block the licensing of strict NPIs in Spanish embedded indicative clauses.

b. *[that Alex will ⌞**breathe a word**⌟ about it].

$\neg\forall w.w \in \mathscr{W}_{\textbf{alex}}.(\ldots \textbf{NPI} \ldots)$ (primary content)

$\underline{\wedge\forall w.w \in \mathscr{W}_{\textbf{alex}}.\neg(\ldots \textbf{NPI} \ldots)}$ (Neg-Raising inference)

$\underline{\wedge\lambda w.(\ldots {}^{*}\textbf{NPI} \ldots) \in \textbf{CPS}}$ (CPS inference)

The non-strict NPI in (30a) is licensed by the negation in the primary content. The utterance content does not play a role for it. The strict NPI in (30b) would be licensed in the primary content and the Neg Raising inference. However, it does not tolerate a non-licensed occurrence in the last conjunct, which is related to the common propositional space.

Blocking case 3: Speech reports block HCs and all NPIs Neither HCs nor any other NPIs can occur in negated speech reports. Hoeksema (2017) argues that the reported utterance itself is not at issue, whereas an NPI must be part of the at-issue content in the constellation in which it is licensed. Montero & Romero (2023) treat speech report on a par with attitude predicates with the only difference being in the modal base (the *reported background* for verbs of saying, following Portner & Rubinstein 2020), but the content of the speech report is not an alternative. Thus, if we adopted Montero & Romero's analysis, we would be in the same situation as for non-factive non-speech-report predicates and wrongly predict NPI licensing and HCs inside speech reports.

Therefore, I will go in a different direction. I propose that the content of the reported speech is simply not in the scope of the matrix negation. I sketch the resulting semantic representation in (31), where I assume that the variable x is of the semantic type u (for "utterance", see Potts 2007b), and the predicate **Content** holds between an utterance x and a proposition p iff p is a propositional content entailed by x. This way, the semantic representation of the embedded clause, p, is not part of the scope of negation.

(31) * Alex didn't say [that Kim had ⌞ever⌟ submited a paper].

$\exists x_u : \textbf{Content}(x, (\ldots \textbf{NPI} \ldots) \wedge \neg \textbf{report}(\textbf{alex}, x)$

Of course, negated reported speech rarely occurs in natural communication. Uses of negated speech report predicates such as those in (14) are far more common and behave with respect to NPI licensing just as other negated matrix predicates discussed in this subsection.

Intermediate summary I have argued that in NI-HC, the fronted constituent has scope over a substantial part of the rest of the clause and that what is in its scope behaves like a strict weak NPI (the NI-HC/NPI-Constraint). In addition, I have amended the NPI licensing theory of Sailer (2021) by adding a licensing condition for strict NPIs: every occurrence of a strict NPI beyond the primary content must be licensed. I propose that since quantification over possible worlds

	utterance content	non-strict	**strict**
Neg Raising	$(\neg\forall w\phi)$	ok	ok
non-factive *know*	$(\neg\forall w\phi)$	ok	ok
factive *know*	$(\neg\forall w\phi)\underline{\wedge\phi}$	ok	*
be of the opinion	$(\neg\forall w\phi)\underline{\wedge\lambda w.\phi \in \mathbf{CPS}}$	ok	*
speech report	$\exists y_u(\mathbf{Cont}(y,\phi)\wedge\neg\mathbf{say}(x,y)$	*	*

Table 3: Types of negated matrix predicates and licensing of an NPI inside ϕ

does not block NPI licensing, licensing from a matrix negation should be possible in general. However, the enrichment through a veridical inference or contextual giveness blocks the licensing of strict NPIs. For the semantics of speech reports I tentatively suggested that the content of the reported utterance is not in the scope of negation at all, which blocks all NPIs from being licensed through a matrix negation. The types of matrix predicates discussed in this paper and the relevant semantic representations are summarized in Table 3. In the following, I extend existing HPSG analyses to formalize this approach.

4 HPSG analysis of NPI licensing

4.1 Previous approaches

Neg Raising Sailer (2006) analyzes Neg Raising as an instance of scope ambiguity, i.e., the negation from the matrix clause can either take scope over the matrix predicate or be in its scope, see (32).

(32) I don't think [that Alex won].

Surface scope: $\neg\forall w(w \in \mathscr{W}_{\mathbf{speaker}} : \mathbf{win}_w(\mathbf{alex}))$

Neg-Raising: $\forall w(w \in \mathscr{W}_{\mathbf{speaker}} : \neg\mathbf{win}_w(\mathbf{alex}))$

This analysis was based on the assumption that Neg Raising is the only possibility for non-local licensing of strict NPIs. We saw, however, that with **G-2**, this analysis is no longer tenable. Instead, a simpler analysis turned out to be more adequate in which only the surface scope representation is needed but the universal quantification over possible world does not block NPI licensing.

NPI licensing Richter & Soehn (2006) propose a collocational treatment of NPIs. They argue that an NPI must be licensed within the semantic representation of a particular syntactic domain containing it. Both, the kind of licenser and the size of the domain can vary. Richter & Soehn use a list-valued feature CONTEXT-OF-LEXICAL-LICENSING (COLL) to express distributional idiosyncrasies of lexical items. The elements on the COLL list specify the syntactic domain

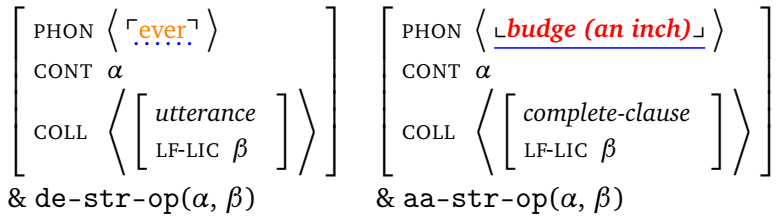


Figure 2: Licensing requirements adapted from Richter & Soehn (2006)

of the licensing by their type – for example *utterance* for licensing within the complete utterance. In addition, there are features for whether this is a semantic, syntactic, or phonological licensing condition. NPIs require semantic licensing, which is expressed by a feature LF-LIC.[7] The LF-LIC value is identical with the CONTENT value of the licensing syntactic object. The concrete licensing strength is encoded as relational constraints. For example, if α is the relevant part of the semantic representation that needs to be licensed by a strong (anti-additive) licenser within the CONTENT of the licensing domain, β, we add a constraint `aa-str-op`(α, β).[8]

I sketch the lexical entries of the NPIs ⌜*ever*⌝ and ⌞***budge (an inch)***⌟ according to Richter & Soehn (2006) in Figure 2. As a non-strict weak NPI, ⌜*ever*⌝ requires a licenser within the CONTENT of the overall utterance that is at least downward entailling. The strict strong NPI ⌞***budge (an inch)***⌟ asks for an anti-additive licenser within the CONTENT value of the smallest clause containing it. Note that the syntactic restriction to a clause-mate licensing for strict NPIs indicates that Richter & Soehn (2006) don't assume **G-2**, i.e., they assume that strict NPIs can only be licensed by a morpho-syntactically higher negation if that negation is interpreted within the clause that contains the NPI.

4.2 Revised analysis of NPI licensing

The aim of the present paper is to modify Richter & Soehn's collocational NPI-licensing theory so that it can express the licensing theory of Section 3. Since this theory relies heavily on pragmatic notions, the notion of an *utterance* is important. I will adopt the architecture of Höhle (2019: 583), who discusses phonological processes at the level of embedded an unembedded signs. Höhle assumes that unembedded signs are of type *unembedded-phrase*. Unembedded signs have both, a traditional PHON feature, and a feature UTTERANCE-PHON. The value of the latter is the result of applying post-lexical phonological processes to the utterance's PHON value.

[7] I use lower-case Greek letters instead of boxed integers as tags for semantic representations.

[8] This is a free adaptation of Richter & Soehn (2006): First, Richter & Soehn assume the semantic combinatorics of *Lexical Resource Semantics* (Richter & Sailer 2004). I am using a more traditional semantic architecture in this paper for better accessibility, though an LRS encoding would be straightforward. Second, they formulate functional rather than relational constraints.

I model the semantics-pragmatics interface from Figure 1 in analogy to Höhle's treatment of phonology: The *primary content* of a sign is its CONTENT value. The *conventional content* is the CONTENT value of an unembedded sign.[9] I introduce a new feature, UTTERANCE-CONTENT (U-CONT) on unembedded signs, whose value is the sign's *utterance content*. The U-CONT is the result of applying semantic/pragmatic enrichment (such as generalized conversational implicatures) to the utterance's CONTENT value. In other words, I assume that generalized conversational implicatures are to meaning what post-lexical phonological processes are to phonology.

I argued in Section 3 that NPI licensing can be checked at the primary content (for non-strict regular NPIs) and at the utterance content (for strict and lexical NPIs) – since the latter are sensitive to conventional and generalized conversational implicatures. Contrary to Richter & Soehn (2006), I do not distinguish between licensing at the clausal level and licensing at the utterance level because at least the NPI types discussed here can all be licensed non-locally. Consequently, I will assume that all NPI-related elements on the COLL list are of type *utterance*. Such *utterance* objects have an additional feature ULF-LIC, whose value is identical with the U-CONT value of the utterance containing the NPI.

I can now provide the HPSG encoding of my NPI-licensing theory. My specification of a weak strict regular NPI such as ⌜*ever*⌝ is just as Richter & Soehn's in Figure 2: The NPI asks for a downward-entailing licenser in the CONTENT value of the utterance containing them.

A lexical NPI need not be licensed in the primary content, but can be licensed anywhere within the enriched semantic representation, i.e., within the U-CONT value of the utterance. The difference between a regular and a lexical NPI is, therefore, just that the lexical NPI looks for its licenser in the U-CONT value.

In order to include the strict/non-strict aspect of NPI licensing, I unpack the notation in Richter & Soehn (2006). Richter & Soehn require an occurrence of the NPI's semantics in the scope of an appropriate licenser, i.e., there is an *existential quantification* over the occurrences of the NPI semantics. This suffices to capture non-strict NPIs. For **strict** NPIs, however, we need a *universal quantification* over the occurrences of the NPI semantics: every occurrence of the NPI semantics in the utterance content (beyond the primary content) must be in the scope of an appropriate licenser. This is implemented in Figure 3. A weak regular strict NPI like ⌜***need***⌝ requires a downward entailing licenser in the primary content (β), by virtue of being a weak regular NPI (the first condition next to the AVM). By virtue of being a strict NPI, any additional occurrence in the utterance content (γ) must also be licensed (the second condition).[10]

[9]The conventional content is the result of enriching the primary content with Conventional Implicatures. While Potts (2005) assumes that this only happens at the utterance level, Bach (1999) shows that many of these inferences can be integrated at embedded signs, as long as these have some kind of speech act operator. I will ignore this complication, but see Asadpour et al. (2022) for the relevance of this distinction also for conventionalized social meaning inferences.

[10]For the time being, I simply postulate that the universal non-at-issue licensing underlies the

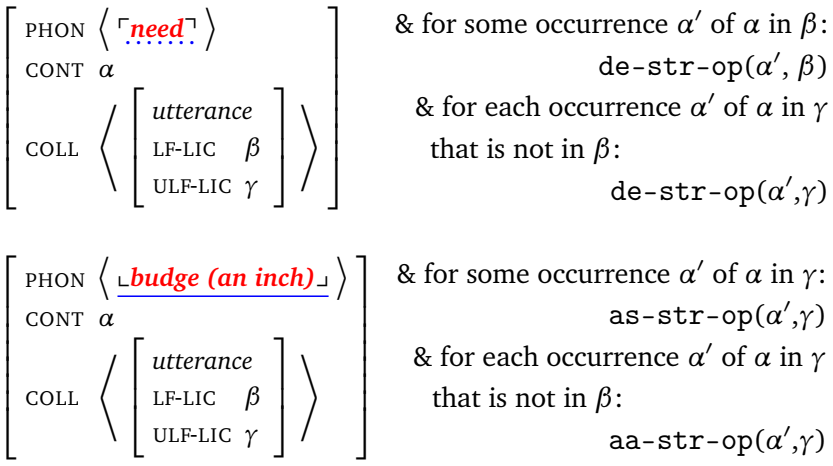


Figure 3: Sketches of lexical entries of example NPIs with locality

I provide the revised lexical entry of ⌞*budge (an inch)*⌟ – a strong lexical strict NPI – below that of ⌜*need*⌝. This NPI asks for an anti-additive licenser in the utterance content γ, instead of the primary content. In addition, since it is a strict NPIs, all occurrences outside the primary content must be licensed.

4.3 HPSG analyses of Negative Inversion/Horn Clauses

My final task is to show how ⌜NI-HC⌝ can be modelled. I will start from the syntactic analysis of NI in Maekawa (2012) and add to it NPI-licensing conditions analogous to those of ⌜*need*⌝.

Maekawa (2012) adopts the classical analysis of NI from Emonds (1970) as a flat structure in which a clause-initial constituent is followed by an inverted auxiliary, its subject, and a VP. The fronted constituent is related to the VP by an unbounded dependency.

Maekawa (2012) restricts the fronted constituent to be marked as negative, which means that it must contain some element that is able to express clausal negation, such as *never*, or *not a single person*. Since we also find inversion with other NPI licensers, see (16), negative inversion cannot be related to morpho-syntactic negativity of the fronted constituent but rather to some semantic properties. In HC, the fronted element is never an inherently negative expression.

To remedy this limitation of Maekawa's analysis, it is sufficient to treat NI and HC as instances of the same construction and to assume that the scope of the fronted constituent is a weak regular strict NPI. Consequently, in NI, the fronted constituent must be a (weak) NPI licenser and, for HCs, the fronted constituent must not block the licensing from the matrix clause. The resulting analysis is summarized in Figure 4, using an oversimplified feature geometry, just to sketch the essence of the analysis.

same strength requirements as the existential licensing.

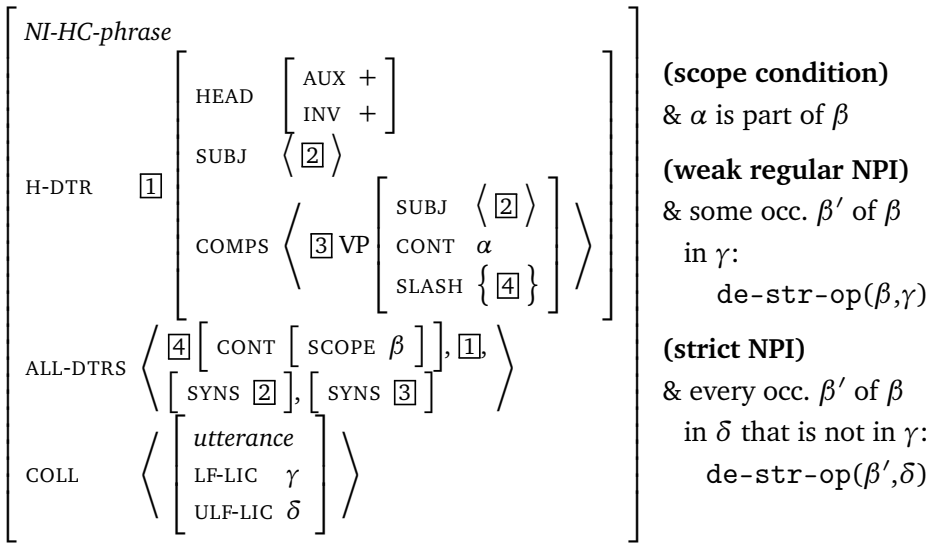


Figure 4: Negative Inversion/Horn Clauses as weak regular strict NPI

I adopt Maekawa's flat syntactic structure. It is a headed phrase, headed by an inverted auxiliary, 1. The head daughter selects for a subject, 2 and a VP complement, 3. The VP has an element, 4, in its SLASH value. This extracted element is realized as the first daughter in the structure (see the ALL-DTRS value), followed by the head daughter, the subject, and the complement VP.

I add the special semantic properties of NI-HC to this basic syntactic structure. First, the fronted element is semantically scopal, indicated by the feature SCOPE. The first condition next to the AVM specifies that the CONTENT value of the head daughter (α) must be in the scope of the fronted constituent (β). According to the NI-HC/NPI-Constraint the scope of the fronted constituent behaves like a weak regular strict NPI. Consequently, the NPI-licensing constraints are stated for β. First, as a regular NPI, β must have a licensed occurrence within the primary content of the utterance containing the *NI-HC-phrase*. Second, being a weak NPI, the licenser need only be downward-entailing. These two licensing aspects are expressed in the second constraint next to the AVM. Third, since β is a strict NPI, any occurrence of it within the utterance content (δ) must be licensed as well. Consequently, the constraint on the sort *NI-HC-phrase* in Figure 4 illustrates nicely how NPI-licensing requirements can be added to individual constructions in a straightforward way.

5 Conclusion

The present paper made four main contributions: First, it argued for a surface-scope oriented approach to phenomena that have been considered strong evidence against surface scope of negation: the licensing of embedded strict NPIs

and Horn clauses. Second, this surface-scope analysis was expressed within existing HPSG analyses of NPI licensing and Negative Inversion. Third, the strict/non-strict distinction of NPIs is interpreted as universal vs. existential quantification over the licensing requirement within an enriched semantic representation. Fourth, Negative Inversion/Horn clauses are analyzed as a constructional NPI. The existence of constructional NPIs should not be surprising, but no such example has been previously discussed in HPSG to my knowledge.

The proposed theory of NPI licensing is a synthesis and further elaboration of previous representational (HPSG) accounts. I used three empirical aspects of NPI-licensing to develop this theory systematically: Strength distinctions are expressed through different licenser requirements in terms of relational constraints (Richter & Soehn 2006). At-issueness follows from whether an NPI needs to be licensed in the primary content or the utterance content of the utterance containing it, i.e., in the utterance's CONT or U-CONT value respectively (Sailer 2021). Finally, locality is a matter of quantification: A non-strict NPI needs just a single, licensed occurrence, a strict NPI requires that all its occurrences (outside the primary content) be licensed.

References

Asadpour, Hiwa, Shene Hassan & Manfred Sailer. 2022. Non-*wh* relatives in English and Kurdish: Constraints on grammar and use. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 29th International Conference on Head-Driven Phrase Structure Grammar, Online (Nagoya/Tokyo)*, 6–26. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2022.1.

Bach, Kent. 1999. The myth of conventional implicature. *Linguistics and Philosophy* 22(4). 327–366. DOI: 10.1023/A:1005466020243.

von Bergen, Anke & Klaus von Bergen. 1993. *Negative Polarität im Englischen*. Tübingen: Narr.

Büring, Daniel. 2004. Negative inversion. In Leah Bateman & Cherlon Ussery (eds.), *Proceedings of NELS 35*, 1–21.

Emonds, Joseph E. 1970. *Root and structure-preserving transformations*. MIT. (Doctoral dissertation).

Francis, Naomi. 2017. Modal scope in negative inversion constructions. In Aaron Kaplan, Abby Kaplan, Miranda K. McCarvel & Edward J. Rubin (eds.), *Proceedings of the 34th West Coast Conference on Formal Linguistics*, 214–221. Somerville, MA: Cascadilla Proceedings Project.

Gajewski, Jon Robert. 2007. Neg raising and polarity. *Linguistics and Philosophy* 30. 289–328. DOI: 10.1007/s10988-007-9020-z.

Giannakidou, Anastasia. 1998. *Polarity sensitivity as nonveridical dependency*. Amsterdam: John Benjamins.

Grice, Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics*, vol. 3, 41–58. New York: Academic Press.

Hoeksema, Jack. 2017. Neg-Raising and long-distance licensing of Negative Polarity Items. In Debra Ziegeler & Zhiming Bao (eds.), *Negation and contact: With special focus on Singapore*, 30–62. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Hoeksema, Jack. 2024. *De negatiief-polaire uitdrukkingen van het Nederlands*. Groningen: University of Groningen Press. DOI: 10.21827/656dea670d2ba.

Höhle, Tilman N. 2019. An architecture for phonology. In Stefan Müller, Marga Reis & Frank Richter (eds.), *Beiträge zur deutschen Grammatik. Gesammelte Schriften von Tilman N. Höhle*, 2nd edn., 571–607. Originally published 1998. Berlin: Language Science Press. DOI: 10.5281/zenodo.2588381.

Homer, Vincent. 2008. Disruption of NPI licensing: The case of presuppositions. In Tova Friedman & Satoshi Ito (eds.), *SALT 18*, 429–446. Ithaca: Cornell University. DOI: 10.3765/salt.v18i0.2483.

Horn, Laurence R. 1978. Remarks on neg-raising. In Peter Cole (ed.), *Pragmatics*, vol. 9 (Syntax and Semantics), 129–220. New York: Academic Press.

Horn, Laurence R. 2014. The cloud of unknowing. In Jack Hoeksema & Dicky Gilbers (eds.), *Black book*, 178–196. Groningen: University of Groningen.

Kastner, Itamar. 2015. Factivity mirrors interpretation: The selectional requirements of presuppositional verbs. *Lingua* 164. 156–188. DOI: 10.1016/j.lingua.2015.06.004.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1. 7–36. DOI: 10.1007/s40607-014-0009-9. http://www.sketchengine.eu.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. London: MIT Press.

Maekawa, Takafumi. 2012. *An HPSG approach to negative inversion constructions*. https://hokusei.repo.nii.ac.jp/?action=repository_action_common_download&item_id=416&item_no=1&attribute_id=45&file_no=1.

Montero, Raquel, Leah Doroski & Maribel Romero. 2024. *Spanish Neg-raising: Mood effects on NPI licensing*. Handout to a talk at the MECORE Closing Workshop, June 20, 2024.

Montero, Raquel & Maribel Romero. 2023. Examining the meaning of polarity subjunctive. In Suet-Ying Lam & Satoru Ozaki (eds.), *Proceedings of NELS 53*.

Portner, Paul & Aynat Rubinstein. 2020. Desire, belief, and semantic composition: Variation in mood selection with desire predicates. *Natural Language Semantics* 28. 343–393. DOI: 0.1007/s11050-020-09167-7.

Potts, Christopher. 2005. *The logic of conventional implicatures* (Oxford Studies in Theoretical Linguistics 7). Oxford: Oxford University Press.

Potts, Christopher. 2007a. Into the conventional-implicature dimension. *Philosophy Compass* 4(2). 665–679. DOI: 10.1111/j.1747-9991.2007.00089.x.

Potts, Christopher. 2007b. The dimensions of quotation. In Chris Barker & Pauline Jacobson (eds.), *Direct compositionality*, 405–431. Oxford: Oxford University Press.

Richter, Frank & Manfred Sailer. 2004. Basic concepts of Lexical Resource Semantics. In Arne Beckmann & Norbert Preining (eds.), *ESSLLI 2003: Course material I*, vol. 5 (Collegium Logicum), 87–143. Vienna: Kurt Gödel Society.

Richter, Frank & Jan-Philipp Soehn. 2006. *Braucht niemanden zu scheren*: a survey of NPI licensing in German. In Stefan Müller (ed.), *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar, Varna 2006*, 421–440. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2006.23.

Romoli, Jacopo. 2013. A scalar implicature-based approach to neg-raising. *Linguistics and Philosophy* 36. 291–353. DOI: 10.1007/s10988-013-9136-2.

Romoli, Jacopo. 2015. The presuppositions of soft triggers are obligatory scalar implicatures. *Journal of Semantics* 32. 173–219. DOI: 10.1093/jos/fft017.

Sailer, Manfred. 2006. *Don't Believe* in underspecified semantics: Neg Raising in Lexical Resource Semantics. In Olivier Bonami & Patricia Cabredo Hofherr (eds.), *Empirical issues in formal syntax and semantics*, vol. 6, 375–403. Paris: CNRS. www.cssp.cnrs.fr/eiss6/sailer-eiss6.pdf.

Sailer, Manfred. 2021. Minimizer negative polarity items in non-negative contexts. In Stefan Müller & Nurit Melnik (eds.), *Proceedings of the 28th International Conference on Head-Driven Phrase Structure Grammar*, 348–368. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2021.19.

Sailer, Manfred. 2022. Use-conditional licensing of strong Negative Polarity Items. In Daniel Gutzmann & Sophie Repp (eds.), *Proceedings of Sinn und Bedeutung 26*, 734–752. University of Cologne. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/issue/view/30.

Sailer, Manfred & Frank Richter. 2002. Not for love or money: Collocations! In Gerhard Jäger, Paola Monachesi, Gerald Penn & Shuly Wintner (eds.), *Proceedings of Formal Grammar 2002*, 149–160. https://www.english-linguistics.de/fr/cards/fg02sr.html.

Sedivy, Julie. 1990. Against a unified analysis of negative polarity licensing. *Cahiers Lingistiques D'Ottawa* 18. 95–105.

van der Wouden, Ton. 1997. *Negative contexts: Collocation, polarity and multiple negation*. London: Routledge.

Zeijlstra, Hedde. 2017. Does Neg-Raising involve Neg-Raising? *Topoi*. DOI: 10.1007/s11245-017-9461-0.

Zwarts, Frans. 1981. Negatief polaire uitdrukkingen. *GLOT* 4(1). 35–132.

Zwarts, Frans. 1986. *Categoriale grammatica en algebraısche semantiek. een onderzoek naar negatie en polariteit in het Nederlands*. Rijksuniversiteit Groningen. (Doctoral dissertation).

Gradient HPSG

Berke Şenşekerci

University of Warsaw

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 173–192

Keywords: HPSG, gradience, coordination of unlikes, experimental syntax

Şenşekerci, Berke. 2024. Gradient HPSG. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 173–192. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.11. CC-BY 4.0

Abstract

Prevailing grammatical frameworks treat grammaticality as a binary concept, despite strong experimental evidence suggesting it is better understood as a gradient notion. This highlights a serious disconnect between linguistic theory and empirical data. While a few truly gradient frameworks have been proposed to bridge the gap, none have been developed within a constraint-based formalism – an approach particularly well-suited for modeling gradient grammaticality. This work formally introduces a gradient version of HPSG and subsequently employs it to analyze acceptability judgment data on unlike coordination phenomena in Turkish, which display distinctly gradient patterns.

1 Introduction

The notion that grammaticality of sentences cannot be neatly divided into two categories has been recognized since the early days of generative linguistics (Bolinger 1961, Chomsky 1961, Chomsky 1965: 148–153) and has found consistent support in subsequent work involving controlled acceptability judgment experiments (Keller 2000, Keller & Alexopoulou 2001, Featherston 2005b, Sorace & Keller 2005, Haegeman et al. 2014, Hofmeister et al. 2014).[1]

Despite substantial support in favor of gradience, prevailing grammatical frameworks persist in upholding a binary view of grammaticality, forcing linguists to rely on arbitrary generalizations when interpreting acceptability judgment data. As a result, these frameworks allow vastly different grammars to emerge from the same data depending on the chosen cutoff point between grammatical and ungrammatical.

To remedy this problem, various proposals have been put forward. Notable among these are Harmonic Grammar (Legendre et al. 1990), Linear Optimality Theory (Keller 2000), and the Decathlon Model (Featherston 2005a).[2] Interestingly, no such attempt has been made within a fully-fledged

[†]I want to thank Adam Przepiórkowski, whose guidance was invaluable during the conception of this work, and Marcin Opacki for reviewing an earlier version. I also extend my thanks to the audience at HPSG 2024 and to the anonymous reviewers. Any remaining errors are my own.

[1]I tentatively attribute the gradience observed in controlled acceptability judgment experiments to gradience in grammar and not to processing effects, as these experiments significantly minimize and control for such confounds. Accordingly, throughout this paper, I refer to gradient *grammaticality*, not gradient *acceptability*. Nevertheless, it is worth noting that the distinction between performance and grammar effects on acceptability has not been thoroughly explored in experimental settings (but see Hofmeister et al. 2014).

[2]Although Optimality Theory and stochastic variants of existing frameworks may appear to be viable options for modeling gradience, they have notable shortcomings. Optimality Theory not only presupposes binary grammaticality but is also fundamentally incompatible with judgment data (Keller & Asudeh 2002). Similarly, stochastic frameworks are specifically designed to model corpus frequencies, which are a distinct type of data that should not be conflated with acceptability judgments (Pullum & Scholz 2001: 31).

constraint-based framework like Head-Driven Phrase Structure Grammar (Pollard & Sag 1994, Müller et al. 2024), although a constraint-based backbone has been considered to be especially suitable for modeling gradience (Pullum & Scholz 2001: §3.1, Sag & Wasow 2011, Wasow 2024).

In light of this gap, the present work proposes a version of HPSG that accommodates the gradient grammaticality observed in acceptability judgment experiments. Subsequently, the proposed framework is utilized to analyze the results of an acceptability judgment experiment investigating unlike coordination phenomena in Turkish.

2 Gradient HPSG

Compelling evidence suggests that the grammaticality of a sentence is a matter of degree, primarily determined by two distinct factors (Keller 2000, Featherston 2005a, Sorace & Keller 2005): 1) the number of violations present, and 2) the relative severity of the violated constraints. To model gradient grammaticality in terms of these two factors, Gradient HPSG introduces two modifications to the model theory of HPSG.[3]

The first modification updates the original definition of an HPSG grammar (Richter 2004: 178) to allow each grammar constraint to be assigned a numeric weight that reflects the severity of its violation:

Definition 1 (grammar) Γ *is a grammar iff*
Γ *is a pair* $\langle \Sigma, \theta \rangle$,
Σ *is a septuple* $\langle S, \sqsubseteq, S_{max}, A, F, R, Ar \rangle$,
θ *is a set of ordered pairs such that:*
$\theta = \{\langle \delta, w \rangle \mid \delta \in D_0^{\Sigma} \wedge w \in \mathbb{R}^+\}$

The original definition of a *signature*, denoted by Σ (Richter 2004: 156), remains unchanged. This essentially means that Gradient HPSG does not introduce gradience to type hierarchies (cf. Brew 1995). However, θ is no longer a set of constraints as originally defined, but instead a set of ordered pairs where each pair consists of a constraint, δ, and its weight, w, which can only be a positive real number.

The second modification concerns the definition of a *model*, which originally classifies a sentence as a well-formed structure within a *model* of a grammar *iff* the sentence satisfies each constraint of the grammar (Richter 2004: 178–179).

By contrast, Gradient HPSG posits that the modelness (or well-formedness) of a sentence is a real number from 0 to negative infinity, where sentences

[3]Throughout the paper, 'model theory of HPSG' refers to *Relational Speciate Reentrant Logic* (RSRL; Richter 2004).

with 0 modelness value are perfect models of the grammar – i.e., they do not violate any constraint of the grammar.

As per the two factors underlying the grammaticality of a sentence, this value is determined on the basis of constraint weights and the number of constraint violations present in a sentence. The following definition of a *model* assumed in Gradient HPSG formalizes this concept (to be revised):

Definition 2 (model; preliminary) *For each grammar* $\Gamma = \langle \Sigma, \theta \rangle$ *and for each* Σ *interpretation* $\mathsf{I} = \langle \mathsf{U}, \mathsf{S}, \mathsf{A}, \mathsf{R} \rangle$
Modelness degree of I *with respect to* Γ *is:*

$$M(\mathsf{I}) = -\sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathsf{U} \setminus \mathsf{D}_{\mathsf{I}}(\delta_i)| \cdot w_i$$

The mathematical function that determines the modelness degree of a sentence is conceptually equivalent to the harmony function operationalized in Linear Optimality Theory (Keller 2000: 253): it computes the weighted sum of constraint violations for each constraint δ_i in a grammar. However, the function used in this definition is model-theoretic, operating strictly on HPSG structures.

The first term following the negated summation, $|\mathsf{U} \setminus \mathsf{D}_{\mathsf{I}}(\delta_i)|$, returns the number of entities that are *not* denoted by δ_i. In simpler terms, this term counts the number of violations that a sentence makes with respect to δ_i. The number of δ_i violations obtained by this term is subsequently multiplied by the weight assigned to δ_i, w_i. For example, if a sentence violates δ_i twice and the weight of δ_i is specified as 0.45 in the grammar, the sentence receives an evaluation of 0.90 with respect to δ_i (2×0.45).

This evaluation procedure is carried out for each and every constraint in the grammar, with the outcomes of each assessment summed. The resulting sum is then negated to render the modelness value more intuitive, as higher values obtained from the weighted sum indicate greater degrees of ill-formedness rather than well-formedness.

Alas, this definition does not work as intended on standard RSRL assumptions regarding the shape of models: it does not compute the modelness degree of an individual sentence with respect to the weighted grammar. The standard model theory of HPSG posits that models reflect language as a whole (King 1999, Richter 2004, 2007), i.e., that they are *exhaustive models*. This assumption implies that models include all possible sentences within a language, as well as various partial HPSG objects, such as SYNSEM objects. Consequently, the function presented in Definition 2 iterates over all such objects instead of a specific sentence.

To ensure that this function takes an individual sentence as its input, Gradient HPSG additionally incorporates Przepiórkowski's (2021) revisions to the model-theory of HPSG that restrict models to correspond strictly to

individual sentences (i.e., rooted, non-exhaustive models).[4] The final definition is presented below, with *interpretation* I now formally defined as a 5-tuple that includes a root element, denoted as r:

Definition 3 (model) *For each grammar* $\Gamma = \langle \Sigma, \theta \rangle$ *and for each* Σ *interpretation* $\mathsf{I} = \langle \mathsf{U}, \mathsf{r}, \mathsf{S}, \mathsf{A}, \mathsf{R} \rangle$
Modelness degree of I *with respect to* Γ *is:*

$$M(\mathsf{I}) = -\sum_{\langle \delta_i, w_i \rangle \in \theta} |\mathsf{U} \backslash \mathsf{D}_{\mathsf{I}}(\delta_i)| \cdot w_i$$

Having established the formal properties of Gradient HPSG, we can now proceed to illustrate its application in the formal analysis of acceptability judgment data.

3 Experiment

3.1 Background

The morphosyntactic properties of coordinate structures have been the subject of prolonged debate. One widely adopted position contends that conjuncts must bear the same syntactic category (Chomsky 1957: 36, Williams 1981: §2, Bruening & Al Khalaf 2020) and grammatical case (Weisser 2020).

Counter-examples to this position, where conjuncts mismatch either in their category, such as (1a)–(1b), or case, as in (1c), have been explained away by invoking various analytical mechanisms, such as supercategories (Bruening & Al Khalaf 2020), ellipsis (Beavers & Sag 2004: 54–56), and allomorphy (Weisser 2020: §2.3).

(1) a. Pat is [[$_{\text{NP}}$ a Republican] and [$_{\text{ADJP}}$ proud of it]].
(Sag et al. 1985: 117, ex. (2b))

b. We walked [[$_{\text{ADVP}}$ slowly] and [$_{\text{PP}}$ with great care]].
(Sag et al. 1985: 140, ex. (57))

c. This is starting to make [him and I] both feel really bad.
(Parrot 2009: 274, ex. (7a))

This position has recently been challenged based on an abundance of attested examples from Polish and English that defy such analyses (Patejuk 2015, Dalrymple 2017, Przepiórkowski 2022, Patejuk & Przepiórkowski 2023),

[4] The analysis in Section 4 further adopts the second-order extension of HPSG's model theory proposed by Przepiórkowski (2021).

suggesting a potential collapse of this generalization in the face of cross-linguistic evidence.

The current debate, however, is limited to English and Polish data. To further challenge this position through an experimental paradigm, a formal acceptability judgment experiment was conducted to gather data from Turkish, an agglutinative and head-final language.[5]

3.2 Methodology

In the experiment, 48 native speakers of Turkish evaluated the acceptability of sentences on a 7-point Likert scale from -3 (completely unnatural) to 3 (completely natural).[6] The experimental hypothesis posited that conjoining unlike categories and cases is acceptable in Turkish, provided that the conjuncts share the same grammatical function.

The experimental design consisted of two blocks: one for unlike categories and another for unlike cases. The category block had a standard 2×2 design, where the two crossed factors were the category of conjuncts (like or unlike: LCAT vs. UCAT) and the grammatical function of conjuncts (like or unlike: LF vs. UF). For the case block, a similar design was pursued – like or unlike cases (LCASE vs. UCASE) and grammatical functions (LF vs. UF). However, in this block, only three levels were feasible, as the construction of LCASE-UF stimuli was limited by the strict mapping between cases and grammatical functions in Turkish.

Sentence stimuli were constructed using the token-set methodology (Cowart 1997). This resulted in 12 token sets per block and a total of 84 sentences ($12 \times 4 + 12 \times 3$). All stimuli were based on examples of unlike coordination extracted from the Turkish Web 2012 corpus (Baisa & Suchomel 2012). To minimize attrition effects, the materials were split into 4 sub-surveys following the Latin square method. As a result, each participant saw 21 target sentences, along with 22 uncontroversially grammatical or ungrammatical fillers and 3 practice sentences.

3.3 Results

3.3.1 Category block

In the 12 token sets in the category block, the UCAT-LF sentences crucial for the hypothesis contained different categories of adjuncts (9 sentences with different categories selected from: AdvP, NP, and PP), arguments (2

[5]I would like to acknowledge the assistance I received from Adam Przepiórkowski, Katarzyna Kuś, Erkan Şenşekerci, and Szymon Talaga during the implementation of the experiment.

[6]This experiment is also described in Şenşekerci & Przepiórkowski (2024), which proposes an LFG analysis of the relevant data under the assumption of binary grammaticality.

sentences of "PP & NP" coordinations), and predicates (1 sentence of "NP & AP" coordination).

As shown in Figure 1, such UCAT-LF sentences received high scores on average. While LCAT-LF sentences, which featured fully parallel coordinations, were rated slightly higher than UCAT-LF sentences, this difference did not reach statistical significance ($p = .11$). A sharp decline in acceptability was observed only in LCAT-UF and UCAT-UF sentences ($p < .001$ w.r.t. UCAT-LF), where the conjuncts had different grammatical functions.

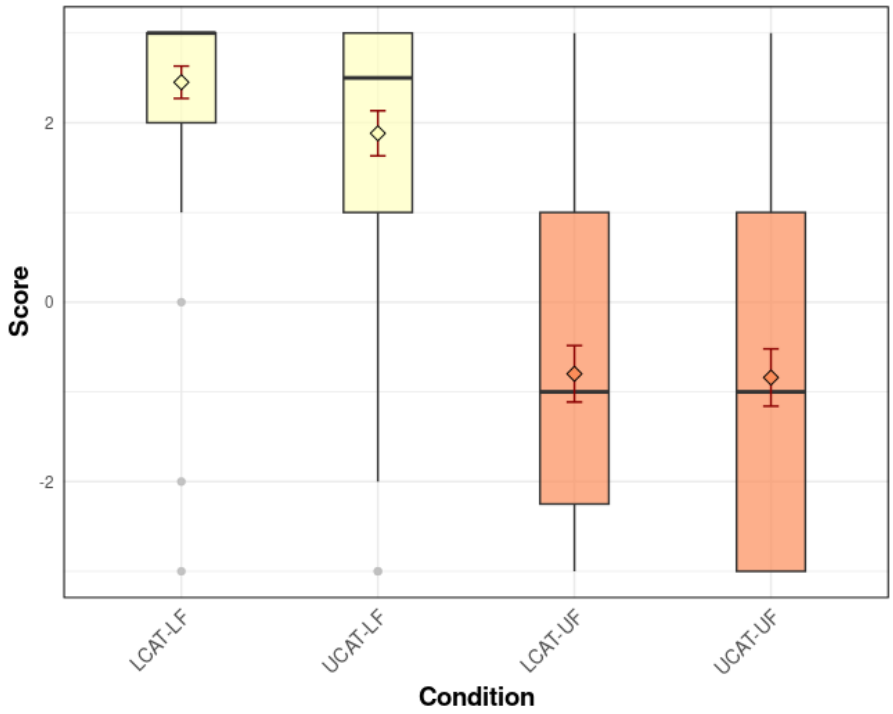


Figure 1: Raw scores of the category block stimuli (y-axis) by sentence type (x-axis), with means indicated by diamonds, and 95% confidence intervals of means by red error bars.

3.3.2 Case block

In the case block, the 12 UCASE-LF sentences with unlike cases but identical adjunct grammatical functions each incorporated cases typical for NP adjuncts: ablative, instrumental, and locative. For example, 4 sentences had coordinations of the type "NP-LOC & NP-ABL".

As shown in Figure 2, these UCASE-LF sentences received significantly lower, yet still positive, judgments compared to LCASE-LF sentences ($p < .001$).

Consistent with the category block results, the average acceptability dropped below zero only for UCASE-UF sentences ($p < .001$ w.r.t. UCASE-LF).

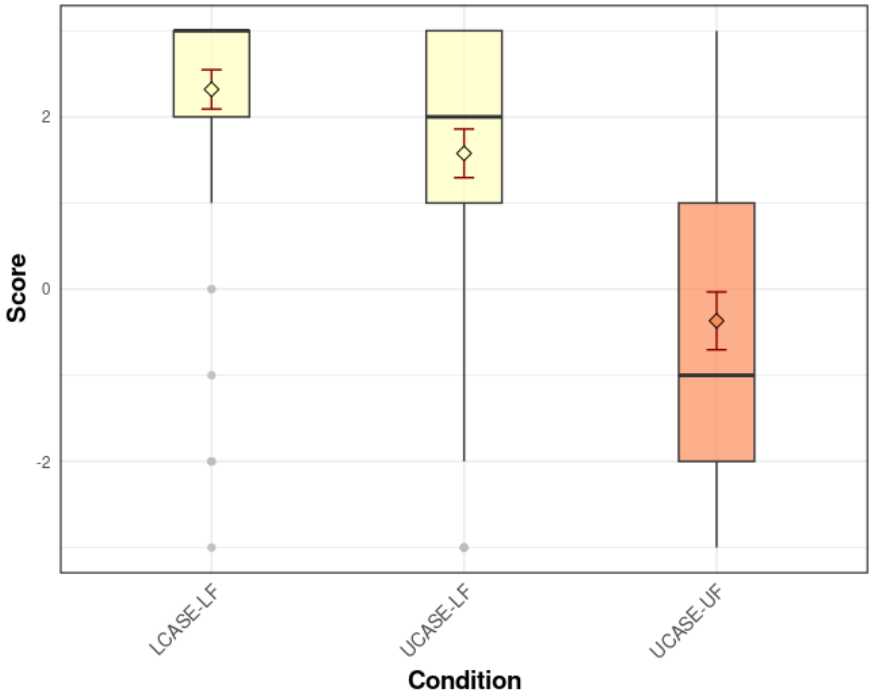


Figure 2: Raw scores of the case block stimuli (y-axis) by sentence type (x-axis), with means indicated by diamonds, and 95% confidence intervals of means by red error bars.

In summary, the results from both experimental blocks support the hypothesis: UCAT-LF and UCASE-LF types of coordination are acceptable. Nevertheless, the fact that such types are not as acceptable as their fully parallel counterparts (i.e., LCAT-LF and LCASE-LF) necessitates a gradient analysis to fully account for the empirical observations.

4 Analysis

As pointed out in the previous section, both coordination of unlike arguments and adjuncts were tested in UCAT-LF and UCASE-LF sentences. Both configurations (i.e., unlike arguments and unlike adjuncts) are acceptable due to the very same reason: satisfaction of disjunctive selectional requirements. Unlike arguments meet the disjunctive requirements imposed on them, while unlike adjuncts modify heads that satisfy the requirements of adjuncts themselves. However, the formal constraints that account for them are different.

4.1 Coordination of unlike arguments

In the case of coordination of unlike arguments, the relevant generalization pertains to the disjunctive requirements imposed by the predicate on the HEAD values of its complements.

For instance, the predicate *sür-* 'last/continue' takes a nominative NP as its subject and a durative complement that can be 1) a nominative NP, as in (2a); 2) a PP projected either by the postposition *boyunca* 'throughout', as in (2b), or *kadar* 'until', as in (2c); or 3) an AdvP, as in (2d).

(2) a. Tahliye çalışma-lar-ı *iki saat* sür-dü.
evacuation work-PL-3.POSS two hour.NOM last-PST
'The evacuation efforts lasted two hours.'

b. Bu kısır döngü *ilk 45 dakika boyunca* sür-dü.
this infertile cycle first 45 minute throughout last-PST
'This vicious cycle continued for the first 45 minutes.'

c. Bu süreç *nisan ay-ı-na kadar* sür-dü.
this phase april month-3P-DAT until last-PST
'This phase lasted until April.'

d. Onlar-ın etki-si *yıl-lar-ca* sür-er.
they-GEN effect-3P year-PL-ADVZ last-AOR
'Their effect lasts for years.'
(Turkish Web 2012; Baisa & Suchomel 2012)

While the coordinated subjects of this verb must be strictly parallel (i.e., all must be nominative NPs), the coordinated complements may mismatch as long as each coordinand satisfies one of the requirements imposed by *sür-*, as in (3).

(3) Bu program [[$_{\text{NP}}$ her hafta] ve [$_{\text{ADVP}}$ saat-ler-ce]] sür-ecek.
this program every week and hour-PL-ADVZ last-FUT
'This program will run every week and for hours.'
(Turkish Web 2012)

To ensure that these selectional requirements are evaluated individually for each conjunct, we employ the c relation (Yatabe 2004, Przepiórkowski 2021), defined in (4). This relation accepts an object and a description as input and checks whether the description holds true for the object.[7] If the

[7]Note that relations can accept descriptions as their inputs in second-order HPSG (Przepiórkowski 2021: 174–178).

input object is a coordination, the relation checks the description against each element in the ARGS list, which contains the HEAD values of each conjunct.

(4) $\forall \boxed{1}_{e}\ \forall \alpha_{et}\ (\ \mathsf{c}(\boxed{1},\ \alpha) \leftrightarrow \alpha(\boxed{1}) \quad \vee$
$\exists \boxed{a_1} \ldots \exists \boxed{a_n}\ (\ \boxed{1}\left[\text{ARGS}\ \left\langle \boxed{a_1},\ \ldots,\ \boxed{a_n} \right\rangle\right] \wedge$
$\mathsf{c}(\boxed{a_1},\alpha) \wedge \ldots \wedge \mathsf{c}(\boxed{a_n},\alpha)\ )\ )$

(Przepiórkowski 2021: 177, ex. (21))

Accordingly, a (simplified) lexical entry for *sür-* can be formalized as shown in (5) where the selectional requirements of *sür-* are checked separately for subject and object position via c relation.[8]

(5) $\begin{bmatrix} \text{PHON} & \langle sür \rangle \\ \text{SYNSEM|CAT|VALENCE} & \begin{bmatrix} \text{SUBJ} & \left\langle \left[\text{CAT|HEAD}\ \boxed{1}\right] \right\rangle \\ \text{COMPS} & \left\langle \left[\text{CAT|HEAD}\ \boxed{2}\right] \right\rangle \end{bmatrix} \end{bmatrix}$

$\wedge\ \alpha_1 \approx (:\sim noun \wedge :\text{CASE} \sim nom)$
$\wedge\ \alpha_2 \approx [(:\sim noun \wedge :\text{CASE} \sim nom) \vee$
$(:\sim postp \wedge (:\text{PFORM} \sim boyunca \vee :\text{PFORM} \sim kadar)) \vee$
$(:\sim adv)]$
$\wedge\ \mathsf{c}(\boxed{1},\ \alpha_1) \wedge \mathsf{c}(\boxed{2},\ \alpha_2)$

4.2 Coordination of unlike adjuncts

An analogous analysis can be applied to unlike adjuncts. However, under standard HPSG assumptions, modifiers select for their heads, which necessitates encoding such disjunctive requirements within the lexical entries of modifiers.

Experimental findings and a related corpus investigation indicate that verbal heads can be modified by 1) any PP, as in (6a); 2) any AdvP, as in (6b); 3) NPs in locative, ablative, or instrumental case, as in (6c); or 4) a coordination of these options, which may involve unlike coordination as in the attested (6d).

[8]The constraint in (5) employs two RSRL operators: '$\sim$' and ':', which are sort assignment and identity functions, respectively. For example, the RSRL description ':CASE $\sim$ *nom*' denotes those objects where the given path, CASE, leads to an object of sort *nom*. Accordingly, the description assigned to α_1 – $(:\sim noun \wedge :\text{CASE} \sim nom)$ – can be informally represented as $\begin{bmatrix} noun \\ \text{CASE}\ \ nom \end{bmatrix}$.

(6) a. Bu ilaç *yemek-ler-den önce* al-ın-malı.
this medicine meal-PL-ABL before take-PASS-NECESS
'This medicine must be taken before meals.'
(Göksel & Kerslake 2010: 103)

b. Adam biz-e *düşman-ca* bak-ıyor-du.
man we-DAT hostile-ADVZ look-PRES.PROG-PST
'(The) man was looking at us with hostility.'
(Göksel & Kerslake 2010: 83)

c. *Son hafta-lar-da* çok yağmur yağ-dı.
last week-PL-LOC a lot rain fall-PST
'It has rained a lot in recent weeks.'
(Göksel & Kerslake 2010: 51)

d. Pamuk-lu çarşaf-lar-ı [yumuşak deterjan-la ve soğuk
cotton-ADJZ sheet-PL-ACC soft detergent-INS and cold
su-da] yıka-yın.
water-LOC wash-2P.IMP
'Wash the cotton sheets with mild detergent and in cold water.'
(Turkish Web 2012)

As for nominal modifiers, they can be 1) any PP, as in (7a); 2) any AdjP, as in (7b); or 3) an unlike category coordination where a PP is coordinated with an AdjP, as in (7c).

(7) a. *Siz-in gibi* insan-lar biz-e yardım ed-ebil-ir-ler.
you-GEN like person-PL we-DAT help do-ABIL-AOR-3PL
'People like you can help us.'
(Turkish Web 2012)

b. *Yeni* bir kitap al-dı-m.
new INDF.DET book buy-PST-1SG
'I bought a new book.'
(Göksel & Kerslake 2010: 83)

c. ... [[$_{PP}$ bir yıl boyunca] ve [$_{ADJP}$ sınır-sız]] gez-me ...
one year throughout and limit-less travel-NMZ
'... limitless sightseeing for a year ...'
(Turkish Web 2012)

Given this highly underspecified relationship between modifiers and their heads – where, for example, practically any PP can modify any verb or a noun – the relevant generalizations can be captured by the following set of

constraints that directly imposes global requirements on the lexical entries of modifiers.

(8) a. $\begin{bmatrix} \textit{postp} \\ \text{MOD}\ \neg\textit{none} \end{bmatrix} \rightarrow \left[\text{MOD}|\text{LOC}|\text{CAT}|\text{HEAD}\ \ \textit{verb} \vee \textit{noun}\right]$

b. $\begin{bmatrix} \textit{adj} \\ \text{MOD}\ \neg\textit{none} \end{bmatrix} \rightarrow \left[\text{MOD}|\text{LOC}|\text{CAT}|\text{HEAD}\ \ \textit{noun}\right]$

c. $\begin{bmatrix} \textit{adv} \\ \text{MOD}\ \neg\textit{none} \end{bmatrix} \rightarrow \left[\text{MOD}|\text{LOC}|\text{CAT}|\text{HEAD}\ \ \textit{verb}\right]$

d. $\begin{bmatrix} \textit{noun} \\ \text{CASE}\ \ \textit{loc} \vee \textit{abl} \vee \textit{ins} \\ \text{MOD}\ \ \neg\textit{none} \end{bmatrix} \rightarrow \left[\text{MOD}|\text{LOC}|\text{CAT}|\text{HEAD}\ \ \textit{verb}\right]$

Crucially, the constraints in (8) merely specify the combinatory possibilities of modifiers but do not alone ensure that only valid instances of like and unlike coordination are licensed in adjunct positions. In any given coordinate structure, all conjuncts must specify the same MOD value and this specification must be shared with the coordination node.

For example, coordination of an AdjP and an AdvP in adjunct position is ill-formed, not because conjuncts have different categories, but because they select different heads – $\left[\text{MOD}\ \ldots\ \textit{noun}\right]$ and $\left[\text{MOD}\ \ldots\ \textit{verb}\right]$, respectively. In order to enforce this parallelism, the following constraint on *coord-phrase* is necessary:

(9) *coord-phrase* $\rightarrow$
$$\left[\text{SYNSEM}|\text{HEAD}\ \begin{bmatrix} \text{MOD}\ \ \boxed{1} \\ \text{ARGS}\ \ \left\langle \left[\text{MOD}\ \boxed{1}\right], \ldots, \left[\text{MOD}\ \boxed{1}\right] \right\rangle \end{bmatrix}\right]$$

(9) ensures not only that all conjuncts have the same MOD value but also that the coordination itself inherits this information. Additionally, if the conjuncts specify $\left[\text{MOD}\ \ \textit{none}\right]$ – i.e., that they are not modifiers – (9) guarantees that the coordination cannot function as a modifier as well. When combined with the *head-adjunct-phrase* constraint illustrated in (10), this analysis now licenses examples like those in (6) and (7).

(10) *head-adjunct-phrase* $\rightarrow$
$$\begin{bmatrix} \text{HD-DTR} & \left[\text{SYNSEM}\ \boxed{1}\right] \\ \text{NON-HD-DTRS} & \left\langle \left[\text{HEAD}\ \left[\text{MOD}\ \boxed{1}\right]\right] \right\rangle \end{bmatrix}$$

(Sag 1997: 475)

4.3 Towards gradience

While the analysis presented thus far accounts for a variety of acceptable configurations of both unlike and like coordination data, it fails to take into account the finding that UCAT-LF and UCASE-LF sentences are somewhat less acceptable than their fully parallel counterparts. Since constraints are violable in Gradient HPSG, this issue can be tackled with two global constraints that would *detect* unlike category and unlike case coordination.

Accordingly, the constraint in (11) checks whether there is a categorical uniformity between the conjuncts: all members of ARGS, which are HEAD values of conjuncts (Yatabe 2004), must uniformly belong to one of the syntactic categories disjunctively specified in the constraint.

(11) *coord-phrase* →
$$\left[\text{HEAD } \boxed{1}\left[\text{ARGS } \left\langle \ldots \right\rangle\right]\right] \wedge \begin{array}{l}[\mathsf{c}(\boxed{1}, (:\sim \textit{noun})) \vee \mathsf{c}(\boxed{1}, (:\sim \textit{adj})) \vee \\ \mathsf{c}(\boxed{1}, (:\sim \textit{postp})) \vee \mathsf{c}(\boxed{1}, (:\sim \textit{adv})) \vee \\ \mathsf{c}(\boxed{1}, (:\sim \textit{verb}))]\end{array}$$

Unlike case coordination can be detected with the constraint in (12), which forces nominal conjuncts to bear the same case only when all the conjuncts are NPs.[9]

(12) *coord-phrase* →
$$\left[\left[\text{HEAD } \boxed{1}\left[\text{ARGS } \left\langle \ldots \right\rangle\right]\right] \wedge \mathsf{c}(\boxed{1}, (:\sim \textit{noun}))\right] \rightarrow \begin{array}{l}[\mathsf{c}(\boxed{1}, (:\text{CASE} \sim \textit{nom})) \vee \\ \mathsf{c}(\boxed{1}, (:\text{CASE} \sim \textit{gen})) \vee \\ \mathsf{c}(\boxed{1}, (:\text{CASE} \sim \textit{acc})) \vee \\ \mathsf{c}(\boxed{1}, (:\text{CASE} \sim \textit{dat})) \vee \\ \mathsf{c}(\boxed{1}, (:\text{CASE} \sim \textit{loc})) \vee \\ \mathsf{c}(\boxed{1}, (:\text{CASE} \sim \textit{abl})) \vee \\ \mathsf{c}(\boxed{1}, (:\text{CASE} \sim \textit{ins}))]\end{array}$$

Marking sentences with these two uniformity constraints for unlike category/case coordination is crucial for obtaining a modelness value that reflects the slightly reduced grammaticality of unlike coordination. However, we still need to establish constraint weights to complete the analysis and obtain modelness values.

4.4 Weight assignment

Assigning weights to specific grammar constraints requires an assumption that each experimental condition corresponds to some grammar constraint. Once

[9]The constraint can potentially be extended to cover configurations where multiple NPs are coordinated with a different syntactic category (e.g., [NP_1, NP_2 & PP]). However, since such configurations were not tested in the experiment, this extension would lack an empirical motivation.

this assumption is made, the quantified impact of an experimental condition on acceptability can be equated with the weight of its formal counterpart in the grammar.

The present analysis assumes that the relative impacts of category and case factors (i.e., LCAT- vs. UCAT- and LCASE- vs. UCASE-) correspond to the weights of the categorial uniformity constraint (see (11)) and the case uniformity constraint (see (12)), respectively. However, the grammatical function factor (i.e., -UF vs. -LF) presents a complex challenge.

Ideally, the grammatical function factor should correspond to a single constraint, much like the one-to-one correspondence between the category and case factors and their respective uniformity constraints. This single constraint would have direct access to the grammatical function of each conjunct and check for functional parallelism between them in a manner analogous to the uniformity constraints.

In the case of uniformity constraints, these checks are relatively straightforward, as the relevant features (i.e., syntactic category and case) are explicitly encoded in the HEAD values of conjuncts. In contrast, grammatical function is a more complex feature which is only implicitly (and partially) encoded in HPSG. Therefore, it is not clear how to formulate a single HPSG constraint that would enforce functional uniformity between conjuncts without revamping core HPSG assumptions. As such, no formal proposal for the functional uniformity constraint is provided in this work.[10]

In summary, the assumed correspondence between conditions and constraints is outlined in Table 1 below.

Conditions	FUNCTIONAL UNIF.	CATEGORICAL UNIF.	CASE UNIF.
LCAT-LF			
UCAT-LF		*	
LCAT-UF	*		
UCAT-UF	*	*	
LCASE-LF			
UCASE-LF			*
UCASE-UF	*		*

Table 1: Summary of the condition-constraint correspondence assumed in the present analysis. '*' indicates a violation of the corresponding constraint.

[10]Alternatively, one could consider a one-to-many correspondence between the relative impact of grammatical function factor and the lexical entries of relevant predicates, such as *sür-* 'last/continue'. However, this approach would not only significantly complicate the analysis and undermine its generalizability, but it would also suggest that the severity of selectional requirement violations (possibly) varies by predicate. While this controversial claim might be true, the relevant experiment does not deal with this question.

As for determining the numeric impacts of experimental conditions, Gradient HPSG does not make an assumption regarding the statistical model utilized for this purpose. However, for methodological soundness, the chosen model must be compatible with a repeated measures design where a participant is tested on a condition multiple times, and, accordingly, consider the dependence between observations. The present analysis relies on linear mixed-effects models to extract weights as such models can take into account the dependence between observations and the individual variability between participants and target sentences.[11]

On the basis of the aforementioned condition-constraint correspondence, a linear-mixed effects model[12] was fitted on the experimental data in question. As standard in experimental syntax, the model treated participants and items as random effects with the sentence type being the sole fixed effect.

According to the fitted model (see Table 2), a sentence that adheres to all constraints (i.e., LCAT-LF and LCASE-LF sentences) is predicted to have an average acceptability of 2.29 (on a scale from −3 to 3), as indicated by the model's intercept. Violating the functional uniformity constraint (denoted by `func_uniformity`) results in an average drop of −2.53 in acceptability, significantly exceeding the individual impacts of categorical and case mismatches, which are −0.33 and −0.62, respectively. In conclusion, these coefficients are assigned to their respective constraints in the grammar.

Fixed Effects	**Coefficients**	**Std. Error**
`(Intercept)`	2.29	0.13
`func_uniformity`	−2.53	0.24
`cat_uniformity`	−0.33	0.11
`case_uniformity`	−0.62	0.14

Table 2: Summary of the fixed effects in the fitted linear mixed-effects model

4.5 Predictions

With the relevant constraints and their weights established, modelness values can now be computed. Consider (13) and (14), which are actual UCAT-LF and UCAT-UF sentences used in the experiment.

[11] For a recent guideline on fitting linear mixed-effects models, refer to Bates, Kliegl, et al. (2015).

[12] The model was trained in R (version 4.1.2; R Core Team 2021) using the `lme4` package (Bates, Mächler, et al. 2015).

(13) Bu isyanlar [[$_{PP}$ yıl-lar boyunca] ve [$_{NP}$ her gün]]
this rebellion-PL year-PL throughout and every day
sür-dü.
last-PST
'These rebellions lasted for years and every day.'

(14) * [[$_{NP}$ Bu savaş-lar] ve [$_{NP}$ toprak-lar-ımız-da]] yıl-lar-ca
this war-PL and land-PL-1PL.POSS-LOC year-PL-ADVZ
sür-dü.
last-PST
'These wars and in our lands lasted for years.'

Table 3 illustrates both the weights (i.e., coefficients extracted from the mixed-effects model) and the modelness values for (13) and (14) based on these weights.

	FUNCTIONAL UNIF. $w = 2.53$	CATEGORICAL UNIF. $w = 0.33$	CASE UNIF. $w = 0.62$	M
(13)	0	1	0	−0.33
(14)	1	0	1	−3.15

Table 3: Modelness of (13) and (14). Numeric values under each constraint column indicate the number of violations per sentence, which is not greater than 1 since the sentences contain no more than one coordinate structure.

The sentence (13) violates only the categorical uniformity constraint, as the conjuncts individually satisfy the disjunctive requirements of *sür-* but bear different categories. Since not all the conjuncts are nominal, the case uniformity constraint is trivially satisfied as well. Thus, the prediction for the modelness degree of (13) is close to 0, which makes it a nearly perfect model of the grammar.

In contrast, the prediction for sentence (14), −3.15, is considerably more negative as (14) violates both functional uniformity and case uniformity constraints because the coordination that occupies the subject position involves a nominative NP (the subject) and a locative NP (an adjunct).

Modelness values can alternatively be interpreted on the original experimental scale by subtracting the non-negated modelness values from the intercept of the mixed-effects model.[13] For instance, (13) is predicted to have an acceptability score of 1.96 (2.29 − 0.33; INTERCEPT − MODELNESS) on

[13]This method does not apply to all statistical models and would not yield the desired results if the original scores are transformed (e.g., into z-scores).

the original scale, while (14) is predicted to score −0.86. These predictions are quite close to the actual mean scores of 1.73 and −0.62 that these sentences received in the experiment.

5 Conclusion

The picture of grammaticality derived from controlled acceptability judgment experiments is inherently gradient, a characteristic also observed in the experiment outlined in this study. To formally analyze the current experimental data within a binary framework of grammar, one would need to posit arbitrary generalizations.

In the Gradient HPSG analysis elucidated here, no such arbitrary measures were needed as the relevant observations could be modeled directly from the experimental data. Consequently, Gradient HPSG presents linguists with a promising avenue to faithfully model their experimental data by utilizing rich representations intrinsic to HPSG.

References

Baisa, Vit & Vit Suchomel. 2012. Large corpora for Turkic languages and unsupervised morphological analysis. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, 28–32. Istanbul, Turkey: European Language Resources Association (ELRA).

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. Parsimonious mixed models. *arXiv*. DOI: 10.48550/arxiv.1506.04967.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.

Beavers, John & Ivan A. Sag. 2004. Coordinate ellipsis and apparent non-constituent coordination. In Stefan Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, 48–69. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2004.3.

Bolinger, D. 1961. *Generality, gradience, and the all-or-none*. The Hague: Mouton & Co.

Brew, Chris. 1995. Stochastic HPSG. In Steven Abney & Erhard Hinrichs (eds.), *Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics*, 83–89. San Francisco, CA: Morgan Kaufmann Publishers Inc. DOI: 10.3115/976973.976986.

Bruening, Benjamin & Eman Al Khalaf. 2020. Category mismatches in coordination revisited. *Linguistic Inquiry* 51(1). 1–36. DOI: 10.1162/ling_a_00336.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton & Co.
Chomsky, Noam. 1961. Some methodological remarks on Generative Grammar. *Word* 17(2). 219–239. DOI: 10.1080/00437956.1961.11659755.
Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. London: SAGE Publications.
Dalrymple, Mary. 2017. Unlike phrase structure category coordination. *Bergen Language and Linguistics Studies* 8(1). DOI: 10.15845/bells.v8i1.1332.
Featherston, Sam. 2005a. The Decathlon Model of empirical syntax. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical and computational perspectives* (Studies in Generative Grammar [SGG] 85), 187–208. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110197549.187.
Featherston, Sam. 2005b. Universals and grammaticality: *wh*-constraints in German and English. *Linguistics* 43(4). 667–711. DOI: 10.1515/ling.2005.43.4.667.
Göksel, Aslı & Celia Kerslake. 2010. *Turkish: An essential grammar* (Routledge Essential Grammars). London: Routledge.
Haegeman, Liliane, Ángel L. Jiménez-Fernández & Andrew Radford. 2014. Deconstructing the subject condition in terms of cumulative constraint violation. *The Linguistic Review* 31(1). 73–150. DOI: https://doi.org/10.1515/tlr-2013-0022.
Hofmeister, Philip, Laura Casasanto Staum & Ivan A. Sag. 2014. Processing effects in linguistic judgment data: (super-)additivity and reading span scores. *Language and Cognition* 6(1). 111–145. DOI: 10.1017/langcog.2013.7.
Keller, Frank. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. University of Edinburgh. (Ph.D. dissertation).
Keller, Frank & Theodora Alexopoulou. 2001. Phonology competes with syntax: Experimental evidence for the interaction of word order and accent placement in the realization of information structure. *Cognition* 79(3). 301–372. DOI: 10.1016/s0010-0277(00)00131-1.
Keller, Frank & Ash Asudeh. 2002. Probabilistic learning algorithms and Optimality Theory. *Linguistic Inquiry* 33(2). 225–244. DOI: 10.1162/002438902317406704.
King, Paul John. 1999. Towards truth in Head-Driven Phrase Structure Grammar. In Valia Kordoni (ed.), *Tübingen studies in Head-Driven Phrase Structure Grammar*, 301–352. Universität Tübingen. http://www.sfs.uni-tuebingen.de/sfb/reports/berichte/132/king/king.dvi.ps (26 November, 2024).

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. 1990. *Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations*. Technical Report CU-CS-465-90. University of Colorado at Boulder, Department of Computer Science.

Müller, Stefan, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.). 2024. *Head-Driven Phrase Structure Grammar: The handbook, Second revised edition* (Empirically Oriented Theoretical Morphology and Syntax 9). Berlin: Language Science Press. DOI: 10.5281/zenodo.13637708.

Parrot, Jeffrey K. 2009. Danish vestigial case and the acquisition of vocabulary in Distributed Morphology. *Biolinguistics* 3(2). 270–302.

Patejuk, Agnieszka. 2015. *Unlike coordination in Polish: An LFG account*. Institute of Polish Language, Polish Academy of Sciences. (Ph.D. dissertation).

Patejuk, Agnieszka & Adam Przepiórkowski. 2023. Category mismatches in coordination vindicated. *Linguistic Inquiry* 54(2). 326–349. DOI: 10.1162/ling_a_00438.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar* (Studies in Contemporary Linguistics). Chicago, IL: The University of Chicago Press.

Przepiórkowski, Adam. 2021. Three improvements to the HPSG model theory. In Stefan Müller & Nurit Melnik (eds.), *Proceedings of the 28th International Conference on Head-Driven Phrase Structure Grammar*, 165–185. Frankfurt am Main: Frankfurt/Main University Library. DOI: 10.21248/hpsg.2021.9.

Przepiórkowski, Adam. 2022. Coordination of unlike grammatical cases (and unlike categories). *Language* 98(3). 592–634. DOI: 10.1353/lan.0.0272.

Pullum, Geoffrey K. & Barbara C. Scholz. 2001. On the distinction between model-theoretic and generative-enumerative syntactic frameworks. In Philippe de Groote, Glyn Morrill & Christian Retoré (eds.), *Logical aspects of computational linguistics*, 17–43. Berlin: Springer.

R Core Team. 2021. *R: A language and environment for statistical computing*. R Project for statistical computing. https://www.R-project.org/.

Richter, Frank. 2004. *A mathematical formalism for linguistic theories with an application in Head-Driven Phrase Structure Grammar*. Eberhard Karls Universität Tübingen. (Ph.D. dissertation).

Richter, Frank. 2007. Closer to the truth: A new model theory for HPSG. In James Rogers & Stephan Kepser (eds.), *Model-theoretic syntax at 10*, 101–110.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33(2). 431–483. DOI: 10.1017/S002222679700652X.

Sag, Ivan A., Gerald Gazdar, Thomas Wasow & Steven Weisler. 1985. Coordination and how to distinguish categories. *Natural Language & Linguistic Theory* 3(2). 117–171. http://www.jstor.org/stable/4047644 (26 November, 2024).

Sag, Ivan A. & Thomas Wasow. 2011. Performance-compatible competence grammar. In Robert Borsley & Kersti Börjars (eds.), *Non-transformational syntax*, 359–377. Chichester: Wiley-Blackwell. DOI: 10.1002/9781444395037.ch10.

Şenşekerci, Berke & Adam Przepiórkowski. 2024. Coordination of unlikes in Turkish. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *The Proceedings of the LFG'24 Conference*, Under review. Konstanz: PubliKon.

Sorace, Antonella & Frank Keller. 2005. Gradience in linguistic data. *Lingua* 115(11). 1497–1524. DOI: 10.1016/j.lingua.2004.07.002.

Wasow, Thomas. 2024. Processing. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean–Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook, Second revised edition* (Empirically Oriented Theoretical Morphology and Syntax 9), 1153–1180. Berlin: Language Science Press. DOI: 10.5281/zenodo.13644940.

Weisser, Philipp. 2020. On the symmetry of case in conjunction. *Syntax* 23(1). 42–77. DOI: 10.1111/synt.12188.

Williams, Edwin. 1981. Transformationless grammar. *Linguistic Inquiry* 12. 645–653.

Yatabe, Shûichi. 2004. A comprehensive theory of coordination of unlikes. In Stefan Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, 335–355. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2004.19.

Topic drop in German: Grammar and usage

Giuseppe Varaschin
Humboldt-Universität zu Berlin

Antonio Machicao y Priemer
Humboldt-Universität zu Berlin

Yanru Lu
Humboldt-Universität zu Berlin

Proceedings of the 31st International Conference on
Head-Driven Phrase Structure Grammar

Palacký University Olomouc, Czech Republic

Stefan Müller, Rui Chaves (Editors)

2024

Frankfurt/Main: University Library

pages 193–217

Keywords: topic drop, german, social meaning, register, HPSG, syntax-pragmatics interface, use-conditions

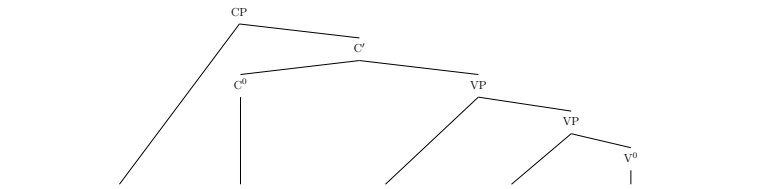
Varaschin, Giuseppe, Antonio Machicao y Priemer & Yanru Lu. 2024. Topic drop in German: Grammar and usage. In Stefan Müller & Rui Chaves (eds.), *Proceedings of the 31st International Conference on Head-Driven Phrase Structure Grammar, Palacký University Olomouc, Czech Republic*, 193–217. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2024.12. CC-BY 4.0

Abstract

German topic drop clauses are a subtype of declarative clauses where the initial position (usually filled by an overt constituent) is left empty. It is often noted that topic drop appears mainly in specific registers (e.g. dialogues), but this claim has neither been previously experimentally validated, nor formally implemented. In this paper, we report the results of a matched-guise study which indicate that the syntactic variation between topic drop and regular V2 declaratives in fact correlates with different social meanings, leading to the register variation postulated in the literature. In order to model German speakers' grammatical and register knowledge about topic drop in HPSG we propose, (i) a unified grammatical constraint that licenses topic drop structures, (ii) a formal theory of register that treats social meanings as a type of use-conditional content subject to compositional rules.

1 Describing topic drop structures

Canonical German declarative clauses consist of a phrase XP in the so called *Vorfeld* (VF) and the finite verb following that constituent in the so called *left bracket* (LB), leading to a verb second (i.e. V2) structure, as Fig. 1 shows (cf. Drach 1937, Wöllstein 2010, a.o.).



Den Aufsatz	*muss*	der Schüler morgen	schreiben.
the essay	must	the student tomorrow	write

'The student must write the essay tomorrow.'

Figure 1: Canonical German declarative clause

The topic drop structure (also called: null topic, pronoun zap, pre-field ellipsis, etc.) being investigated in this paper is a subtype of declarative clauses in German, cf. Fig (1), with a V1 structure (1a), similar to polar questions (1b), but with assertive meaning (cf. Huang 1984, Fries 1988, Cardinaletti 1990, Wöllstein 2010, Müller 2014, Frick 2017, Schäfer 2021).

[†]We would like to thank the audience at the 31st HPSG Conference (Palacky University in Olomouc) and our three anonymous reviewers. We profited from invaluable conversations with David Adger, Emily Bender, Ray Jackendoff, Elin McCready, Stefan Müller, Daniela Palleschi, and Manfred Sailer. All remaining errors are ours. This research was funded by the Deutsche Forschungsgemeinschaft (DFG) – SFB 1412, 416591334, Project A04.

(1) a. [—]$_{\text{VF}}$ muss der Schüler ~~den Aufsatz~~ morgen schreiben.
must the student the essay tomorrow write
'The student must write the essay tomorrow.'

b. [—]$_{\text{VF}}$ Muss der Schüler den Aufsatz morgen schreiben?
must the student the essay tomorrow write
'Does the student have to write the essay tomorrow?'

In contrast to canonical declarative clauses, cf. Fig. (1), one constituent must be dropped (1a), otherwise making the clause ungrammatical (2a). Furthermore, the VF must be left empty in these structures (1a) vs. (2b), although in canonical declaratives this position can be filled by any constituent, see for instance (2c). In movement based analyses (cf. Huang 1984, Cardinaletti 1990; a.o.), this fact has been taken as evidence for an empty element occupying the VF in topic drop structures, and hence not allowing another constituent to occupy this position – hence (2b) – since the German VF can be occupied by only one XP (cf. Machicao y Priemer 2022).[1]

(2) a. * [—]$_{\text{VF}}$ muss der Schüler *den Aufsatz* morgen schreiben.
must the student the essay tomorrow write

b. * [*Morgen*]$_{\text{VF}}$ muss der Schüler ~~den Aufsatz~~ schreiben.
tomorrow must the student the essay write

c. [*Morgen*]$_{\text{VF}}$ muss der Schüler den Aufsatz schreiben.
tomorrow must the student the essay write
INTENDED: 'The student must write the essay tomorrow.'

A further restriction for this construction concerns the information-structural status of the deleted constituent. The dropped XP has to be contextually salient for the purpose of recoverability, cf. (3a) vs. (3b), but it can't be focal, cf. (4). It must be known in the utterance situation, cf. context in (3). Hence, the dropped XP is assumed to be a topic. This differentiates topic drop from pro-drop, which does not posit a topic restriction on the dropped constituent.[2]

(3) A: What's going on with *the essay*?

a. B: [—]$_{\text{VF}}$ muss der Schüler ~~*den Aufsatz*~~ morgen schreiben.
must the student the essay tomorrow write

b. B: * [—]$_{\text{VF}}$ muss ~~der Schüler~~ *den Aufsatz* morgen schreiben.
must the student the essay tomorrow write
INTENDED: 'The student must write the essay tomorrow.'

[1] It is also worth mentioning that topic drop is only possible in main clauses, i.e. when the verb is in the LB, and not in embedded clauses with a complementizer in the LB.

[2] It has been assumed that the German VF is a preferred position for topics (cf. Fries 1988: 24; Wöllstein 2010: 89).

(4) A: Who has seen Lou?

B: * [—]$_{VF}$ hab' ~~*ich*~~ sie gesehen.
have I her seen

INTENDED: '(I) have seen her.'

Even when two constituents are previously mentioned and contextually salient, in a topic drop construction only one constituent (5) can be deleted (cf. Ross 1982, Huang 1984).

(5) A: What's *the student* doing with *the essay*?

B: [—]$_{VF}$ muss *er* ~~*den Aufsatz*~~ morgen schreiben.
must he the essay tomorrow write

B: [—]$_{VF}$ muss ~~*der Schüler*~~ *ihn* morgen schreiben.
must the student it tomorrow write

B: * [—]$_{VF}$ muss ~~*der Schüler*~~ ~~*den Aufsatz*~~ morgen schreiben.
must the student the essay tomorrow write

INTENDED: '(The student) must write (the essay) tomorrow.'

There are also restrictions w.r.t. morphosyntactic and semantic properties of the elements that can(not) be dropped. For instance, while personal pronouns can be omitted (6a), anaphors cannot (6b). From a semantic point of view, pronouns without semantic content cannot be deleted either (7).

(6) A: I've shaved Tim and Tom already, and what about you?

a. B: [—]$_{VF}$ hab' ~~*ich*~~ mich schon rasiert.
have I myself already shaved

b. B: * [—]$_{VF}$ hab' ich ~~*mich*~~ schon rasiert.
have I myself already shaved

INTENDED: 'I have already shaved myself.'

(7) A: How is the weather over there?

B: * [—]$_{VF}$ schneit ~~*es*~~ im August!
snows it in August

INTENDED: '(It) snows in August!'

Therefore, to provide an adequate analysis of topic drop, all of these restrictions must be accounted for. In Sec. 2, we present our analysis dealing with the grammatical aspects of the construction. In Sec. 4, we complement this analysis with constraints concerning the usage of the construction.

2 Licensing topic drop

There are two main ways to account for the grammatical properties of topic drop structures. The first proposal consists of assuming a phonologically

empty category in the VF (*pro* or *Op*), as shown in Fig. 2.[3] This empty category then binds another empty element inside the VP, i.e. the deleted XP (cf. Huang 1984, Cardinaletti 1990; a.o.). There are several difficulties with this approach, for instance how to avoid the realisation of several empty pronouns, how to avoid reflexives and semantically empty expletive pronouns to be realized as phonologically empty elements, and how to restrict the presence of an empty element only to the VF.

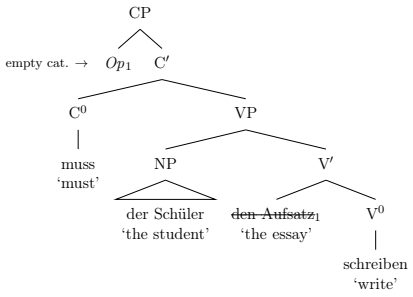


Figure 2: Empty pronoun

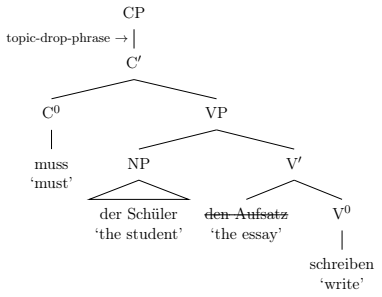


Figure 3: Phrasal constraint

The other proposal, the one followed here, assumes a phrasal constraint named *topic-drop-phrase* (based on Müller 2014: 101), cf. Fig. 3. We enhance this constraint with information-structural details and with restrictions for the deleted element, in order to account for the data presented in Sec. 1, and in Sec. 4.2, it will be further complemented by constraints on usage.

(8) *topic-drop-phrase* ⇒

$$\left[\begin{array}{ll}
\textsc{synsem}|\textsc{context}|\textsc{infostr}|\textsc{topic} & \langle \boxed{2} \rangle \\
\textsc{head-dtr}|\textsc{synsem} & \left[\begin{array}{ll}
\textsc{local}|\textsc{cat} & \left[\begin{array}{ll} \textsc{head} & \left[\begin{array}{ll} \textit{verb} & \\ \textsc{vform} & \textit{fin} \\ \textsc{initial} & + \end{array}\right] \\ \textsc{comps} & \langle\rangle \end{array}\right] \\
\textsc{nonloc} & \left[\begin{array}{ll} \textsc{inher}|\textsc{slash} & \left\langle \boxed{1} \left[\textsc{cont} \left[\begin{array}{ll} \textit{ppro} & \\ \textsc{index} & \boxed{2}\ \textit{ref} \end{array}\right]\right] \right\rangle \\ \textsc{to-bind}|\textsc{slash} & \langle \boxed{1} \rangle \end{array}\right]
\end{array}\right] \\
\textsc{non-head-dtrs} & \langle\ \rangle
\end{array}\right]$$

The *topic-drop-phrase*, cf. (8), is reminiscent of the Head-Filler Rule, proposed in Pollard & Sag (1994: 164), binding off the trace of an element that is being expected in the structure ('the essay' in Fig. 3). In contrast to structures licensed by the Head-Filler Rule, where the LOCAL value of the non-head daughter is token identical with the element in the SLASH list of the head daughter – i.e. a filler and a head are combined, the *topic-drop-phrase* binds

[3] Contrary to Cardinaletti (1990), we do not assume a different syntactic treatment for subject and object topic drop. But, as we show, the two have different use conditions.

off the trace of the element in the SLASH list of the head daughter ('the essay' in Fig. 3), but without combining the head daughter with another element.[4] The Non-Local Feature Principle (Pollard & Sag 1994: 164) guarantees that the mother node in a topic drop structure has an empty SLASH value. In other terms, the *topic-drop-phrase* just eliminates the requirement to overtly attach an XP (co-indexed with the trace), hence the VF is left empty and no empty category is needed, cf. (2). Due to the restriction of the SLASH list to a singleton list (cf. Pollard & Sag 1994: 161, 170) it is ensured that only one constituent can be dropped, cf. (2) & (5), and multiple topics (be they dropped or not) are not allowed in a clause.

The constraint in (8) restricts topic drop structures to main clauses due to the restriction of the HEAD value of the head daugther to INITIAL +. Therefore, embedded clauses in German, i.e. with verb final position, cannot show a topic drop structure (cf. footnote 1). We also adopt the theory of information-structure features in Paggio (2009), where attributes like TOPIC take indices as their values and impose the restriction that the dropped constituent must be a topic in the clause, and for instance not focal, cf. (3) & (4). Furthermore, restricting the CONT value of the element in the SLASH list to *personal-pronoun* (*ppro*) ensures that reflexives (i.e. elements of type *anaphoric*) are ruled out, cf. (6). We also account for the fact that semantically empty pronouns cannot be deleted in a topic drop structure by constraining the INDEX value of the dropped XP to *ref*(*erential*), cf. (7).

That is, only the constraint in (8) is needed in order to account for the grammatical aspects of the construction. The *topic-drop-phrase* restricts quite precisely the realisation of topic drop structures in German, and every other part of the clause up to the verb in second position (cf. C′ in Fig. 3) follows the general constraints related to the German grammar. That is, besides the *topic-drop-phrase* no further stipulations are needed. Furthermore, as will be shown in Sec. 4 the empirical facts related to the usage of the construction and its association with register (cf. Sec. 3) can be formalised within a single model, as has been discussed in Bender (2001, 2007), Paolillo (2000), Asadpour et al. (2022), Machicao y Priemer et al. (2022); a.o.

3 A matched-guise experiment

3.1 Hypotheses and predictions

To investigate the social meanings of topic drop in German, we conducted a matched-guise experiment (Lambert et al. 1975, Bender 2005, Campbell-Kibler 2007). Specifically, we hypothesize that listeners assign different characteristics to topic drop users as opposed to full form users. Thus, we

[4] As a side note, Fries (1988: 24–25) assumes a transformation rule deleting the constituent in the VF, i.e. also without a base-generated empty element.

expect a main effect of topic drop on the ratings of the characteristics of the speakers. Furthermore, we predict that the ratings for speakers who drop the subject are different from those who drop the object, i.e. showing an interaction between the variables topic drop and the topicalized argument.

3.2 Design and procedures

The experiment has a 2×2 within-subjects, within-items design with two independent variables with two levels each: TOPIC DROP (topic drop (TD) vs. full form (FF)) and the topicalized ARGUMENT (subject (S) vs. object (O)). The conditions are illustrated in (9). The materials consist of 8 items (each appears in the four different conditions) and 32 fillers. All items are in the form of written dialogues like (9)[5] and were presented to each participant in a fully randomized order. Participants were tasked to rate speaker B in each dialogue on a 6-point scale (1 = e.g. not friendly at all, 6 = e.g. very friendly) in terms of the following characteristics: *höflich* 'polite', *formell* 'formal', *gebildet* 'educated', *wortgewandt* 'articulate', *freundlich* 'friendly', *pingelig* 'pedantic', *arrogant* 'arrogant', *locker* 'relaxed'. The choice of these characteristics is based on those used in the matched-guise experiment in Beltrama (2018) and on results of prior qualitative interviews with a small group of native speakers addressing attitudes towards topic drop. The experiment was conducted online on the platform Ibex farm.[6]

(9) A: Hast du letzte Woche den Brief geschrieben?
have 2SG.NOM last week DEF.SG.ACC letter written
'Did you write the letter last week?'

B: a. Ich kann ihn morgen schreiben. [FF×S]
1SG.NOM can 3SG.ACC tomorrow write

b. Kann ihn morgen schreiben. [TD×S]
can 3SG.ACC tomorrow write

c. Den kann ich morgen schreiben. [FF×O]
DEM.3SG.ACC can 1SG.NOM tomorrow write

d. Kann ich morgen schreiben. [TD×O]
can 1SG.NOM tomorrow write

'I can write (it) tomorrow.'

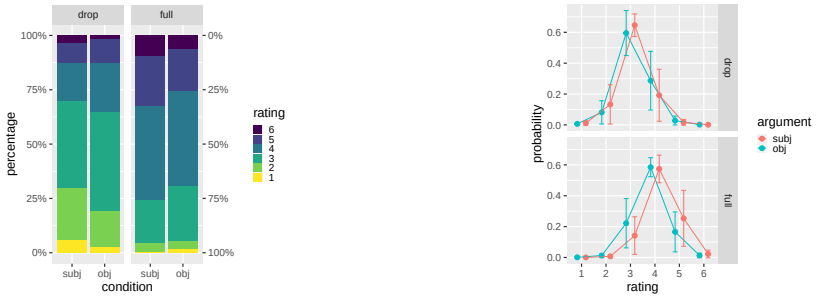
[5] As a reviewer points out, the use of written instead of spoken stimuli may have an influence on participants' perception of topic drop, as it is arguably a phenomenon associated with conceptually spoken language. As a first step, the dialogue form adopted here aims to simulate a spoken conversation as far as possible. Further studies using spoken stimuli are definitely worth carrying out. However, as several corpus studies show, topic drop is also widely used in the medium of informal written communication, e.g. text messages, chats, mails (Frick 2017, Schäfer 2021, a.o.). The stimuli we used are also compatible with such kinds of contexts.

[6] `https://korpling.german.hu-berlin.de/ibex/`

23 self-reported German native speakers (17 female, 3 male, 1 diverse, 2 not specified) participated in the experiment. 21 are between the age 18 and 25, one is between 36 and 45 and another between 56 and 65.

3.3 Data analysis and results

The data is analyzed in R (R Core Team 2023) with cumulative link mixed models (CLMMs) for ordinal data (Christensen 2022). The model includes TOPIC DROP and ARGUMENT as main effects, the effect of their interaction, and PARTICIPANTS and ITEMS as random effects, including both by-participant and by-item random intercepts and slops, based on the maximal random effects approach recommended by Barr et al. (2013).[7] The ratings on each scale are analyzed in a separate univariate analysis. The results show a main effect of TOPIC DROP for the scales polite ($\chi^2 = 15.30$, $p < 0.01$), formal ($\chi^2 = 20.51$, $p < 0.01$), educated ($\chi^2 = 18.99$, $p < 0.01$), articulate ($\chi^2 = 22.12$, $p < 0.01$), friendly ($\chi^2 = 246.38$, $p < 0.01$) and pedantic ($\chi^2 = 4.71$, $p = 0.03$), but not for the scales arrogant and relaxed. Specifically, participants find speakers who use topic drop less polite, less formal, less educated, less articulate, less friendly and less pedantic compared to their counterparts who use the full form. Furthermore, we only find an interaction between TOPIC DROP and ARGUMENT for the scales polite ($\chi^2 = 7.66$, $p < 0.01$) and formal ($\chi^2 = 4.89$, $p = 0.03$). Participants rate speakers who use subject topic drop as even less polite and less formal than those who use object topic drop. Figure 4–11 illustrate the data of each scale. The (a) figures show the percentage of each rating by condition. The (b) figures present the predicted probability for each rating.

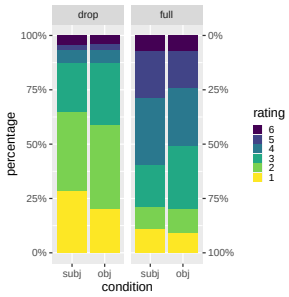


(a) % of rating by conditions

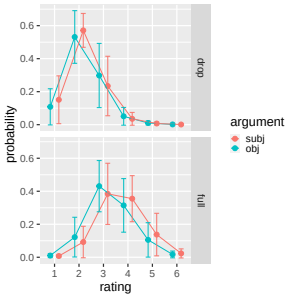
(b) Predicted probability for rating

Figure 4: Data of the polite scale.

[7] clmm(ratings ~ conddrop * condarg + (1 + conddrop * condarg | participant) + (1 + conddrop * condarg | item), data = data_polite)

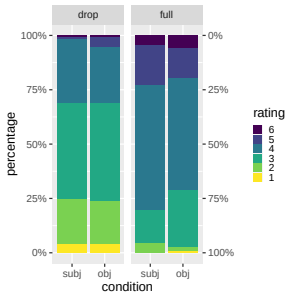


(a) % of rating by conditions

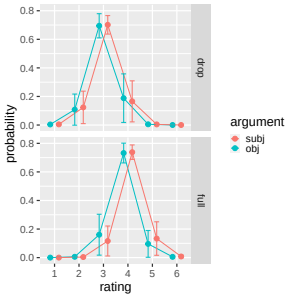


(b) Predicted probability for rating

Figure 5: Data of the formal scale.

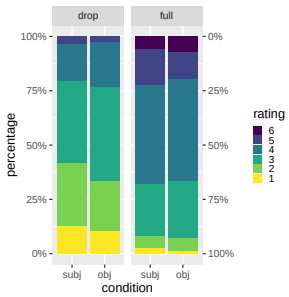


(a) % of rating by conditions

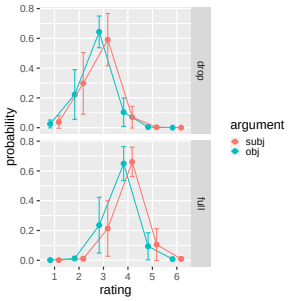


(b) Predicted probability for rating

Figure 6: Data of the educated scale.

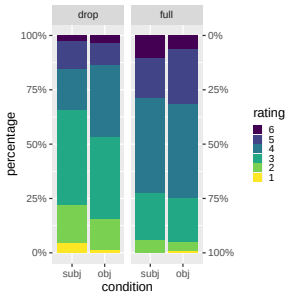


(a) % of rating by conditions

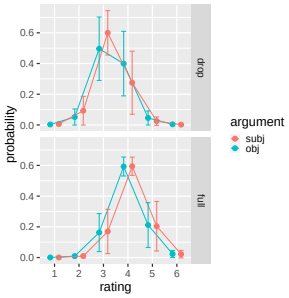


(b) Predicted probability for rating

Figure 7: Data of the articulate scale.

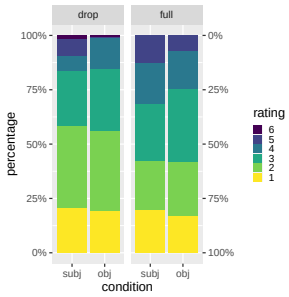


(a) % of rating by conditions

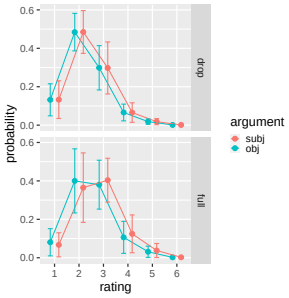


(b) Predicted probability for rating

Figure 8: Data of the friendly scale.

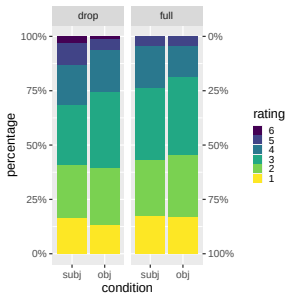


(a) % of rating by conditions

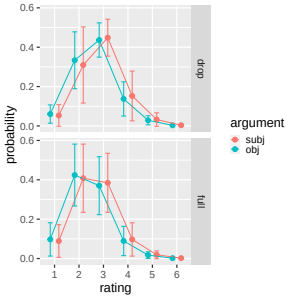


(b) Predicted probability for rating

Figure 9: Data of the pedantic scale.

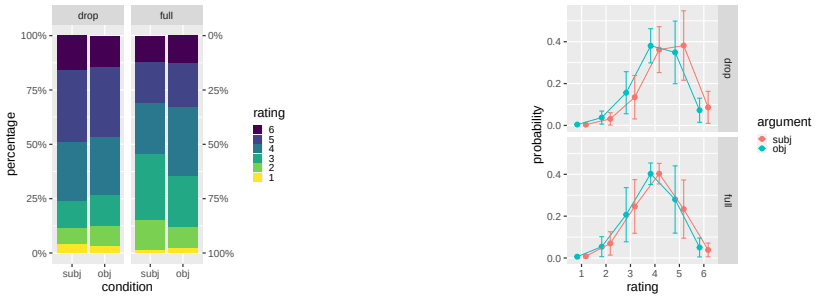


(a) % of rating by conditions



(b) Predicted probability for rating

Figure 10: Data of the arrogant scale.



(a) % of rating by conditions

(b) Predicted probability for rating

Figure 11: Data of the relaxed scale.

4 A model of register competence

These results make it clear that German speakers know more about topic drop than what the structural licensing conditions proposed in Sec. 2 suggest. In addition to being able to form grammatical topic drop sentences with a corresponding at-issue meaning, speakers also know how and by whom topic drop is typically used. In this section, we propose a model that allows us to unify these two kinds of knowledge – structural and use-conditional register knowledge – under a single competence theory using HPSG.

4.1 Grammar and use conditions

In order to attach register information to linguistic structures we assume that, along with constraints like (8) above, grammars of natural languages include use-conditional constraints (UCCs), with the overall form in (10).

(10) *description of linguistic structure* $\mathcal{S} \Rightarrow$ *description of a context for* $\mathcal{S}$

The descriptions in the antecedent of a UCC specify a class of independently licensed structures on which the consequent imposes a contextual appropriateness condition. The antecedents need not always be primitive types, but can be complex descriptions, since use-conditional content can be indexed by structures that are larger than the minimal pieces needed for basic grammatical rules (cf. Bender 2001: 281–282; Bender 2007: 368–370).

The contextual descriptions in the consequent of UCCs can be modeled as restrictions on the values of the CONTEXT attribute, in line with previous formalizations of register in HPSG (Wilcock 1999, Paolillo 2000, Bender 2001, 2007, Asadpour et al. 2022). Crucially, we assume that register-sensitive forms like topic drop constrain their contexts through the expression of conventionalized social meanings (SMs) (Bender 2001, 2007, Burnett 2019, Taniguchi 2019, Beltrama 2020, Asadpour et al. 2022, Salmon 2022).

We use the term SM in a broad sense to denote any kind of non-at-issue content that indexes some socially relevant property of (at least) one of the context coordinates, i.e. the values for C-INDEX features (SPEAKER, ADDRESSEE, etc). According to the results of our experiment, an utterance like B's in (11), conveys both the at-issue meaning (11a) and a SM that can be paraphrased as (11b). Since we take SMs to be conventionalized, it is not necessary for B to have a conscious intention to convey (11b) – all that matters is that (11b) is consistently associated with topic drop by speakers.

(11) B: Muss sie morgen verkaufen.
must 1SG.F.NOM tomorrow sell
a. 'She must sell (it) tomorrow.'
b. 'I am not formal, not friendly, or not articulate...'

The appropriateness of SMs in a situation depends on interlocutors' beliefs, goals and intentions. We can think of a register as a cluster of linguistic constraints whose associated models carry SMs that are appropriate in the same types of situations. For example, the constraints responsible for topic drop and other constructions/lexical items compatible with SMs like those in (11b) could be thought of as belonging to a common 'informal' register.[8]

With respect to their projective and compositional properties, SMs pattern with expressive meanings and other types of conventional implicatures (McCready 2019, Taniguchi 2019, Asadpour et al. 2022, Salmon 2022). The following are some of the formal properties of SMs that we want to capture.

(12) a. **Independence**: SMs contribute to a dimension of meaning that is separate from the at-issue content of the utterance. This means that the basic content of an utterance can remain the same regardless of which SM is expressed and that SMs are not affected by truth-conditional operators (e.g. modals, negation).
b. **Indexicality**: SMs predicate something of the present utterance situation or its participants. That is, SMs always describe an individual/situation that is contiguous with the utterance.
c. **Gradability**: Contexts can be distinguished in terms of SMs at a fine-grained level. The applicability of a SM is not a matter of all-or-nothing; rather, SMs hold of individuals to different degrees, which may change gradually as the dialogue progresses.

[8]Most constraints are underspecified for register because UCCs only associate SMs to a small subset of linguistic objects. So, for example, the combination of the definite determiner *das* and *Buch* 'book' as a *head-specifier-structure* is arguably register neutral. This view of registers is more flexible and multidimensional than the one in Machicao y Priemer et al. (2022), because the grammar signature does not need to commit to a finite set of registers. Registers are epiphenomenal to SMs and any combination of SMs can potentially be used to define a register, provided the SMs are appropriate in similar contexts. Furthermore, in contrast to Wilcock (1999), there is no grammar-internal requirement that different parts of a sentence need to share exactly the same SM or register.

d. **Underspecification**: Forms subject to register variation are not associated with a unique and specific SM, but with a set of inferentially related SMs – i.e. an indexical field (Eckert 2008)

Property (12a) is modeled by representing SMs as values of the C(ONVENTIONAL)I(MPLICATURE) attribute inside CONTEXT, following Asadpour et al. (2022).[9] We assume that operators over at-issue content (e.g. negation, modals, interrogatives) only pick out their scopal arguments from the set of relations under CONTENT|RELS. Property (12b), which is related to the nondisplaceability property in Potts (2007: 169–173), is captured by requiring all SM relations to have a C-INDEX value as one of their arguments. We can have SMs that predicate something of the SPEAKER, the ADDRESSEE, and possibly also of the TIME and SITUATION where the utterance took place; e.g. forms like *thou* in English, which (for present speakers) arguably encode a SM to the effect that the utterance has taken place in the distant past.[10]

In order to model property (12c), we require SMs to take a DEGR(EE) argument (an interval from 0 to 1), similar to the approach in Potts & Kawahara (2004: 261) and McCready (2019: 29). This non-discrete continuous encoding captures the fact that we can make comparative judgments about SMs (e.g. speaker S_1 is more formal than speaker S_2) and even have an intuition that an expression has a particular SM to a higher extent than another expression. Finally, we address property (12d) by representing the indexical field conventionally associated to each linguistic variant as multiple inheritance hierarchy of SMs, like the one in Fig. 12.[11] To enforce (12d), all we need to do is state our grammar so that UCCs always associate structural descriptions with abstract SM types. These types are only resolved to maximal SM sorts in concrete communicative situations, in accordance with probabilistic principles (see Burnett 2017, 2019 for a proposal).

[9] Asadpour et al. (2022) also propose that the ascription of a SM is embedded under attitude predicates expressing that such ascription is relative to the speakers' beliefs about what the communicative norms in a linguistic community are (Green 1994). For reasons of space, we do not explore this possibility here and assume simpler SM structures.

[10] This does not imply that *thou* can only be used in the distant past. Rather, we can exploit the association between *thou* and *pastness* to convey a stylistic effect, which can be interpreted as ironic given the incongruence between this SM and the present context.

[11] We opted for a simple formulation of the hierarchy where each of the adjectives we tested reflects a property of the speaker (given the nature of the matched-guise task) and is also a maximal sort. This is intended as a crude approximation only – there are arguably more realistic alternatives. It is likely that *polite* is underspecified, in that it can be interpreted either as property of the speaker (e.g. as equivalent to *formal*) or as a relational property indicating (social or psychological) distance (McCready 2019: 28–29). Similarly, *friendly* can plausibly be reduced to something like *psych-prox*. Manfred Sailer (p.c.) also suggested to us the idea that some of the *cognitive* traits are contextual inferences drawn from a single general SM – e.g. something like *explicit-expression*. The fact that many of the maximal SM sorts in Figure 12 tend to occur together could also follow from a theory of communicative stereotypes, which should be part of the story about how speakers resolve underspecified SM sorts in concrete communicative situations.

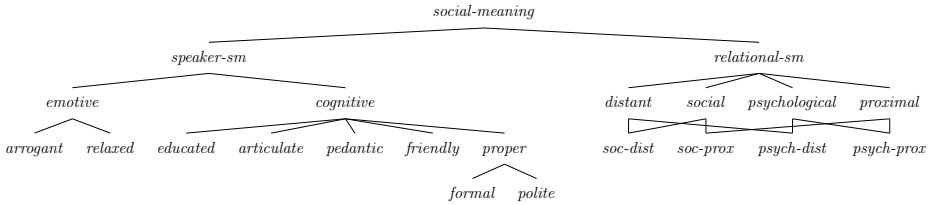


Figure 12: Social meaning hierarchy

The major distinction in Fig. 12 is between SMs that concern the way the speaker presents themselves (*speaker-sm*) and those that say something about the relation between speaker and some other individual, typically the hearer (*relational-sm*). The SM of topic drop is of the former kind, while the latter is arguably what is grammaticalized by honorific pronouns. This explains, for instance, why we do not see a SM clash when topic drop structures are used in conjunction with honorific forms like the 2P pronoun *Sie*; see (18) below. Kaur & Yamada (2022) make a similar proposal to explain the interaction between allocutive markers and 2P honorifics in Japanese.

4.2 The social meanings of topic drop

With this basic toolkit in place, we can begin to state our account of the usage preferences and SMs of topic drop. Since our analysis of topic drop appeals to a new phrasal type (as opposed to a phonologically empty pronoun), we cannot attach the SM of topic drop to a lexical item. Rather, we have to associate the SM to the phrasal type itself. In the case of topic drop with subjects, we need not introduce an ad hoc phrasal type, but can simply identify the range of structures that grammaticalize the relevant SMs as structures of the type *topic-drop-phrase* with NP_{nom} in their SLASH list.

In order to represent SMs introduced at a phrasal level, we introduce a C(ONSTRUCTIONAL)-CI feature taking a list of SMs as value. This feature plays an analogous role to C(ONSTRUCTIONAL)-CONT in the composition of at-issue content (Copestake et al. 2005: 319–321). We propose to model the results of the experiment in Sec. 3.1 by means of the UCCs in (13)–(14). These constraints act solely on the situational level, requiring a *n(on-)e(mpty)* list of SMs to be present inside the CONTEXT attribute of signs.[12]

(13) *topic-drop-phrase*⇒
$$\left[\text{CTXT}\left[\begin{array}{l}\text{C-INDS}|\text{SPEAKER}\ \boxed{1}\\ \text{C-CI}\ \textit{nelist}\left(\left[\begin{array}{ll}\textit{cognitive}\\ \text{ARG}\ \ \boxed{1}\\ \text{DEGR}\ \ (0,\ .5]\end{array}\right]\right)\end{array}\right]\right]$$

[12]Different maximal sorts of the relevant SM types may be chosen for each utterance (but always at least one such sort), depending on the particulars of the situation. For instance, if a speaker is talking to close friends, it is likely that an utterance of topic drop would merely convey low levels of *pedantic* or *formal* – as opposed to low *education*.

(14) $\left[\begin{array}{l}\textit{topic-drop-phrase}\\ \textsc{hd-dtr}|\textsc{slash}\ \left\langle \mathrm{NP}_{\textit{nom}} \right\rangle\end{array}\right] \Rightarrow \left[\textsc{ctxt}\ \left[\begin{array}{l}\textsc{c-inds}|\textsc{speaker}\ \boxed{1}\\ \textsc{c-ci}\ \textit{nelist}\left(\left[\begin{array}{l}\textit{proper}\\ \textsc{arg}\ \ \boxed{1}\\ \textsc{degr}\ (0,\ .3]\end{array}\right]\right)\end{array}\right]\right]$

The UCC in (13) captures one half of the main effect we observed for the TOPIC DROP variable in our experiment: namely, the association of topic drop structures with contexts where the speaker is presenting as having a low degree for at least one of the *cognitive* traits in Figure 12. The UCC in (14) captures the overall interaction effect between our two experimental variables (TOPIC DROP and ARGUMENT).

Regarding association between V2 structures and a high degree of *cognitive* traits (the other half of the main effect we observed in our experiment), there are two alternatives. The most straightforward one would be to posit a UCC parallel to (13) but where the antecedent identifies declarative V2 and the consequent requires CIs with *cognitive* predications whose DEGR values are on the opposite of those required by topic drop. One of the reviewers of this paper has convinced us that this is probably not the best approach. First of all, the fact that the two variants of the variable (V2 and topic drop) are associated with complementary intervals on the same scale ends up being entirely accidental. This fails explain why SMs are generally tied to variability, with each variant of a variable expressing a different partition of the scalar property corresponding to the SM (Eckert 2008, Oushiro 2019).

Following the reviewer's suggestion, we model this effect in pragmatic terms. We propose that declarative V2 is neutral with respect to SMs as far as the grammar is concerned. The perception that V2 structures index higher degrees for *cognitive* traits than topic drop is an implicature that emerges from an interaction between the SMs of its salient alternative (topic drop) and an extension of *Maximize Presupposition!* (Heim 1991, Singh 2011) to all non-at-issue meanings – as proposed in McCready (2019: 53) and Oshima (2021: 179). We can state the principle informally as follows:

(15) **Maximize Non-At-Issue Content!**
If an expression $\mathcal{E}_1$ has the same at-issue content as $\mathcal{E}_2$ and the conventional non-at-issue content of $\mathcal{E}_2$ is stronger than that of $\mathcal{E}_1$ and appropriate in a context c, then $\mathcal{E}_1$ should not be used in c.

Topic drop and filler-gap constructions are equivalent in terms of descriptive at-issue CONTENT because neither of the schemas licensing these structures introduce constraints on this level. Therefore, they are alternatives subject to (15). Since topic drop has a stronger CI content than V2 (i.e. it expresses a low degree for *cognitive* SMs, while the CI content of V2 is vacuous), topic drop must be used when appropriate. If it is not used, the hearer can infer that it is not appropriate – i.e. that the context is not such that the speaker is presenting as having a low degree for *cognitive* SMs.

Now that we know what SMs look like and which SMs are expressed by topic drop, we turn to the question of how SM projection works – i.e. how the SMs words and constructions interact to form the SM of a full utterance.

4.3 Social meaning composition

Taking this into account, we propose the principle in (16) to account for the projection of SMs at the level of a single utterance. Following Potts' (2007, 185) theory of expressive meanings, our principle distinguishes two basic cases of SM composition: one where the SMs to be composed are independent and another when they involve repeated predications. By repeated predications we mean SM predications of the same type (as per Figure 12) and with the same ARG values, but possibly different DEGR values.

(16) **Local CI Projection Principle**

a. For each phrase, if its C-CI value and the CI values of its daughters do not have **repeated predications**, then the CI value of the phrase is the concatenation of the CI values of its daughters and its own C-CI value.

b. For each phrase, if its C-CI value and the CI values of its daughters have **repeated predications** $SM_1, \ldots SM_n$ then the CI value of the phrase is the concatenation of the CI values of its daughters and its C-CI value **minus** $\langle SM_1 \rangle, \ldots \langle SM_n \rangle$ **plus** a list of predications of the same type and with the same ARG values as $SM_1, \ldots SM_n$, but with a DEGR value consisting in the intersection between the DEGR values of $SM_1, \ldots SM_n$.

Figure 13 illustrates a case where (16a) is relevant. Figure 14 shows a case where (16b) applies, giving rise to a structure where the degree of formality of the mother is an intersection of that of its daughters. The intersection of DEGREE values δ_1, δ_2 is the set of points that are in both δ_1 and δ_2. Clause (16b) also imposes consistency requirement on SMs: repeated SMs in the C-CI of a phrase or in the CI values of its daughters must intersect, otherwise no CI value is defined for the mother, cf. Figure 15.[13]

[13]Since our UCCs (13)–(14) only require non-empty lists of SMs predications of the relevant types, it is in principle possible to circumvent clashes by simply resolving the underspecification to a different maximal sort. For instance, if one structure is required to have a *formal* SM with a DEGRE value [.7, .9], it is in principle still possible to use it in combination with topic drop as long as the underspecified SMs of topic drop are resolved to other maximal sorts (e.g. low degrees for *polite* or *friendly*). The result might still be pragmatically odd due to our stereotypes about communicative situations – i.e. it is not common for a person to be at the same time very formal and very impolite. But this outcome is not blocked by the grammar itself. We think this flexibility is desirable, especially when we compare the SMs of topic drop with more 'grammaticalized' SMs like those of Thai honorifics (McCready 2019). In the latter case, conflicting forms are much

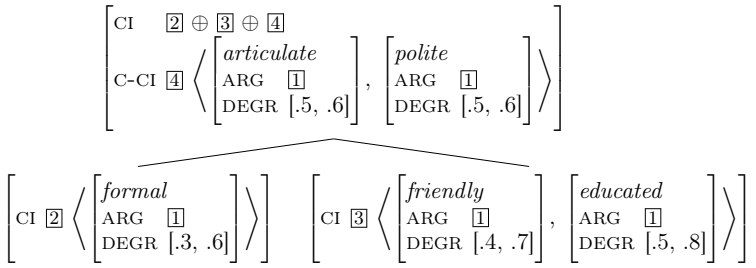


Figure 13: Simple SM composition

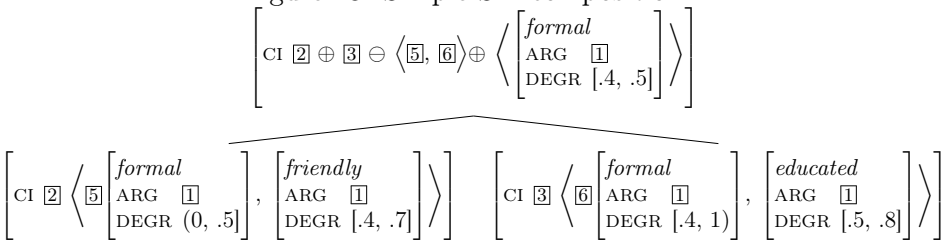


Figure 14: Complex SM composition

According to both clauses in (16), the amount of information in CI values increases monotonically as one moves upwards in the tree. The constraint in (16a) preserves the SMs of the daughters and adds them to the mothers. (16b) either preserves or narrows the DEGREE interval in the (repeated) SMs of the daughters, imposing further constraints on the contexts where the mother can be admissibly uttered. This captures another property of SMs, namely, the fact that repetition of SMs does not give rise to redundancy, but to reaffirmation or specification, unlike what we see when at-issue meanings are repeated (Potts 2007, Smith et al. 2010). This property can be tested with a context like (17), adapted from Taniguchi (2019: 19–20). A's impatient response only makes sense if A perceives B as being redundant. This is not the case when topic drop is used twice, as in B's reply.

(17) B: Muss sie verkaufen. Kann ihr vielleicht helfen.
must 3SG.F.NOM sell can 3SG.F.DAT maybe help
'She must sell (it). Maybe I can help her (with it).'

A: # Ja ja, ich verstehe. Du sprichst informell!
INTER 1SG.NOM understand 2SG.NOM speak informally
'Ugh. I get it already! You are speaking informally.'

As an illustration of how SM composition works in a specific structure, consider the example of object topic drop in (18).

more prone to give rise to ungrammatical clashes. We can model this by stating UCCs that assign to alternative honorific forms fully specified SMs pertaining to complementary intervals in a single dimension (e.g. low and high degrees of *polite*). In that case, it would not be possible to avoid a clash by resolving the SMs to a different maximal sort.

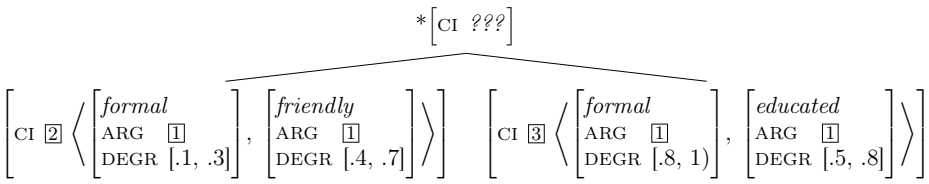


Figure 15: Undefined social meaning composition

(18) B: Muss sie Ihnen morgen geben.
must 3SG.F.NOM 2SG.DAT tomorrow give
'She must give (it) to you tomorrow.'

For the sake of simplicity, we omit most of the syntactic details, and focus on the core parts of the CI composition in Figure 16. In the middle field, there is an honorific 2P pronoun (*Ihnen*), which we assume encodes as a SM a moderate-to-high degree of social distance. In the CI value of the mother, this CI is appended to the CI enforced by the topic drop construction itself (13) – i.e. a non-empty list of *cognitive*-typed predications.

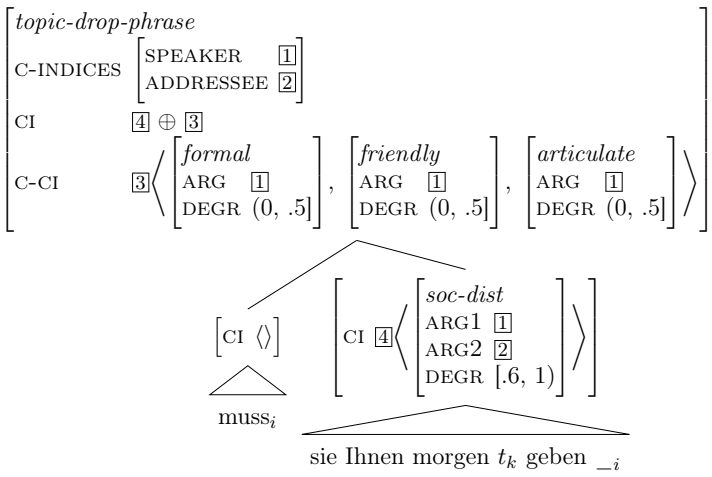


Figure 16: SM composition with Object Topic Drop

Subject topic drop (19) functions in a similar way. Figure 17 illustrates how the SMs project. We assume a context where two *formal* SMs are chosen as part of the C-CI value, corresponding to (13) and (14).

(19) B: Muss es heute verkaufen.
must 3SG.NEUT.ACC today sell
'(She) must sell it today.'

As Paolillo (2000) notes, there is only one type of context where SMs do not project from daughters to their mothers: direct speech reports or

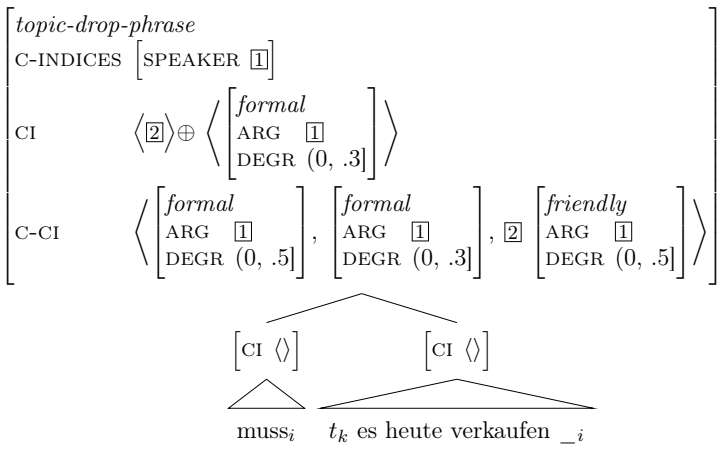


Figure 17: SM composition with Subject Topic Drop

quotation. This exceptional behavior follows from (16) if we assume that such contexts involve some kind of unary rule that shifts the C-INDEX of a clause to some individual other than the speaker of the full utterance (e.g. the AGENT of a quotative verb like *say*). Since SMs get their ARG value from C-INDEX|SPEAKER, the SM under the CI of a quote will not get attributed to the speaker of the full utterance, but to the speaker of the quotation itself.

The felicity of SMs is constrained by the prior global context, which we can think of as containing a list of SMs. The 'register appropriateness' of utterances carrying SMs is constrained by the context in the following way:

(20) **Felicity constraint** (based on McCready 2019: 31)
For every utterance U expressing a SM α, if the **prior context** of U is contains a SM α', where α and α' are repeated predications, then the DEGREE values of α and α' **have to intersect**.

All things being equal, the more the SMs in the prior global context and those in U's CONTEXT|CI value match (the more SM predications they have in common and, for each of these, the more their DEGREE intervals overlap relative to their total size), the more appropriate U is with respect to the context. On this view, topic drop is more frequent in 'less formal' contexts because it is grammatically constrained to have SMs that largely match with the SMs that define these contexts (e.g. low values for *educated* or *formal*). We assume that if the output of Local CI Composition for an utterance U is felicitous, it updates the global context, dynamically pushing it in the direction of the SM contribution of U (McCready 2019).[14]

[14]We called the SM composition principle in (16) "local" because we assume that the way SM composition works locally (i.e. in the context of a single utterance) is fundamentally different from the way it works in the context of a larger discourse. The latter needs to be

This view entails that registers are not theoretical primitives. The only primitives are SMs, which are features of situational contexts and are also contributed by linguistic signs (and constrained by UCCs). However, registers can be reconstructed as clusters of linguistic constraints whose associated models are required (by virtue of UCCs) to carry SMs that are appropriate in the same global contexts. Since SMs are gradable, whether or not a form 'belongs' to a register R, is also essentially a matter of degree: it depends on how much its SMs match the contextual parameters associated with R.[15]

5 Concluding remarks

In this paper, we proposed an HPSG model that integrates structural and usage knowledge into a single theory of linguistic competence, building on previous efforts (Paolillo 2000, Bender 2007, Asadpour et al. 2022). We explored these issues in connection to a specific instance of variation: the realization of declarative clauses in German with and without topic drop.

On a grammatical level, the analysis we proposed here entails that a single grammar licenses both attested variants of a variable (topic drop, V2) as different phrasal constructions. To account for the usage preferences we proposed a mapping between linguistic features and situational parameters that uses the same descriptive vocabulary used for grammatical constraints. These UCCs relate independently licensed structures to underspecified SMs, which function as abstractions over the details of particular situations.

Note that the schemas that constrain SMs are kept separate from those responsible for the basic aspects of the form and meaning of utterances – i.e. the 'core grammar' of a language. For instance, we do not build in the SMs of topic drop directly into the schema (8); rather, we posit the separate UCCs in (13)–(14). Though the two options are often extensionally equivalent, there are good reasons for keeping things separate in this way.

First, as we mentioned above, the pieces of structures SMs get attached to are not always ones that are described by independently required constraints

handled by a separate principle that takes the yield of (16) as its input, and outputs a modified global context. This immediate context-shifting effect is what makes SM-bearing elements similar to performatives (Potts 2007: 179–181). A distinction between sentence-level and discourse-level composition of SMs is also proposed in Paolillo (2000: 243) and McCready (2019: 31–33). For instance, SM clashes (e.g. mixing formality and informality markers) are much more prone to give rise to reduced acceptability at the utterance level than across longer stretches of discourse (Wilcock 1999). For discourse, something like the averaging approach that McCready (2019: 32) proposes for local composition could turn out to be better than the intersective approach we formulate in (16).

[15]This is an idealization. Other factors are arguably relevant to define which SMs are appropriate in which situations in addition to overlap with the prior context in the sense of the Felicity Condition in (20). One important additional set of such factors are speakers' goals and beliefs, which are estimated on probabilistic grounds, as described in Burnett (2017, 2019). If we take these into account, registers have to be inferred on the basis of the cues provided by SMs and their probabilistic interplay with such situational parameters.

on lexical or phrasal types. For example, as far as we can see, there is no constraint in the core grammar of German that specifically needs to refer to subject topic drop structures. Nonetheless, such structures are still targeted for the expressions of particular SMs, as (14) makes clear.

Second, keeping constraints on SMs in a separate 'corner' of the grammar is more in line with what is known about how register is processed in real time. The perception of register incongruence tends to come about later (and with less intensity) than that of grammatical errors (Münster & Knoeferle 2018, Almeida 2023, Plesca et al. 2024, i.a.). One way of capturing this is by formulating a model of processing where speakers first parse utterances according to their core grammar and only then evaluate them according to UCCs, checking for contextual felicity – an inherently graded notion.

In spite of this separation, UCCs use the same descriptive vocabulary and implicational format as standard HPSG rules. In this sense, use-conditional 'register' knowledge and core grammar are subsumed under one and the same competence theory. Insofar as SMs are modeled as part of individuals' linguistic competence, we predict speakers to be capable of manipulating variants in order to actively construct new registers and personal linguistic styles. Furthermore, the division of labour between the grammar (which deals with SMs) and probabilistic usage preferences (which relate SMs to concrete situations) simplifies the task of defining the register potential of complex expressions from that of their parts – one of the main challenges for register modeling in HPSG noted by Machicao y Priemer et al. (2022).

The analysis we propose here also has consequences for the standard view about which linguistic variables are visible to sociolinguistic evaluation. Since topic drop is licensed by a phrasal construction, our results imply that speakers' evaluations are sensitive to abstract syntactic variables (see also Bender 2007, Robinson 2022). This is a departure from some sociolinguistic literature (Labov 2001, Meyerhoff & Walker 2013, Eckert & Labov 2017, i.a.), which claims that only phonological or lexical variables can be socially monitored. Our analysis is also incompatible with a model of grammar that only attaches SMs and other CI-meanings to vocabulary items or surface realizational patterns in PF (Adger 2006, Saab 2021). Rather, we require a more flexible architecture that can represent the social information speakers indexically associate with any linguistic structure.

HPSG is especially well-suited for this, given the fact that all types of linguistic information are modeled with a single unified formalism. Therefore, it can naturally express the fact that any linguistic unit (words, unary lexical projections, phrases and even PHON strings) can be associated with a SM. Furthermore, HPSG allows SMs to be arranged in a sortal hierarchy, which makes it possible to capture the fact that variants are often underspecified with respect to the conditions they can be felicitously used in.

References

Adger, David. 2006. Combinatorial variability. *Journal of Linguistics* 42(3). 503–530. DOI: 10.1017/s002222670600418x.

Almeida, Wellington Couto de. 2023. Não chame de erro o que a linguística chama variação: processamento de variação linguística e de agramaticalidade no âmbito da concordância verbal variável. *RELIN* 31(2). 510–550.

Asadpour, Hiwa, Shene Hassan & Manfred Sailer. 2022. Non-*wh* relatives in English and Kurdish: Constraints on grammar and use. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 29th International Conference on Head-Driven Phrase Structure Grammar, Nagoya University & National Institute for Japanese Language and Linguistics*, 6–26. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2022.1.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). DOI: 10.1016/j.jml.2012.11.001.

Beltrama, Andrea. 2018. Precision and speaker qualities: The social meaning of pragmatic detail. *Linguistics Vanguard* 4(1). 20180003. DOI: 10.1515/lingvan-2018-0003.

Beltrama, Andrea. 2020. Social meaning in semantics and pragmatics. *Language and Linguistics Compass* 14(9). 1–20. DOI: 10.1111/lnc3.12398.

Bender, Emily M. 2001. *Syntactic variation and linguistic competence: The case of AAVE copula abscence*. Stanford, CA: Stanford University. (Doctoral dissertation).

Bender, Emily M. 2005. On the boundaries of linguistic competence: Matched-guise experiments as evidence of knowledge of grammar. *Lingua* 115(11). 1579–1598. DOI: 10.1016/j.lingua.2004.07.005.

Bender, Emily M. 2007. Socially meaningful syntactic variation in Sign-Based Grammar. *English Language & Linguistics* 11(2). 347–381. DOI: 10.1017/S1360674307002286.

Burnett, Heather. 2017. Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of Sociolinguistics* 21(2). 238–271. DOI: 10.1111/josl.12229.

Burnett, Heather. 2019. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy* 42(5). 419–450.

Campbell-Kibler, Kathryn. 2007. Accent, (ING), and the social logic of listener perceptions. *American Speech* 82(1). 32–64. DOI: 10.1215/00031283-2007-002.

Cardinaletti, Anna. 1990. Subject/object asymmetries in German null-topic constructions and the status of specCP. In Joan Mascaró & Marina Nespor (eds.), *Grammar in progress*, 75–84. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110867848.75.

Christensen, Rune Haubo Bojesen. 2022. *ordinal: Regression models for ordinal data*. R package version 2022.11-16. `https://cran.r-project.org/package=ordinal` (11 October, 2024).

Copestake, Ann, Dan Flickinger, Carl J. Pollard & Ivan A. Sag. 2005. Minimal recursion semantics: An introduction. *Research on Language and Computation* 3(4). 281–332.

Drach, Erich. 1937. *Grundgedanken der deutschen Satzlehre*. 1st edn. Frankfurt am Main: Verlag Moritz Diesterweg.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453–476. DOI: 10.1111/j.1467-9841.2008.00374.x.

Eckert, Penelope & William Labov. 2017. Phonetics, phonology and social meaning. *Journal of Sociolinguistics* 21(4). 467–496. DOI: 10.1111/josl.12244.

Frick, Karina. 2017. *Elliptische Strukturen in SMS: Eine korpusbasierte Untersuchung des Schweizerdeutschen* (Empirische Linguistik / Empirical Linguistics 7). Berlin: De Gruyter. DOI: 10.1515/9783110517859.

Fries, Norbert. 1988. Über das Null-Topik im Deutschen. *Sprache und Pragmatik* 3. 19–49.

Green, Georgia M. 1994. The structure of context: The representation of pragmatic restrictions in HPSG. In James H. Yoon (ed.), *Proceedings of the 5th Annual Conference of the Formal Linguistics Society of Mid-America*, vol. 1/2 (Studies in the Linguistic Sciences), 215–232. Urbana, IL: Department of Linguistics, University of Illinois.

Heim, Irene. 1991. Artikel und Definitheit. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, 487–534. Berlin: De Gruyter.

Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15(4). 531–574.

Kaur, Gurmeet & Akitaka Yamada. 2022. Honorific (mis)matches in allocutive languages with a focus on Japanese. *Glossa: A Journal of General Linguistics* 44(1). DOI: 10.16995/glossa.7669.

Labov, William. 2001. *Principles of linguistic change: social factors*, vol. 2. Malden, MA: Blackwell.

Lambert, Wallace E, Howard Giles & Omer Picard. 1975. Language attitudes in a French-American community. *Linguistics* 13(158). 127–152. DOI: 10.1515/ling.1975.13.158.127.

Machicao y Priemer, Antonio. 2022. Konstituententest. In Stefan Schierholz & Pál Uzonyi (eds.), *Grammatik: Syntax*, 482–487. Berlin: De Gruyter. DOI: 10.1515/9783110698527.

Machicao y Priemer, Antonio, Stefan Müller, Roland Schäfer & Felix Bildhauer. 2022. Towards a treatment of register phenomena in HPSG. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 29th International Conference on Head-Driven Phrase Structure Grammar, Nagoya*

University & National Institute for Japanese Language and Linguistics. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2022.5.

McCready, Elin. 2019. *The semantics and pragmatics of honorification: Register and social meaning* (Oxford Studies in Semantics and Pragmatics 11). Oxford: Oxford University Press.

Meyerhoff, Miriam & James A. Walker. 2013. An existential problem: The sociolinguistic monitor and variation in existential constructions on Bequia. *Language in Society* 42(4). 407–428.

Müller, Stefan. 2014. Elliptical constructions, multiple frontings, and surface-based syntax. In Paola Monachesi, Gerhard Jäger, Gerald Penn & Shuly Wintner (eds.), *Proceedings of Formal Grammar 2004, Nancy*, 91–109. Stanford, CA: CSLI Publications.

Münster, Katja & Pia Knoeferle. 2018. Extending situated language comprehension (accounts) with speaker and comprehender characteristics: toward socially situated interpretation. *Frontiers in psychology* 8. 1–12.

Oshima, David Y. 2021. When (not) to use the Japanese particle *wa*: groundhood, contrastive topics, and grammatical functions. *Language* 97(4). 320–340.

Oushiro, Livia. 2019. A computational approach for modeling the indexical field. *Revista de Estudos Linguagem* 27(4). 1737–1786. DOI: 10.17851/2237-2083.0.0.1737-1786.

Paggio, Patrizia. 2009. The information structure of Danish grammar constructions. *Nordic Journal of Linguistics* 32(1). 137–164. DOI: 10.1017/S0332586509002066.

Paolillo, John C. 2000. Formalizing formality: An analysis of register variation in Sinhala. *Journal of Linguistics* 36(2). 215–259. DOI: 10.1017/S0022226700008148.

Plesca, Ana-Maria, Katja Maquate & Pia Knoeferle. 2024. Register and standard language knowledge in interaction. *Register Aspects of Language in Situation* 3(2). 1–35. DOI: 10.18452/28042.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press & CSLI Publications.

Potts, Christopher. 2007. The expressive dimension. *Theoretical Linguistics* 33(2). 165–198. DOI: 10.1515/TL.2007.011.

Potts, Christopher & Shigeto Kawahara. 2004. Japanese honorifics as emotive definite descriptions. In Robert B. Young (ed.), *Proceedings of the 14th Semantics and Linguistic Theory Conference (SALT)*, vol. 14, 253–270. Northwestern University: The Linguistic Society of America. DOI: 10.3765/salt.v14i0.2917.

R Core Team. 2023. *R: A Language and Environment for statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. `https://www.R-project.org` (29 July, 2024).

Robinson, Mary. 2022. *Negative concord as a window into social perception of morphological and syntactic variables*. New York, NY: NYU. (Doctoral dissertation).

Ross, John R. 1982. *Pronoun deleting processes in German*. Paper presented at the annual meeting of the Linguistic Society of America. San Diego, California.

Saab, Andrés. 2021. On the locus of expressivity: Deriving parallel meaning dimensions from architectural considerations. In Eleonora Orlando & Andrés Saab (eds.), *Slurs and expressivity: Semantics and beyond*, 17–44. Lanham, MA: Lexington Books.

Salmon, William. 2022. Social markers and dimensions of meaning. *Journal of Pragmatics* 192. 98–115. DOI: 10.1016/j.pragma.2022.02.014.

Schäfer, Lisa. 2021. Topic drop in German: Empirical support for an information-theoretic account to a long-known omission phenomenon. *Zeitschrift für Sprachwissenschaft* 40(2). 161–197. DOI: 10.1515/zfs-2021-2024.

Singh, Raj. 2011. Maximize presupposition! and local contexts. *Natural Language Semantics* 19. 149–168.

Smith, E. Allyn, Kathleen Currie Hall & Benjamin Munson. 2010. Bringing semantics to sociophonetics: social variables and secondary entailments. *Laboratory Phonology* 1(1). 121–155.

Taniguchi, Ai. 2019. Social meaning at the semantics-sociolinguistics interface. Unpublished Manuscript.

Wilcock, Graham. 1999. Lexicalization of context. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and constructional aspects of linguistic explanation*, 373–387. Stanford, CA: CSLI Publications.

Wöllstein, Angelika. 2010. *Topologisches Satzmodell* (Kurze Einführungen in die germanistische Linguistik 8). Heidelberg: Winter.