Non-*wh* relatives in English and Kurdish: Constraints on grammar and use

Hiwa Asadpour
Goethe-University Frankfurt, JSPS International Fellow Tokyo University

Shene Hassan
Rackow Schule Frankfurt

Manfred Sailer
Goethe-University Frankfurt a.M.

Proceedings of the 29th International Conference on
Head-Driven Phrase Structure Grammar

Online (Nagoya/Tokyo)

Stefan Müller, Elodie Winckel (Editors)

2022

Frankfurt/Main: University Library

pages 6–26

Keywords: implicature, prescriptive language, non-restrictive "that" relative, non-standard language, register, social meaning, Sorani Kurdish, Mukri, Selemani

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.

Abstract

The paper looks at constraints on non-*wh* relatives in Sōrānī Kurdish (Iranian) and English (Germanic). We argue that some of them are grammatical, whereas others introduce social meaning. We present a basic, lexicalist syntactic analysis and expand it with social meaning constraints. We propose that classical sociolinguistic variables have the status of conventional implicatures and the overall assessment of a style is treated as a particularized conversational implicature.

1 Introduction

Sōrānī Kurdish (Iranian) has two formal types of relative clauses: bare relatives and relatives introduced by a relativizer, *ka*. This situation is analogous to what we find with non-*wh* relatives in English (Germanic), see (1).[1]

(1) a. Ali kətebakay (ka) Rezān nūsīwyetī deyxwenetawa
Ali book.DEF.EZ (that) Rezān wrote.3SG read.3SG
b. Ali read the book (that) Rezān wrote.

In both languages, the variation between the two types is subject to grammatical constraints, but also to regional variation, register variation, and prescriptive constraints, i.e., the choice between them carries *social meaning*.

We will first describe the situation in Sōrānī Kurdish (Section 2) and then look at the English data in Section 3. We will briefly look at previous HPSG work on relative clauses and sketch our own syntactic analysis in Section 4. Similarly, Section 5 contains a short review of previous work on social meaning in HPSG, followed by our own proposal. We end with a conclusion in Section 6.

2 Sōrānī relative clauses

Sōrānī is also known as Central Kurdish (MacKenzie, 1961). We will look at two varieties of Sōrānī: Mukrī and Səlemānī. Mukrī is a regional minority language in Northwest Iran, and, as such, in contact with Persian (Iranian) as superstrate language (Asadpour, 2021). Səlemānī is a regional majority language in Iraqi-Kurdistan and in contact with Iraqi Arabic (Semitic).

Sōrānī has no *wh* relatives, but bare relatives and relatives introduced by a non-inflecting particle *ka* (regionally: *ke*/*we*). Hassan (2021, Chapter 7) shows

[†]We would like to thank the reviewers and the audience for their comments, in particular Bob Borsley, Antonio Machicao y Priemer, and David Y. Oshima. Linda Kremer and Pascal Hohmann-Huet were of tremendous help with getting access to the literature. All errors are ours. Hiwa Asadpour was supported by the DAAD-JSPS grant PE21779 during his work on this paper.

[1]We follow the Leipzig glossing rules (https://www.eva.mpg.de/lingua/resources/glossing-rules.php). The following abbreviations appear in our glossings: COP copula, DEF definite, EZ Ezafe, IPFV imperfective, PL plural, PRS present tense, PVB preverb, SG singular.

that there is a striking similarity between Sōrānī and English concerning the alternation of embedded bare and *ka/that* clauses. As indicated above in (1), both *ka* and *that* are optional with restrictive relative clauses. Example (2) shows that both function words are also possible for declarative complement clauses. In (3) we see that the bare form is excluded in non-restrictive relatives. Finally, as indicated in (4), only the *ka/that*-marked form can be extraposed.

(2) Ali bīr dakā (ka) Rezān bərduyatyawa
Ali think does (that) Rezān won.3SG
'Ali thinks (that) Rezan won.'

(3) Ānnā, *(ka) kəč=ī mən=a, lera=ya
Anna (that) daughter=EZ I=is here=is
'Anna, *(who) is my daughter, is here.'

(4) šuša-ka šəkā *(ka) to kərību=t bo=m.
glass-DEF broke.3SG (that) you bought=2SG for=1SG
'The bottle broke *(that) you bought for me.'

Kim (2010) claims that *ka* is preferred when the relativized element is the local subject of the relative clause. However, Saady (2020) and Hassan (2021) find examples like (5) natural.

(5) aw šofer=a=y (ka) ba hewāšī otōmbel
DEM driver=DEM=EZ (that) with slowly automobile
le-da-xuř-et səłāmat=a.
PVB-IPFV-drive-3.SG safe=COP.PRS.3.SG
'The driver who drives cars slowly is safe.' (Saady (2020, 114), transliteration and glosses adapted from Hassan (2021, 216))

Hiwa Asadpour, a native speaker of Mukrī, conducted two informal interviews with Mukrī and Səlemānī speakers to explore the acceptability of bare or *ka*-marked restrictive relatives. He conducted the first interview with 40 Mukrī speakers and 20 Səlemānī speakers. He wrote down the relative clauses that they produced and asked them spontaneously whether they considered the bare form acceptable in formal and/or colloquial contexts. The results are given in Table 1. Səlemānī speakers generally reject bare relatives in formal situations, while Mukrī speakers are less categorical. Mukrī speakers accept bare relatives in colloquial situations. While Səlemānī speakers largely reject bare relatives in colloquial situations, their acceptance increases compared to formal situations.

For the second interview, Hiwa Asadpour isolated some of the spontaneously produced restrictive relative clauses from the first interviews and discussed them explicitly with ten Mukrī and ten Səlemānī informants. As before, all informants considered *ka* relatives adequate in both formal and colloquial settings. Their assessment of bare relatives is given in Table 2. The second interview

	formal		colloquial	
	Mukrī	Səlemānī	Mukrī	Səlemānī
bare?	($N = 40$)	($N = 20$)	($N = 40$)	($N = 20$)
✓	3 (7.5%)	0 (0%)	39 (97.5%)	5 (25%)
✗	37 (92.5%)	20 (100%)	1 (2.5%)	15 (75%)

Table 1: Acceptability of bare relatives, assessed in passim

	formal		colloquial	
	Mukrī	Səlemānī	Mukrī	Səlemānī
bare?	($N = 10$)	($N = 10$)	($N = 10$)	($N = 10$)
✓	3 (30%)	0 (0%)	5 (50%)	3 (30%)
✗	7 (70%)	10 (100%)	5 (50%)	7 (70%)

Table 2: Acceptability of bare relatives in explicit metalinguistic discussion

confirms the results of the first: A clear majority of our Mukrī informants prefer *ka* relatives in formal situations, but half of them are willing to accept bare relatives in colloquial situations. The Səlemānī speakers categorically rejected bare relatives in formal situations, but are more accepting in colloquial situations.

It should be noted that the results of these interviews should not be taken to mean that Səlemānī speakers generally reject bare relatives. Rather, the explicit metalinguistic nature of the interview can be taken as an indication that the judgments were given in the light of a prescriptive perspective. The fact that the researcher himself is a Mukrī speaker might have influenced Səlemānī speakers to give even more prescriptively influenced assessments.

Hiwa Asadpour's small scale informal interviews show that *ka* relatives are considered the prescriptively preferred form in both varieties, and that the use of bare relatives signals a colloquial way of speaking. Səlemānī speakers seem to have an additional constraint that bare relatives should be avoided in situations in which "proper," "correct" language use is considered adequate.

3 English bare and *that* relatives

English has *wh* relatives in addition to bare and *that* relatives. However, only *wh* relatives allow for complex relative constituents, see (6).[2]

[2]The only exception seems to be a possessive use of *that* as in "*the pencil* [*that's lead is broken*]" (Sag, 1997, 463). We lack precise data on the distribution of possessive relative *that's*. We tentatively propose here that it is a separate lexical element that is a determiner and marked as highly non-standard.

(6) the student [to whom/*to that I talked] ...

English *that* relatives can occur with a wide variety of antecedents and, in most constellations, there is free variation between the bare and the *that* form. As in Sōrānī, bare relatives are excluded in extraposed position, see example (4), and English also categorically excludes bare non-restrictive relatives, see Fabb (1990, 72) or Huddleston & Pullum (2002, 1056). Based on these similarities, Hassan (2021) pursues a parallel analysis for Sōrānī relatives and English non-*wh* relatives. Another similarity between the two languages is that the bare form is usually associated with more casual, informal speech situations and simpler/shorter sentences – see, for example, Finegan & Biber (1994) or Huddleston & Pullum (2002, 1056). Finegan & Biber (1994) characterize the bare form as *implicit*, whereas the *that* form is called *explicit* as it overtly marks a clause boundary. They argue that implicit forms are typical for colloquial, spontaneous, and spoken language use.

Many formal descriptions implement two additional restrictions on English relatives, which are absent from Sōrānī. First, *that* relatives are banned from non-restrictive uses, for example in the analysis in Arnold (2004). Second, bare relatives are excluded when the relativized element is the local subject of the relative clause – see, for example the analysis in Pollard & Sag (1994). We will argue that in both cases, we are dealing with extra-grammatical constraints, i.e., that the "banned" cases exist, but are associated with a social meaning that is not compatible with the variety often assumed as the basis for formal linguistic studies. We will start by looking at non-restrictive relatives.

The status of non-restrictive *that* relatives is not fully clear in the literature. Quirk et al. (1972, 872) say they occur "occasionally," providing example (7).

(7) I looked at Mary's sad face, that I had once so passionately admired.

Additional authentic instances of non-restrictive *that* relatives are given by Huddleston & Pullum (2002, 1052), Carey (2013), and Hassan (2021, Section 6.2). Hassan (2021, Section 6.3) presents a small-scale questionnaire study on the naturalness of some of these cases, such as example (8). It was judged as "natural" by the majority of her informants (Hassan, 2021, 181). Sentence (8) contains a restrictive and a non-restrictive *that* relative. We mark the latter in (8). This marking was, of course, not in the questionnaires. Taken together, these various sources confirm that non-restrictive *that* relatives are part of the English relative clause system.

(8) The big topic this week was this video that Mitt Romney uploaded on YouTube, [**that**, according to reliable sources, had been filmed during a private party] ... (COCA, Davies (2008–2017))

Nonetheless, a description of English relative clauses needs to capture the fact that speakers (and linguists) tend to exclude and/or avoid non-restrictive *that* relatives. For example, Biber (2010, 616) found no instance of them

in British and American news corpora. Before doing her questionnaire study, Shene Hassan posted some of her example sentences in a facebook group to get acceptability intuitions from native speakers. The majority of answers included very explicit prescriptive comments, advising her strongly against the use of such constructions. This shows that non-restrictive *that* relatives are seen as indications of a non-prescriptive language use.

Let us turn to the second constraint mentioned above. While excluded prescriptively, bare subject relative clauses appear to be grammatical in some colloquial varieties. Huddleston & Pullum (2002, 1055) provide the examples in (9), though they indicate them as not generally acceptable, and they consider (9c) as more non-standard than the other two examples. Pollard & Sag (1994, 222: fn. 6, and 350: fn. 7) and Arnold & Godard (2021, 632: fn. 45) acknowledge the existence of bare relatives with local subject gaps in some varieties of English and propose different constraints for those varieties.

(9) a. ? It was my father [did most of the talking].
b. ? There is someone at the door [wants to talk to you].
c. ?? Anyone [wants this] can have it.

In this paper, we will assume that, just like Sōrānī, the grammar of English allows for bare relatives with relativized local subjects in principle. However, such cases carry a strong social meaning – as highly colloquial and, probably in addition, as regional. This social meaning makes them inappropriate for situations requiring a more general form of English.

To sum up, bare relatives are excluded from extraposed uses and from non-restrictive uses. In other cases, they are considered an "implicit" form compared to *that* relatives. Bare relatives with local subject gaps are strongly marked as specific to a variety that is not generally used. The grammar allows for *that* relatives throughout, but non-restrictive uses are prescriptively banned.

4 The syntax of non-*wh* relatives

In this short paper, we cannot possibly do justice to the rich research on relative clauses in HPSG, let alone in formal grammar. We can merely try to justify some of the analytic decisions that we have made and refer to Arnold & Godard (2021) for an overview of HPSG approaches to relative clauses and to Hassan (2021) for a more in-dept presentation of our analysis. Consequently, we will only provide some pointers to the previous literature in Section 4.1 and sketch our own syntactic analysis in Section 4.2.

4.1 Previous HPSG analyses of non-*wh* relatives

Taghvaipour (2004, 2005) analyzes restrictive relative clauses in Persian. Like Sōrānī, Persian does not have *wh* relatives, but the non-inflecting function words

ka in relative clauses. Taghvaipour treats *ka* as a relativizer that acts as the head of a relative clause. While Taghvaipour (2005) only considers relatives with an overt relativizer, there is a regional register variation with respect to the presence or absence of the relativizer (Majidi & Naghzguy-Kohan, 2020). It is unclear if Taghvaipour would have assumed an empty relativizer for bare relatives. We will follow Taghvaipour's basic analysis in our treatment of Sōrānī, but add a phonologically empty relativizer for bare relatives. We acknowledge that a constructional variant might be equally conceivable, as proposed for English in Sag (1997) or Hoffmann (2010).

Matters are more complex in English because as it has *wh* relatives in addition to bare and *that* relatives. Some basic research questions include: (i) Do all non-*wh* relatives pattern alike or are some more closely related to *wh* relatives? (ii) How uniform can or should an analysis be and which tools should be employed (empty heads, special phrasal constructions, ...)?

As to the first question, Hoffmann (2010, Sections 5.1, 5.2) shows experimentally that *that* relatives pattern with bare relatives rather than with *wh* relatives: While pied piping is excluded for both relative *who* and relative *that* (*the student to whom/ *to who/ *to that/ *to* ∅ *I talked*), the judgements for P+*that* and P+∅ are like those for ungrammatical sentences, whereas P+*who* sequences are judged significantly more acceptable. Hoffmann (2010, 250) proposes that pied piping is not a grammatical option for bare and *that* relatives, but that pied piping with relative *who* is just stylistically marked. This suggests that non-*wh* relatives form a natural class, contrasting with *wh* relatives.

The second question can, probably, not be answered on purely empirical grounds. Pollard & Sag (1994, Chapter 5) present an HPSG analysis that uses a number of different empty heads for English relatives. They distinguish three types of English non-*wh* relatives: bare relatives (which exclude bare relatives with relativized local subjects), *that* relatives with a relativized local subject and other *that* relatives. In their approach, *that* is a relative pronoun in local subject relatives, but a complementizer in other cases. This heterogeneous analysis is not very appealing (Arnold & Godard, 2021, 621: fn. 35). Sag (1997) and Hoffmann (2010) pursue a constructional approach – with all occurrences of relative *that* being analyzed as a pronoun by Sag (1997, Section 5.4), and as a relativizer in Hoffmann (2010, 251–252). In both constructional approaches, special constructions are postulated for relative clauses with relativized local subjects. This might, again, be considered an undesirable aspect of the analyses – in particular since Levine & Hukari (2006) show that there is no need for such a fundamental distinction between the extraction of a local subject and the extraction of other constituents of a clause.

Local subjects behave in a special way in English and other languages, also outside the domain of relative clauses or unbounded dependencies in general. In particular, the information on the subject of a verb is sometimes needed by an element selecting a fully saturated projection of that verb. Höhle (2019, 558–559) introduces a head feature SMOR (for "subject morphology") to identify

the local subject of a verb. Sag (2012, 84) uses the feature XARG ("external argument") for the same purpose. In Sag's (2012, 84) implementation, the XARG value of a word is *none* if the word's SUBJ list is empty and identical with the element on the word's SUBJ list otherwise. This feature can be used, for example, to relate the subject of a tag question and the subject of the main clause, see (10). In Sag (2012, 151), the two clauses have NPs with identical indices as their XARG value. In our analysis, we will use this independently motivated feature XARG to model the restriction on local subject relatives.

(10) [S: [S: There$_i$ are two possibilities], [S: aren't there$_i$/*they$_j$]]?

Before closing this subsection, we need to say a few words on the distinction between restrictive and non-restrictive relative clauses. Arnold (2004, 2007) argues that the two types of relative clauses do not differ fundamentally in their internal and external syntax. According to him, restrictive relatives modify any nominal category, and non-restrictive relatives can modify any saturated phrase (Arnold, 2004, 43). The main difference between the two is, instead, semantic. Arnold (2004, 43) introduces subtypes of relative clause signs to encode this: *intersective-semantics* for restrictive relative clauses, and *global-scope-semantics* for non-restrictive relative clauses.

In this subsection, we mentioned the analytic ideas that have influenced our analysis and their main competitors.

4.2 Analysis of bare and *ka/that* relative clauses

We can now present our analysis of Sōrānī and English non-*wh* relatives. As the main focus of this paper is on the interplay of grammar and constraints on social meaning, we do not strongly commit to particular aspects of the syntactic analysis, though we propose one that is compatible with the data and parallel for the two investigated languages. We largely follow Hassan (2021) in the syntactic analysis and will gloss over all details of the semantic analysis. Her theory also extends to *wh* relatives.

Fattah (1997) argues in favor of an analysis of Sōrānī *ka* as a complementizer/relativizer rather than a relative pronoun. Therefore, we follow Taghvaipour (2005) and assume a functional head for relative clauses, which can be phonologically empty or realized as *ka*. The similarities between English *that* relatives and their Sōrānī counterparts, together with Hoffmann's (2010) experimental data support a relativizer analysis for *that* as well.

As mentioned above, Taghvaipour (2005) does not discuss bare relatives in Persian. To keep the structure of bare and non-bare relatives maximally similar, we assume bare relatives to be introduced by a phonologically empty relativizer.

The lexical entry of the Sōrānī and English relativizer is sketched in Figure 1. It is either phonologically empty or has the PHON value *ka/that*. The relativizer modifies some constituent with which it shares the INDEX value, $\boxed{1}$. It selects a clause on its COMPS list. This clause contains a gap that has the index $\boxed{1}$ as

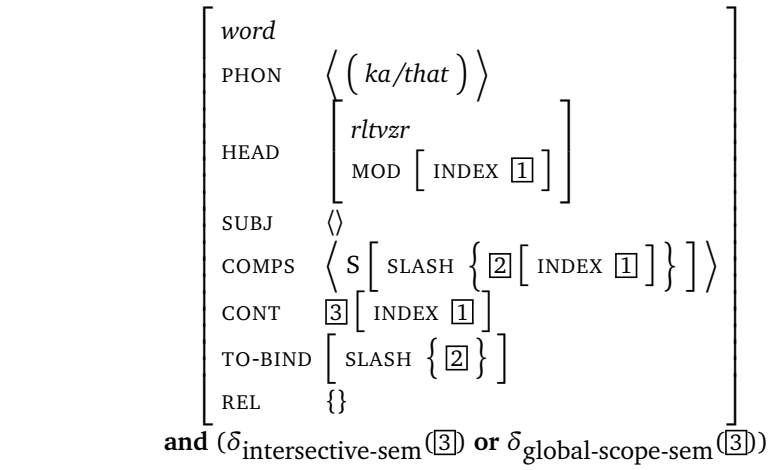


Figure 1: Lexical entry of the relativizer (Sōrānī and English)

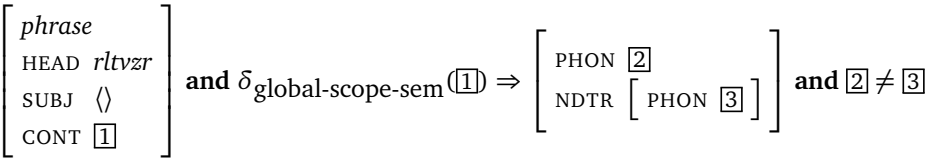


Figure 2: Ban on bare non-restrictive relatives (Sōrānī and English)

well.[3] The REL value of the relativizer is empty, as it does not contain a complex relative phrase. We add that the relativizer has a restrictive or a non-restrictive semantics. As we largely ignore the semantic analysis, we simply assume that there are descriptions $\delta_{\text{intersective-sem}}$ and $\delta_{\text{global-scope-sem}}$ that identify the type of content of the relativizer.[4]

The constraint in Figure 2 excludes an empty relativizer for non-restrictive relatives. It requires that, in a phrase that is headed by a relativizer with a non-restrictive semantics, the PHON value of the mother must not be identical with that of the nonhead daughter.[5]

This basic analysis allows for the full range of non-*wh* relatives discussed in Sections 2 and 3: Bare relatives are only allowed as restrictive relatives, but there is no constraint on the grammatical function of the relativized element.

[3] In Sōrānī the "gap" can take the form of a resumptive pronoun, see Fattah (1997, 254) and Hassan (2021, 220–225).

[4] See Hassan (2021, 249 and 263) for a concrete proposal for such descriptions.

[5] This constraint is compatible with an analysis of English *wh* relatives in which the fronted *wh* phrase is syntactically treated as a subject of the relativizer, as in Pollard & Sag (1994, 216).

5 Towards a modelling of constraints on social meaning

5.1 Previous HPSG approaches to social meaning

Pollard & Sag (1994) propose a basic treatment of context-dependent linguistic effects through their feature CONTEXT (CXT). This includes information on the participants of a discourse in the C-INDICES value, and a set of backgrounded propositions in the BACKGROUND (BGR) feature. In this architecture, only lexical elements introduce BGR elements. We will call this "lexical introduction." All such lexically introduced backgrounded propositions percolate to the highest node in a structure by the PRINCIPLE OF CONTEXTUAL CONSISTENCY. We will refer to this property as "global percolation." In the following, we will discuss these two properties and the general question of what information should be encoded as the social meaning contribution of a linguistic expression.

It is important to consult the sociolinguistic literature when addressing this general question. Current, third-wave, sociolinguistic research assumes that linguistic expressions are associated with some elements of social meaning and that the overall register or style assessment is a complex inference, influenced also by non-linguistic, contextual factors (Eckert, 2012). This means that we have a two-level system consisting first of individual linguistic properties, the classical sociolinguistic variables (Labov, 1984), and second of an overall categorization of a variety. Bender (2007) points out that the way in which third-wave sociolinguistic theory interprets such a system makes it very apt for an integration into a formal linguistic framework like HPSG: Speakers are seen as having (implicit) knowledge of the social meaning of individual sociolinguistic variables. They combine them in order to achieve a particular style in a given communicative situation.

Green (1994) proposes a model of speaker attitude and interlocutor relation within HPSG's CONTEXT value. Adopting the overall architecture of Pollard & Sag (1994), she encodes speaker attitudes as elements of the BGR set, i.e., as backgrounded propositions. For example, the word *dog* comes with a *mutual belief* among the speaker and the addressee that it is *normally believed* by members of the English speech community that the predicate **dog** is true for the INDEX value of the word.

While Green (1994) does not discuss social meaning as such, a similar, BGR-based system is used by Paolillo (2000) to model diglossia in Sinhala (Iranian). He explicitly distinguishes between the marking for a communicative attitude expressed by an individual lexical expression and the overall inference of a particular register (the High or Low variety, in the case of diglossia). The communicative attitudes express properties such as *edited*, *interactive*, *public*, and others. They are contributed by individual lexical signs in the form of a Green-style element into the BGR set. At the overall utterance level – and, more generally, at the text level – these attitudes are evaluated and lead to a highly context-specific assessment of the register. Paolillo describes this assessment as an implicature.

Paolillo (2000) encodes the social meaning of linguistic variables inside the grammar, but puts the level of the style evaluation outside grammar.

A different way of encoding is chosen in Wilcock (1999) and Bender (2007). They propose a feature REGISTER or SOCIAL respectively, whose value is atomic and expresses the result of the overall register or style assessment. Individual linguistic forms, i.e. the variants of sociolinguistic variables, constrain which REGISTER/SOCIAL values they are compatible with. For example, in Wilcock (1999), the relative pronoun *whom* comes with the REGISTER value *formal*. Similarly, Bender (2007) assigns the phonologically empty version of the copula in African American English a certain SOCIAL specification. Wilcock (1999, his (11)) introduces a REGISTER AMALGAMATION CONSTRAINT which states that a head and all its dependents have the same REGISTER value. This explicitly excludes combining elements with conflicting register specifications.

Machicao y Priemer et al. (2022) work with a register value as well, but it is non-atomic, containing attributes REGISTER1, REGISTER2, ...– one attribute for each register. The value of each of these is a factor that determines the likelihood that a sign is used in that register. The occurrence of elements that are indexical of a particular register will boost its factor and may have a negative effect on the factors for other registers. The authors derive the number of registers and the register factors associated with individual linguistic forms from corpora. While Machicao y Priemer et al.'s paper is not very explicit with respect to the percolation of register values, it is clear that their architecture is more flexible than Wilcock's and Bender's. If an utterance contains only elements that agree with respect to the register they are pointing to, the likelihood that it is used in that register will be very high. If we have conflicting elements, this could promote or lower various mutually incompatible register factors. In either case, the utterance would be well-formed but would show an unclear result for which register it is most likely to occur in. Our main concern with that approach is that it might not be able to model social meaning variation that is more fine-grained than what is comprised under their notion of register.

To summarize the previous HPSG literature on social meaning with respect to what is encoded as social meaning, we find two camps: one that encodes the meaning of individual sociolinguistic variables explicitly in the grammar, and one that encodes register/style explicitly. In our interpretation of third-wave sociolinguistic literature, we tend to side with the first group.

Let us now turn to the question of which elements can contribute social meaning. Green (1994) briefly addresses the potential problem of lexical introduction and expresses the hope that this would not be an obstacle as HPSG is a lexicalist framework. In the analysis of relative clauses proposed in Section 4.2, we propose a single lexical entry for the non-*wh* relativizer, be it phonologically empty or filled. However, we saw in Sections 2 and 3 that bare and *ka/that* relatives come with different speaker attitudes. We also saw that, in English, non-restrictive *that* relatives have a different social meaning than restrictive *that* relatives. Nonetheless, we do not need distinct lexical entries for these

and, therefore, a Green-style lexical analysis is problematic.

Wilcock (1999) and Bender (2007) take a different approach. As they assume that all signs in an utterance have the same REGISTER/SOCIAL value, that value can be constrained anywhere in the structure. In their examples, they always attach a constraint on the register value on a linguistic type. In other words, every linguistic element that contributes to social meaning needs to be associated with a node in the type hierarchy. Bender (2007, 269–370) addresses this aspect and explicitly defends what she calls *redundant types* as cognitively motivated. We will show in Section 5.2 that this is not necessary and that we can add social meaning constraints to the analysis from Section 4.2 without introducing new constructional types. Furthermore, we suspect that the type-based approach faces difficulties when looking at social meaning that is not associated with complete signs but rather with parts of them – such as social meaning of particular phoneme realizations (like r-fullness or r-lessness, Labov 1966) or with particular interpretation strategies (like negative concord, Labov 1969).

Machicao y Priemer et al. (2022) argue that constructions can change the register value. In their example (7), it looks as if there can only be one register-sensitive constraint on any given phrase. This is potentially problematic in the light of our analysis of English bare relatives above: We would need one constraint on all bare relatives to promote the register factor for colloquial registers. If the relative clause, in addition, has its local subject as relativized element, the factors for all formal, prescriptive registers need to be strongly demoted.

We conclude that constraints on social meaning are not restricted to lexical items. Existing HPSG approaches seem to be limited in their ability to cope with the flexibility of attaching social meaning to any aspect of a linguistic expression.

Let us finally look at the question of global percolation. Paolillo (2000) shows that global percolation is at the same time not enough and too much. First, he argues that style is not a property of individual sentences but rather of an entire discourse. He proposes a DISCOURSE COHERENCE principle according to which the communicative attitude information percolates beyond individual sentences. Second, his data contain examples of Low-variety use in a quotation, embedded into an otherwise High-variety discourse – such as direct speech embedded in novels. He suggests some sort of EXCEPTIONAL INHERITANCE for these cases, i.e., a situation in which the communicative attitude does not project outside of a quote or embedded speech report.

This overview leaves us with three desiderata and a big question: First, we lack a clear notion of how to relate the two levels of analysis, the social value of individual variables and the overall assessment of a register, style, or variety. Second, we lack a possibility to assign social meaning to linguistic expressions independently of the rest of the grammar. Third, we need to account for percolation of social meaning as going both beyond the utterance level and below it. The big question is if the mechanisms needed to model social meaning are new or are rather instances of already established notions.

5.2 Formulating social meaning constraints

We propose that social meaning can be adequately modeled as various types of implicatures as presented in Grice (1975). In particular, we assume that linguistic expressions can trigger *conventional implicatures* (CI) which express the social meaning associated with a sociolinguistic variable. We will show that CIs have exactly the percolation properties within an utterance needed for our purpose. The overall, context-dependent evaluation of the social meaning will be treated as a *particularized conversational implicature* (PCI).

The CIs expressing social meaning have the form proposed in Green (1994), see (11). X and Y are typically the speaker and the addressee, Z is a speech community as conceptualized by the attitude holder(s), E is any linguistic sign or part thereof, and ϕ is an arbitrary statement. In this paper, we will work with statements of the form "E signals colloquial use," or "E is incompatible with prescriptive use." However, other forms are possible such as some inference about the social relation between speaker and addressee (as with honorifics).[6]

(11) (X believes that) X and Y mutually believe that community Z normally believes that expression E signals ϕ.

Let us look at the projective properties of CIs. As noted in Grice (1975) and Potts (2005), CIs obligatorily project over negation and belief predicates. While is it often said that they project globally, Bach (1999) and others have pointed out that CIs don't necessarily project over embedded speech reports. Example (12) is taken from Bach (1999, 339). The word *but* triggers the CI that being huge and being agile are normally not compatible with each other. This contrast is inferrable in (12a), but not necessarily when the CI trigger is inside a speech report as in (12b). Bach (1999, 340: fn.19) shows that CIs do not project in direct quotes of individual words either.

(12) a. Shaq is huge but he is agile.
b. Marv said that Shaq is huge but he is agile.

We can show that social meaning has the same projective behavior. In English, the word *baba* 'bottle' comes with the social meaning that it indicates communication with a small child. Example (13a) shows that this word is inappropriate ("$") in inter-adult talk in simple affirmative or negated sentences, or when used under a belief predicate. However, when used as a direct quote, as in (13b), there is no inference that the speaker is addressing a small child.

(13) Two adults talking to one another:
a. $ (Alex believes that) Kim should (not) buy a new baba.
b. Kim should buy a new "baba."

[6]Thanks to Antonio Machicao y Priemer for suggesting the "X believes" part in (11). We will leave this implicit in the following, sticking closer to the formulation in Green (1994).

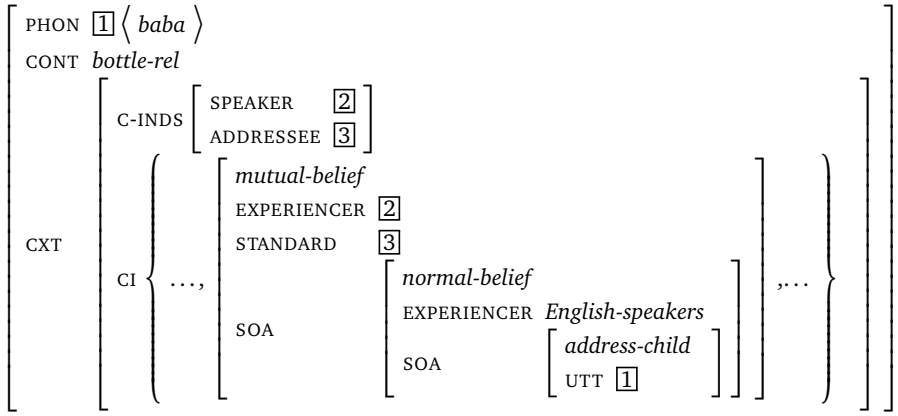


Figure 3: Lexical entry of the word *baba* 'bottle' including its social meaning

CIs have, of course, other properties as well. In particular, they are conventionally attached to a linguistic expression and, typically, express speaker-oriented side messages. These properties are also shared by the social meaning attached to linguistic expressions. The insight that the social meaning associated with a particular linguistic form has the status of a CI makes it unnecessary to postulate a special mechanism for the purpose of social meaning.

So far, there is little work on CIs in HPSG. We will simply follow the assumption in Sailer & Am-David (2016) that the BGR set of Pollard & Sag (1994) should be split into different sets, one for each type of projective meaning. In this paper, we assume a set-valued feature, CI, whose value contains all CIs attached to a sign. In (14) we formulate the CI PROJECTION PRINCIPLE. It determines that all CIs contributed by a phrase's daughters will project, unless they are retrieved, which can only happen in embedded speech constellations. The phrase can freely add more CIs.

(14) For each phrase, the CI value of the phrase is a superset of the union of the CI values of the daughters minus those that are integrated into the phrase's semantic representation.
CIs can be integrated into a semantic representation only in the scope of a speech act operator.

We can now look at an example encoding of the social meaning of the word *baba* as used in (13). The lexical entry of the word is sketched in Figure 3. Note that the element specified in the CI value has the form given in (11). It states that by using the word *baba* in the meaning of 'bottle', (the speaker believes that) speaker and addressee mutually believe that the English speech community normally believes that this word is used while talking to a child. This expresses an Eckert (2012)-style community belief of speaker/situation indexing as part of the linguistic competence of an individual. We think that this is in line with Bender's (2007) perspective of a formal integration of social meaning.

In this framework, we keep accumulating CIs with social meaning. Once we are at the level of the utterance, all of these CIs will be integrated into the utterance content, as proposed in Sailer (2021). The resulting enriched semantic representation will, then, be subject to evaluation of its discourse adequacy. In other words, it will be evaluated assuming Grice's COOPERATIVE PRINCIPLE through the application of *particularized conversational implicatures* (PCI). If a sign has several markers indicating colloquial speech and is used in a colloquial situation, then the Gricean maxims are met and no special PCI is inferred. If, however, the situation would require a more formal way of speaking, it will be assumed that the speaker is flouting maxims to achieve a particular effect.

Similarly, some registers seem to be mutually exclusive, such as the written formal and highly colloquial register, as also discussed in Paolillo (2000). Utterances with properties of both of them are not ungrammatical in our architecture. Instead, Gricean reasoning will be triggered to resolve this conflict by PCIs. However, the set of social meaning CIs can also be perceived as communicatively unresolvable, which then makes the utterance inappropriate.

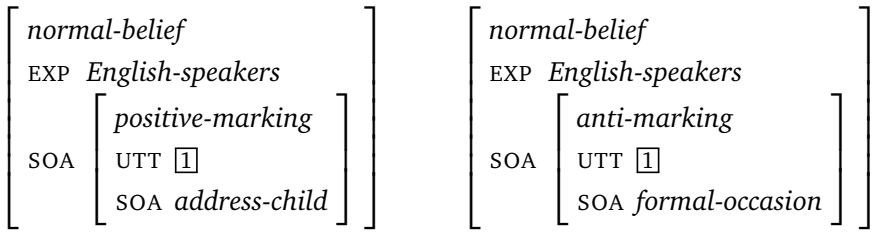
Example (15) illustrates the last situation. The sentence contains the word *baba* which triggers a CI of child-directed speech. At the same time, it contains a *wh* relative with a pied piped relative phrase containing *whom* and the rather technical term *dehydrated*. Unless we are in a very specific situation, sentence (15) is probably judged inadequate as there is no obvious way to imagine a situation in which all the social meaning CIs it contributes are satisfied.

(15) \$ The person to whom I passed the baba nearly dehydrated.

Our inclusion of PCIs into the picture of social meaning also captures the discourse effect observed in Paolillo (2000), i.e., the idea of projection beyond individual utterances: Cooperative speakers are expected to utter sentences that are in line with the properties of dialogue participants and situation.

In order to model the social meaning restrictions relevant for non-*wh* relatives, we need to address a further detail of what social meaning CIs can look like. Linguistic expressions can be used to signal adequacy for a particular social meaning aspect, but also incompatibility with it. For example, elements of child-directed speech such as the word *baba* might also be marked as incompatible with use in formal occasions.[7] In order to express this positive or negative marking, we include yet another layer into our social meaning CIs. For example, the *normal-belief* object from Figure 3 needs to be changed into the left AVM in Figure 4. There would also be a further element on the CI list which specifies the normal believe of English speakers of an *anti-marking* object. This object indicates that the utterance of the word is anti-marked for a state of affairs of type *formal-occasion*, see the right AVM in Figure 4.

[7]Some social meaning categories might be ordered along a Horn scale. Then, incompatibility inferences might have the status of *generalized conversational implicatures* (Grice 1975, Levinson 2000). Sailer (2021) proposes how this type of implicature can be integrated into the present architecture as well. We are grateful to David Oshima (p.c.) for discussing similar cases with us.



a. Marking for child-directed speech b. Anti-marking for formal situations

Figure 4: Positive and anti-marking for social meaning for *baba* 'bottle'

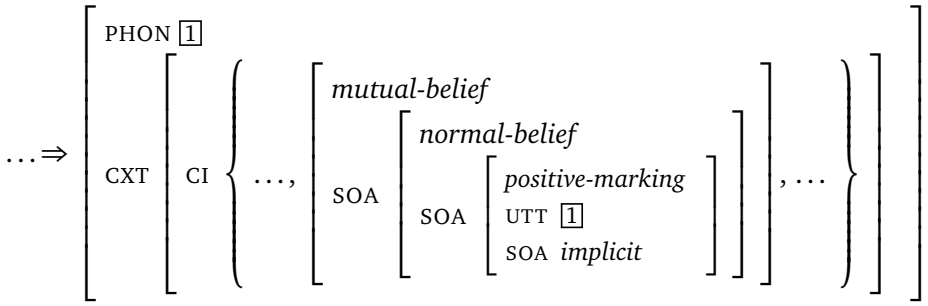


Figure 5: Constraint marking bare relatives as implicit (Sōrānī, English)

We can now formalize the social meaning constraints for Sōrānī and English non-*wh* relatives that we established in Sections 2 and 3. First, in both Sōrānī and English, bare relatives are marked as implicit forms and *ka/that* relatives as more explicit forms. We illustrate the constraint for bare relatives. A bare relative can be identified syntactically as a clause with HEAD value *rltvzr* whose PHON value is identical with that of its non-head daughter. This is the antecedent of the required constraint whose consequent we depict in Figure 5. The constraint for *ka/that* relatives will be analogous, just specifying non-identical PHON values in the antecedent and a positive marking for *explicit*.

In addition to these general constraints, bare relatives in Mukrī are also marked positively as *colloquial*. In Səlemānī, they are not only positively marked as *colloquial*, but also anti-marked for *prescriptive*. These constraints capture the data summarized in Section 2.

For English bare relatives, we want to formalize the constraint that they are a marker of a colloquial and non-prescriptive speech when used with a relativized local subject. To express this constraint, information on the local subject must be available at the clause level. We can use the feature XARG mentioned in Section 4.1 to identify a relative clause with a relativized local subject: Its lexical head's COMPS list contains a clause whose SLASH element is identical with its XARG value. In Figure 6 we only show the antecedent of this constraint. The consequent specifies two elements of the CI set, one with a positive marking for colloquial speech, one with an anti-marking for prescriptive speech.

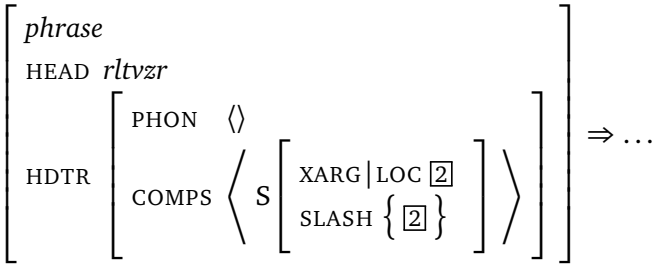


Figure 6: Antecedent of the constraint on English bare local subject relatives

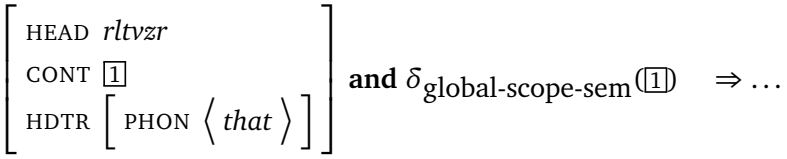


Figure 7: Antecedent of the constraint on English non-restrictive *that* relatives

Finally, we can turn to non-restrictive *that* relatives in English. A non-restrictive *that* relative can be identified as a phrase with HEAD value *rltvzr*, a head daughter with the phonology *that* and a global scope semantics, i.e., a CONTENT value that satisfies the description $\delta_{\text{global-scope-sem}}$. This antecedent is shown in Figure 7. The consequent of the constraint contains an anti-marking CI for prescriptive language use.

Before closing this section, a technical remark on the CI percolation mechanism should be made. In Section 5.1, we argued against the restriction to lexical introduction and against introducing new construction types for social meaning. In the principle in (14) we stated that the CI value of a phrase is a superset of the union of its daughters' CI sets (unless retrieved). We also always leave open whether there are elements on the CI set beyond the ones we specify in a constraint on phrases. This allows for phrases that contain elements on their CI set that are not inherited from the words they dominate, and yet we don't require explicit constructional types in the type hierarchy. For example, any non-restrictive *that* relative will not only have on its CI set the anti-marking for prescriptive from the constraint in Figure 7, but also the marking for explicit expression from the constraint mentioned in the discussion of Figure 5.

However, this also allows for additional background assumptions to be freely inserted anywhere in the structure. We propose to block this through the model theory of the grammar. The standard assumption in HPSG is that we consider all utterance-representing signs in a (minimal) exhaustive model of our grammar as constituting the described language (Richter, 2007, 2021). In such a model, we will have many signs representing the same utterance which are isomorphic except for their CI values. Among such signs we select only those that have a minimal number of elements in their CI value. This guarantees that register constraints that are enforced lexically or through constraints of the grammar always

appear, but randomly added ones are filtered out. Such a model-theoretic treatment seems justified as the CI value is assessed outside the grammar through PCIs. In other words, we are dealing with a phenomenon at the interface between grammar and the extra-linguistic interpretation of linguistic structures.[8]

6 Conclusion

We argued for a parallel treatment of the basic grammar of Sōrānī and English non-*wh* relatives. There are grammatical constraints – such as the ban on pied piping in non-*wh* relatives, and the ban on non-restrictive bare relatives. In addition, there are socially conditioned constraints: forms can be marked as signals of a particular register, but also as being incompatible with a certain register. We showed examples of either type of constraint.

While we restricted ourselves to non-*wh* relatives, English *wh* relatives can be included straightforwardly by allowing the relativizer to select the fronted relative constituent via its SUBJ value, analogously to the treatment in Pollard & Sag (1994). The constraint in Figure 2 predicts that *wh* relatives are compatible with the empty relativizer also in non-restrictive relatives.

Our formalization of social meaning as various types of implicatures seems to be in line with the two levels of analysis (variables and styles) of current variationist sociolinguistics. At the same time, it allows us to treat social meaning with formal tools that are needed independently.

References

Arnold, Doug. 2004. Non-restrictive relative clauses in construction based HPSG. In Stefan Müller (ed.), *Proceedings of HPSG 2004*, 27–47. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/5/arnold.pdf.

Arnold, Doug. 2007. Non-restrictive relative clauses are not orphans. *Journal of Linguistics* 43. 271–309.

Arnold, Doug & Danièle Godard. 2021. Relative clauses in HPSG. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 595–663. Berlin: Language Science Press. doi:10.5281/zenodo.5599844.

Asadpour, Hiwa. 2021. Cross-dialectal diversity in Mukrī Kurdish I: Phonological and phonetic variation. *Journal of Linguistic Geography* 9(1). 1–12. doi:10.1017/jlg.2021.1.

[8]Note that this solution is not directly compatible with the suggestion in Przepiórkowski (2021) to look at individual utterance-representing signs as models of the grammar, because we need to compare different utterance-representing signs.

Bach, Kent. 1999. The myth of conventional implicature. *Linguistics and Philosophy* 22(4). 327–366.

Bender, Emily M. 2007. Socially meaningful syntactic variation in sign-based grammar. *English Language and Linguistics* 2(11). 347–381. doi:10.1017/S1360674307002286.

Biber, Douglas (ed.). 2010. *Longman grammar of spoken and written English*. Harlow: Longman.

Carey, Stan. 2013. Non-restrictive 'that', that can be ambiguous. Blog entry from 2013 with more recent comments. https://stancarey.wordpress.com/2013/12/15/non-restrictive-that-that-can-be-ambiguous/.

Davies, Mark. 2008–2017. The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at english-corpora.org/coca.

Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of variation. *Annual Review of Anthropology* 41. 87–100. doi:10.1146/annurev-anthro-092611-145828.

Fabb, Nigel. 1990. The difference between English restrictive and nonrestrictive relative clauses. *Journal of Linguistics* 26. 57–78.

Fattah, Muhammad Maruf. 1997. *A generative grammar of Kurdish*: University of Amsterdam dissertation.

Finegan, Edward & Douglas Biber. 1994. Register and social dialect variation: An integrated approach. In Douglas Biber & Edward Finegan (eds.), *Sociolinguistic perspectives on register*, 315–347. Oxford: Oxford University Press.

Green, Georgia M. 1994. The structure of CONTEXT: The representation of pragmatic restrictions in HPSG. *Studies in the Linguistic Sciences* 24(1/2). 215–232.

Grice, Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics*, vol. 3, 41–58. New York: Academic Press.

Hassan, Shene. 2021. *Syntactic and semantic aspects of supplementary relative clauses in English and Sōrānī Kurdish*. Frankfurt a.M.: University Library. https://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/61371.

Hoffmann, Thomas. 2010. *Preposition placement in English: A usage-based approach*. Cambridge: CUP.

Höhle, Tilman N. 2019. Featuring creatures of darkness. In Stefan Müller, Marga Reis & Frank Richter (eds.), *Beiträge zur deutschen Grammatik. Gesammelte*

Schriften von Tilman N. Höhle, 553–569. Berlin: Language Science Press 2nd edn. doi:10.5281/zenodo.2588379. Previously unpublished article of a talk given 7.9.1994.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: CUP.

Kim, Danielle H. 2010. A basic guide to Kurdish grammar: A self-study reference and practice book. https://kurdishcentral.org/wp-content/uploads/2020/05/Kims_Kurdish_Grammar.pdf.

Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45(4). 715–762.

Labov, William. 1984. Field methods of the project on linguistic change and variation. In John Baugh & Joel Sherzer (eds.), *Language in use*, 28–53. Englewood Cliffs: Prentice-Hall.

Levine, Robert D. & Thomas E. Hukari. 2006. *The unity of unbounded dependency constructions*. Stanford: CSLI Publications.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge: MIT Press.

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 HPSG 2022*, 86–101. Frankfurt a.M.: University Library. doi:10.21248/hpsg.2022.5.

MacKenzie, David N. 1961. *Kurdish dialect studies*, vol. 1. Oxford: OUP.

Majidi, Shaho & Mehrdad Naghzguy-Kohan. 2020. A survey on relative clause construction in Kurmanji. *Journal of Westerrn Iranian Languages and Dialects* 8(4). 57–76.

Paolillo, John C. 2000. Formalizing formality: An analysis of register variation in Sinhala. *Journal of Linguistics* 36(2). 215–259. doi:10.1017/S0022226700008148.

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

Potts, Christopher. 2005. *The logic of conventional implicatures* (Oxford Studies in Theoretical Linguistics 7). Oxford: Oxford University Press.

Przepiórkowski, Adam. 2021. Three improvements to the HPSG model theory. In Stefan Müller & Nurit Melnik (eds.), *Proceedings of HPSG 2021*, 165–185. Frankfurt a.M.: University Library. doi:10.21248/hpsg.2021.9.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1972. *A grammar of contemporary English*. London: Longman.

Richter, Frank. 2007. Closer to the truth: A new model theory for HPSG. In James Rogers & Stephan Kepser (eds.), *Proceedings of the ESSLLI 2007 MTS@10 Workshop*, 101–110. Dublin: Trinity College Dublin.

Richter, Frank. 2021. Formal background. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 89–124. Berlin: Language Science Press. doi: 10.5281/zenodo.5599822.

Saady, Dilovan Sayfuddin. 2020. Reduced clauses in English and Kurdish. *Journal of Tikrit University for Humanities* 27(5). 106–118. doi:http://dx.doi.org/10.25130/jtuh.27.2020.7.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33. 431–483.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: An informal synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, 69–202. Stanford: CSLI Publications.

Sailer, Manfred. 2021. Minimizer negative polarity items in non-negative contexts. In Stefan Müller & Nurit Melnik (eds.), *Proceedings of HPSG 2021*, 348–368. Frankfurt a.M.: University Library. doi:10.21248/hpsg.2021.19.

Sailer, Manfred & Assif Am-David. 2016. Definite meaning and definite marking. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the HeadLex 2016*, 641–661. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/2016/headlex2016-sailer-am-david.pdf.

Taghvaipour, Mehran A. 2004. An HPSG analysis of Persian relative clauses. In Stefan Müller (ed.), *Proceedings of HPSG 2004*, 274–293. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/2004/taghvaipour.pdf.

Taghvaipour, Mehran A. 2005. *Persian relative clauses in Head-driven Phrase Structure Grammar*: University of Essex dissertation.

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: CSLI Publications.