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

Manfred Sailer Hiwa Asadpour Shene Hassan

Goethe-University Frankfurt/

Rackow Schule Frankfurt Goethe-University Frankfurt

JSPS Internation Fellow University of Tokyo

1 Introduction An intriguing property of relative clauses in English (Germanic) is not only the wide variety of different constructions (Sag, 1997), but also their sensitivity to semantic and register factors (Hoffmann, 2010). We will look at two special cases: bare relatives with local subject gaps and non-restrictive that relatives. Both constructions have been excluded in the formal modelling of English relatives. We will argue that including them leads to a more natural description of the inventory of English relative clauses. Cross-linguistic support for this comes from Sōrānī Kurdish (Iranian).¹ Sōrānī relatives are strikingly similar to English non-wh relatives. We will, therefore, propose a uniform analysis for the two languages. The basic syntactic and semantic analysis will be lexicalist in nature, but will be enhanced with phrasal, constellation-specific, register constraints.

2 English bare and that relatives English that relatives can occur with a wide variety of antecedents. The main restriction is that there cannot be pied piping or the formation of any complex relative constituent ("the student *[to that I talked]"). Bare relatives are compatible with mostly the same type of gaps inside the relative clause as that relatives. However, bare subject relative clauses are restricted to some, colloquial varieties – see the register comments in Huddleston & Pullum (2002, 1055) on the examples "?It was my father did most of the talking."/ "?Anyone wants this can have it" and also in Arnold & Godard (2021, fn 45). The main difference between bare and that relatives is semantic: bare relatives are categorically excluded from non-restrictive relative clauses, whereas that relatives can have such an interpretation, though it is marked and prescriptively frowned upon. Nonetheless Quirk at al (1972, 871) and Huddleston & Pullum (2002, 1052) provide examples, such as (1). Carey (2013), and corpus and acceptability data collected in Hassan (2021) further confirm that non-restrictive that relatives are part of the English relative system, see (2) for a corpus example that was also judged as natural by the majority of Hassan's (2021) informants.

3 **Sorānī relative clauses** Sorānī has no *wh* relatives, but bare relatives and relatives introduced by a noninflecting particle ka (regionally: ke/we). Hassan (2021) shows that there is a striking similarity between Sōrānī and English concerning the alternation of embedded bare and ka/that clauses: both are possible for declarative complement clauses, (3), and restrictive relative clauses, (4); only the ka/that-marked form occurs in non-restrictive relatives, (5). Finally, only the ka/that-marked form can be extraposed, (6). In his fieldwork, Hiwa Asadpour identified factors for the presence or absence of ka. While there is regional variation, morphological marking on the antecedent (so-called ezafe, Samvelian 2007), and factors similar to those for English bare relatives facilitate the absence of ka. On the other hand, bare relatives are generally disprefered in formal written context.

The strong similarities between Sorani and English non-wh relatives make a parallel basic analysis at least worth exploring. Under such a view, the difference between the two languages lies in additional constraints on the use rather than in the syntactic structures themselves: (i) Both languages categorically exclude bare non-restrictive relatives. (ii) English bare relatives with a relativized local subjects are strongly marked as colloquial. (iii) English, but not Sorānī, has a prescriptive ban on non-restrictive that relatives.

4 Previous HPSG analyses of relative clauses The analytic options include parameters like: Do that relatives pattern with wh relatives (i.e., is relative that a relative pronoun?) as in Sag (1997) or do they pattern with bare relatives (Huddleston & Pullum, 2002; Hoffmann, 2010)? Is there a (possibly phonologically empty) functional head mediating the syntactic and semantic connection between the relative clause and the antecedent (Pollard & Sag, 1994a) or is this achieved through a construction (Sag, 1997; Hoffmann, 2010)? Hoffmann (2010, §§ 5.1, 5.2) provides experimental evidence that shows that that relatives pattern with bare relatives rather than with *wh* relatives.

Arnold (2004, 2007) provides an analysis of non-restrictive relatives based on Sag (1997). The assumptions that relative *that* is a pronoun and that there are no non-restrictive *that* relatives force Arnold to change the REL value from being a set of *index* objects to a set of *npro* objects, as is proposed for the QUE value in Pollard & Sag (1994a). This would not have been necessary had Arnold (2004) included non-restrictive that

¹Also known as Central Kurdish (MacKenzie, 1961). We mainly refer to varieties of Mukri in Iran and Silemanī in Iraqi Kurdistan.

relatives. Instead, he could just require that all non-restrictive relatives be finite head-filler phrases (i.e. of the sort *fin-head-filler-phrase*).

Taghvaipour (2004, 2005) analyzes restrictive relatives in Persian. Like Sōrānī, Persian does not have *wh* relatives, but a relativizer *ka*. Taghvaipour treats this relativizer 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.

5 Previous HPSG/constructional approaches to register restrictions Green (1994) proposes a first model of speaker attitude and interlocutor relation within HPSG's CONTEXT value. Green encodes speaker attitudes and transferred references as elements of the BACKGROUND set (BGR), i.e., as backgrounded propositions. These have the status of presuppositions or, often, conventional implicatures. For example, the word *dog* comes with a *mutual believe* 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. Paolillo (2000) adapts this system to model diglossia in Sinhala (Iranian), and illutrates his system also with the register constraint on pied-piping in English *wh* relatives (Paolillo, 2000, 254).

Sag (1997) suggests to apply this technique to block pied piping of relative *who* ("*a student* *[*to that*]/*[*to who*]/[*to whom*] *I talked*"). This is worked out in Wilcock (1999). Wilcock introduces a feature REGISTER on context objects which has the values *informal* or *formal*. He assigns accusative *who* the REGISTER value *informal* and introduces a constructional constraint that a finite *wh* relative with a preposition phrase as its relative phrase must be marked as *formal*. While Wilcock shows how lexical and construction-specific register constraints can interact, his assumption of just two registers and of a uniform register marking on all expressions in an utterance are overly simplistic. Bender (2007) provides an in-depth motivation for modelling sociolinguistic variation in HPSG, but, eventually, uses a system with the same problems as Wilcock's.

Peterson (2016), based on *Diasystematic Construction Grammar* (Höder, 2012, 2018), sketches a more refined model of the co-existence of different varieties and the constraints on their uses within a constructional framework. The basic assumption in a diasystematic approach is that we can find instances of different grammar systems within a single utterance. Constructions in a multilectal speaker's grammar can be marked or unmarked for a particular language or variety. Example (7) shows the mixture of words from colloquial Lower German (in italics), from Standard German (underlined), and, the other words, belonging to both.

Peterson (2016, 140) moves from atomic marking of varieties or languages in multilectal individuals' grammars to a more refined system of features for speech-situational factors, based on Biber (1988). An utterance with conflicting contextual markings would, then, not be excluded – as in Wilcock's system – but simply contain potentially conflicting marking, which reduces the range of felicitous speech situations for that utterance. A situation that is not taken into account in the mentioned diasystematic approaches is exactly what we find for non-restrictive *that* relatives: A particular construction is prescriptively excluded. This means that in addition to marking a sign positively as belonging to a particular variety or register, it must be possible to mark it explicitly as not being part thereof.

6 Analysis of bare and *ka/that* **relative clauses** We develop a lexical analysis of English and Sōrānī non-*wh* relatives. Our main focus is on the interplay of grammar and register constraints. Therefore we do not strongly commit to a particular syntactic analysis, though we propose one that is compatible with the data, and parallel for the two investigated languages.² We adopt Arnold's (2004) semantic analysis and express the difference between restrictive and non-restrictive modifier semantics as two subtypes of *sign: intersective-semantics* and *global-scope-semantics* respectively. Syntactically, restrictive relatives modify any nominal category, and non-restrictive relatives can modify any saturated phrase (Arnold, 2004, 43).

We assume a functional head for relative clauses, which can be realized as ka/that – just as Taghvaipour (2005) for Persian. With the absence of *wh* relatives in Sōrānī, a pronominal analysis of *ka* is not plausible. The similarities between English *that* relatives and their Sōrānī counterparts, together with Hoffmann's (2010) empirical 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 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,

²For more details and an extension to wh relatives see Hassan (2021).

$$\begin{bmatrix} word \\ PHON & \left\langle \left(ka/that \right) \right\rangle \\ HEAD & \begin{bmatrix} rltvzr \\ MOD & [INDEX 1] \end{bmatrix} \\ SUBJ & \left\langle \right\rangle \\ COMPS & \left\langle S & [SLASH \left\{ 2 & [INDEX 1] \right\} \end{bmatrix} \right\rangle \\ CONT & [INDEX 1] \\ TO-BIND & [SLASH \left\{ 2 &] \\ REL & \left\{ \right\} \end{bmatrix}$$
 and (intersective-sem or global-scope-sem)

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

 $\begin{bmatrix} phrase \text{ and } global-scope-sem \\ HEAD \ rltvzr \\ SUBJ \ \langle \rangle \end{bmatrix} \Rightarrow \begin{bmatrix} PHON \ 1 \\ NDTR \ PHON \ 2 \end{bmatrix} \text{ and } 1 \neq 2$

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

1. It selects a clause on its COMPS list. This clause contains a gap that has the index 1 as well.³ As there is no pied piping in non-*wh* relatives, the REL value is empty. We add that the relativizer has a restrictive or a non-restrictive, i.e. global scope, semantics.

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.

7 Formulating holistic register constraints Schütze (2016) shows that register affects acceptability judgments, which indicates that register-awareness is clearly part of a speaker's linguistic competence. To model register competence, we combine the approaches mentioned in Section 5. We assume some classification of registers. We only need the registers *colloquial* and *prescriptive* here, though a more refined Peterson-style encoding is desirable. In line with Eckert (2012, 2019), we assume that individual linguistic expressions are associated with a speaker's register attitude. We will show that register attitude markings have the properties of *conventional implicatures* (CI), as formulated in Grice (1975) and Potts (2005): They encode a speaker attitude (on the appropriateness of an utterance for a particular register); they cannot be cancelled; they are conventional; and the are detachable, i.e., they only arise if a particular wording is chosen. As CIs, they express non-at issue side messages a speaker commits to.

In an utterance, all these "register side messages" are collected. The result will then be assessed as a *particular-ized conversational implicature*, in the sense of Grice (1975). For example, when a communicating individual consistently uses forms that contribute marking for a particular register, their utterance will be assessed as expressing this register. If there are markers of different registers within an utterance, this can usually be resolved as a special/individual style of speaking, in Eckert's sense. Paolillo (2000) characterized particular registers (such as *formal spoken*) as combinations of markers for individual aspects (for *formal spoken* these would be: *interactive*, *public*, *correct*, but not *edited*).

Some registers seem to be mutually exclusive, such as the written formal and highly colloquial register, as also discussed in Paolillo (2000). Utterances bearing properties of both of them are, then, not ungrammatical in our architecture but bear strong markings of different registers. This can be perceived as a communicatively unresolvable conflict, which makes the utterance inappropriate in any imaginable context. Alternatively, it can be construed a multilectal utterance, analogously to (7).

This architecture expresses Eckert's concept of individual style – and its HPSG-compatible rendering in Bender (2007): individual elements of an utterance can contribute different register-related markers (as conventional

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

implicatures). The overall combination of these, then, constitute the *style* of a speaker (which is inferred as a particularized conversational implicature). For our integration into HPSG, we subscribe to the semantics-pragmatics interface as used in Sailer (2021), based on Levinson (2000): conventional implicatures are part of the semantic representation of a linguistic sign, but flagged as speaker-oriented side messages. Particularized conversational implicatures, on the other hand, are not part of the grammar – just as Grice's *Cooperative Principle* is assumed to be a general principle of social interaction.

We can now formalize the constraint that a bare relative with local subject as relativized element is a marker of a colloquial register, see Figure 3. To express this constraint, information on the local subject must be available at the clause level. This is the case in Höhle (2019), through a feature SMOR, and in Sag (2012), through a feature XARG. 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. We can identify a relative clause with a relativized local subject because its lexical head's COMPS list contains a clause whose SLASH element is identical with its XARG value.

Instead of excluding such a constellation, we consider it a marker of colloquial speech. Using Green's (1994) feature geometry, we specify that there is a mutual believe (a *m*-believe relation) among speaker and addressee that it is normally believed (*n*-believe) among the relevant speech community that the phonology of a sign (locally available through the feature UTT) is connected with a some marking for a register, here the colloquial register. For the sake of simplicity, we leave open Green's specifications of whose mutual believe (typically the speaker and addressee) and whose normal believe (i.e. the relevant speech community).

As in Figure 3, register side messages are usually formulated as positively indicating what register an expression is an exponent of. However, we also find *incompatibility* or *taboo* marking, i.e. within a speech situation, a particular expression is considered incompatible with a certain register. Non-restrictive *that* relatives in English are a case in point. While attested and in use, there is a strong prescriptive constraint against them. This is formulated in Figure 4: Non-restrictive relatives headed by *that* are marked as non-acceptable (taboo) in prescriptive registers. We use the relation *register-taboo* for such a register-incompatibility marking. This contrasts with the relation *register-marking*, which is used to express a positive register marking.

In Sōrānī, bare relatives are subject to an analogous taboo constraint in formal writing. The constraint looks just like the one in Figure 3, but there is no need to further specify the properties of relativizer's complement.⁴ Let us close with a remark on the BGR percolation mechanism. Green (1994) and Paolillo (2000) only exemplify lexically introduced background assumptions. Wilcock (1999) and Bender (2007) can encode constructionally triggered register marking, but only because they work with a single register value shared by all signs within an utterance. We assume, in slight modification of Pollard & Sag's (1994a) principle, that the BGR value of a phrase is a superset of the union of its daughters' BGR sets. This allows for phrases to introduce new elements on their BGR set without requiring explicit constructional subtypes. For example, any utterance that contains a phrase that matches the antecedent of the constraint in Figure 4 will have a taboo marking for the prescriptive register in its BGR, though its daughters need not contain this background assumption.

This allows for additional background assumptions to be freely inserted anywhere in the structure. We propose to block this at 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 a huge number of signs representing the same utterance which are isomorphic except of their BGR values. Among such signs we select only those that have a minimal number of elements in their BGR value. This guarantees that register constraints that are enforced through constraints of the grammar always appear, but randomly added ones are filtered out. Such a model-theoretic treatment is well motivated as the BGR value is assessed outside the grammar through particularized conversational implicatures. In other words, the phenomena we are dealing with here are at the interface between grammar and the extra-linguistic interpretation of linguistic structures.⁵

8 Conclusion We argued for a parallel treatment of the basic grammar of English and Sōrānī non-*wh* relatives. There are grammatical constraints – such as the ban on pied piping in non-*wh* relatives, and the ban on bare non-restrictive relatives. In addition, there are register-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 register marking constraint.

⁴The same seems to hold for Persian according to Hiwa Asadpour's fieldwork.

⁵Note 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 utternance-representing signs.



Figure 3: Positive register constraint on bare local subject relatives (English)



Figure 4: Taboo constraint on non-restrictive that relatives in prescriptive register style (English)

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 (1994a). The constraint in Figure 2 predicts that *wh* relatives are compatible with the empty relativizer also in non-restrictive relatives We can formulate a further taboo constraint to ban the co-occurrence of a fronted *wh* phrase and *that*. Examples like (8) (from an interview with an American football coach) show that this might be a register constraint as well, not a strict grammatical constraint. The inclusion of register into our description of language has long been a desideratum in formal grammar. Our proposal shows how concepts from research on register as well as on multilingual and multilectal communication combines fruitfully with formal pragmatics to arrive at a leaner description of the grammatical system and, at the same time, at a robuster modelling of linguistic competence.

Examples

- (1) I looked at Mary's sad face, that I had once so passionately admired. (Quirk et al., 1972, 872)
- (2) 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)
- (3) Ali bīr dakā (ka) Rezān birduyatyawaAli think does (that) Rezān won.3sg 'Ali thinks (that) Rezan won.'
- (4) Al kitebakay (ka) Rezān nūsīwyetī deyxwenetawaAl book.DEF.EZ (that) Rezān wrote.3SG read.3SG 'Ali read the book (that) Rezān wrote.'
- (5) Ānnā, *(ka) kič=ī min=a, lera=ya Anna (that) daughter=3sG I=is here=is 'Anna, who is my daughter, is here.'
- (6) šuše-ke šika *(ka) to kirību=t bo=m. glass-DEF broke.3SG (that) you bought=2SG for=1SG
 'The bottle broke *(that) you bought for me.'
- (7) An *dat* Licht kann *de* <u>Hausmeister</u> nix *ännern*.
 at the light can the caretaker nothing change
 'The caretaker cant do anything about the light.' (Höder, 2012, 244)
- (8) He was a guy who that we absolutely had a major priority on ... (enTenTen20)

References Arnold, Doug. 2004. Non-restrictive relative clauses in construction based HPSG. In St. Müller (ed.), Proceedings of HPSG 2004, 27–47. CSLI. ♦ 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 St. Müller et al (ed.), Head-Driven Phrase Structure Grammar: The handbook, 595–663. LSP. ♦ Bender, Emily M. 2007. Socially meaningful syntactic variation in sign-based grammar. English Language and Linguistics 2(11). 347-381. doi: doi:10.1017/S1360674307002286. ♦ Biber, Douglas. 1988. Variation across speech and writing. CUP. ♦ Carey, Stan. 2013. Non-restrictive that, that can be ambiguous. Blog entry. URL https://stancarey.wordpress.com/ 2013/12/15/non-restrictive-that-that-can-be-ambiguous/. ♦ Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of variation. Annual Review of Anthropology 41. 87–100. ♦ Eckert, Penelope. 2019. The limits of meaning: Social indexicality, variation, and the cline of interiority. Language 95(4). 751–776. ♦ Fattah, Muhammad Maruf. 1997. A generative grammar of Kurdish: University of Amsterdam dissertation. \blacklozenge Green, Georgia M. 1994. The structure of CONTEXT. 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 Sorāni Kurdish. Frankfurt/Main: Goethe University dissertation. URL https: //publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/61371. Heta Höder, Steffen. 2012. Multilingual constructions: A diasystematic approach to common structures. In K. Braunmüller & Ch. Gabriel (eds.), Multilingual individuals and multilingual societies, 241–257. Benjamins. ♦ Höder, Steffen. 2018. Grammar is community-specific. In H. C. Boas & St. Höder (eds.), Constructions in contact, 37–70. Benjamins. ♦ Hoffmann, Thomas. 2010. Preposition placement in English. CUP. ♦ Höhle, Tilman N. 2019. Featuring creatures of darkness. In St. Müller et al (ed.), Beiträge zur deutschen Grammatik, 553–569. LSP 2nd edn. ♦ Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. CUP. ♦ Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, Massachusetts; London, England: MIT Press. ♦ MacKenzie, David N. 1961. Kurdish dialect studies, vol. 1. 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. • Peterson, John. 2016. Multilingualism, multilectalism and register variation in linguistic theory: Extending the diasystematic approach. In J. Fleischhauer et al (ed.), *Explorations of the syntax-semantics interface*, 109–147. DUP. \blacklozenge Pollard, Carl & Ivan A. Sag. 1994a. Head-Driven Phrase Structure Grammar. UCP. ♦ Pollard, Carl & Ivan A. Sag. 1994b. Head-Driven Phrase Structure Grammar. Chicago and London: University of Chicago Press. ♦ Potts, Christopher. 2005. The logic of conventional implicatures. OUP. • Przepiórkowski, Adam. 2021. Three improvements to the HPSG model theory. In St. Müller & N. Melnik (eds.), Proceedings of HPSG 2021, 165–185. ♦ Quirk, Randolph et al. 1972. A grammar of contemporary English. Longman. A Richter, Frank. 2007. Closer to the truth: A new model theory for HPSG. In J. Rogers & St. Kepser (eds.), Model-theoretic syntax at 10: Proceedings of the ESSLLI 2007 MTS@10 Workshop, 101–110. Dublin: Trinity College. ♦ Richter, Frank. 2021. Formal background. In St. Müller et al. (ed.), Head-Driven Phrase Structure Grammar: The handbook, 89–124. Berlin: LSP. • 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 H. C. Boas & I. A. Sag (eds.), Sign-Based Construction Grammar, 69–202. CSLI. ♦ Sailer, Manfred. 2021. Minimizer NPIs in nonnegative contexts. In St. Müller & N. Melnik (eds.), Proceedings of HPSG 2021, 348–368. Samvelian, Pollet. 2007. The Ezafe as a head-marking inflectional affix: Evidence from Persian and Kurmanji Kurdish. In S. Karimi et al (ed.), Aspects of Iranian linguistics, 339–361. Cambridge Scholars. ♦ Schütze, Carson T. 2016. The empirical base of linguis*tics: Grammaticality judgments and linguistic methodology.* LSP. Originally published 1996. ♦ Taghvaipour, Mehran A. 2004. An HPSG analysis of Persian relative clauses. In St. Müller (ed.), Proceedings of HPSG 2004, 274–293. ♦ Taghvaipour, Mehran A. 2005. Persian relative clauses in HPSG: Essex dissertation. ♦ Wilcock, Graham. 1999. Lexicalization of context. In G. Webelhuth et al (ed.), Lexical and constructional aspects of linguistic explanation, 373–387. CSLI.