Arabic Relative Clauses in HPSG[1]

Abdulrahman Alqurashi[a] and Robert D. Borsley[b]
[a]University of Essex and King Abdulaziz University
aaaalq@essex.ac.uk, aaalqurashi@kau.edu.sa
[b]University of Essex
rborsley@essex.ac.uk

Summary. In HPSG relative clauses have been analyzed in terms of phonologically empty heads in Pollard and Sag (1994) and in terms of a complex system of phrase types in Sag (1997). Modern Standard Arabic has a distinction between relatives with a definite antecedent, which are introduced by a special complementizer, and relatives with an indefinite antecedent, which are 'bare' clauses. An analysis assuming a complex system of phrase types faces a number of problems. An analysis in which relatives with an indefinite antecedent contain a phonologically empty complementizer is more satisfactory. Thus, in the case of Arabic, the approach of Pollard and Sag (1994) seems preferable to the approach of Sag (1997).

Keywords: Arabic, relative clauses, empty categories.

1. Background

Pollard and Sag (1994) develop an analysis of relative clauses employing a number phonologically empty heads. Sag (1997) rejects empty heads and instead makes use of a complex system of phrase types. Thus, for any language, major questions about relative clauses are: what phrase types are necessary? and are any empty heads necessary? In this paper we consider the implications of Modern Standard Arabic for these questions.

2. The basic data

Arabic has two main types of relative clauses. With a definite antecedent a relative clause consists of the element *ʔallaði* and a verb-initial clause containing either a gap or resumptive clitic, as in (1).

(1) a. jaaʔa l-walad-u llaði qaabala ___ l-malik-a
came.3.M.SG the-boy- NOM that. M.SG met.3.M.SG the-king-ACC
'The boy who met the king came,'

b. wajadtu l-kitab-a llaði tuħib-**hu**
found.1.SG the-book-ACC that. M.SG like.1.SG-3.M.SG
'I found the book that I like'

Here and subsequently we mark gaps by '___' and place resumptive clitics in bold.

With an indefinite antecedent *ʔallaði* does not appear. We just have a 'bare' verb-initial clause with either a gap or resumptive clitic, as in (2):

(2) a. jaaʔa walad-un qaabala ___ l-malik-a
came.3.M.SG a-boy-NOM met.3.M.SG the-king-ACC
'A boy who met the king came,'

[1] HPSG 2012 Conference/Ellipsis Workshop, July 18-21, 2012, Chungnam National University, Daejeon, Korea

b. wajadtu kitab-an tuħib-**hu**
found.1.SG book-ACC like.1.SG-3.M.SG
'I found a book that I like'

In definite relatives, *ʔallaði* is inflected for number, gender and case, making it look like a pronoun. However, its case is that of the antecedent and not that of the position relativized, as (3) shows.

(3) a. jaaʔa l-waladaani llaðaani qaabala-**humaa**
came.3.M.SG the-boy-DUAL.NOM that.M.DUAL.NOM met.3.M.SG-3.DUAL
l-malik-u
the-king-NOM
'The two boys whom the king met came.'
b. raʔaytu l-waladayni llaðayni qaabala-**humaa**
saw.1.SG the-boy-DUAL.ACC that.M.DUAL.ACC met.3.M.SG-3.DUAL
l-malik-u
the-king-NOM
'I saw the two boys whom the king met.'

Moreover it is never part of a larger clause initial phrase. Thus, for example, (4a) is ungrammatical. Instead we have (4b).

(4) a. *l-wallad-u [$_{PP}$ maʕ llaði] takallamta
the-boy-NOM with that talked.2.M.SG
'*The boy with that you talked.'
b. l-wallad-u llaði takallamta maʕ-**hu**
the-boy- NOM that talked.2.M.SG with-3.M.SG
'The boy that you talked with.'

Similarly, (5a) is ungrammatical and instead we have (5b).

(5) a. *ʔaʕrifu r-rajul-a [[ʔom llaði] maatat]]
know.1.M.SG the-man-ACC mother that.M.SG died.3.F.SG
'I know the man whose mother died.'
b. ʔaʕrifu r-rajul-a [llaði maatat ʔom-**hu**]
know. 1.M.SG the-man-ACC that.M.SG died.3.F.SG mother-3.M.SG
'I know the man whose mother died.'

Thus, *ʔallaði* is quite different from an interrogative pronoun, which can be part of a complex clause initial phrase, as the following show:

(6) [$_{PP}$ maʕa man] takallamta
with who talked.2.M.SG
'With whom did you talk?'

(7) [$_{NP}$ ʔom man] maatat
mother who died.3.F.SG
'Whose mother died?'

We conclude that *ʔallaði* is not a pronoun but a special inflected complementizer. It is confined to relative clauses, free relatives, and certain *wh*-questions and does not occur in complement clauses.

We will not try to deal with the distribution of gaps and resumptive clitics here. We note, however, that there is evidence from examples like the following with a gap in one conjunct and a resumptive clitic in the other that they are similar elements:

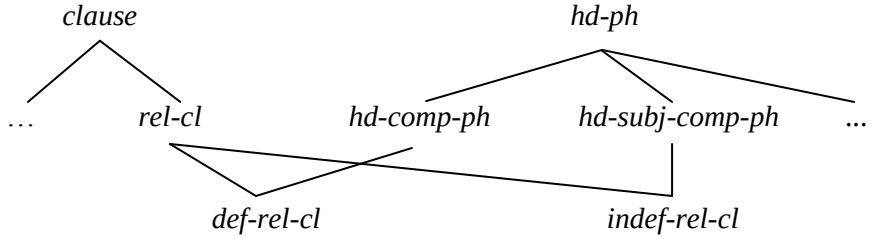
(8) l-fatatu llati ʔu_ibu ___ wa ʔa_ras ʕalay-**ha**
the-girl- NOM that- F.SG love.1.M.SG and care.1.M.SG about-3.F.SG
'The girl that I love and care about'

We will assume that both are realizations of SLASH and hence that there is no need to invoke a separate RESUMP feature as in Vaillette (2000).

3. An analysis

If we accept the conclusions reached above and also assume that indefinite relatives are bare clauses, we might propose an analysis with the following system of phrase types:

(9)



Among other things, this ensures that relatives with a definite antecedent are head-complement phrases and that relatives with an indefinite antecedent are head-subject-complement phrases (because they are verb-initial). Ignoring questions about CONTENT we might propose the following constraint on relative clauses.

(10) *rel-cl* $\Rightarrow$ $\begin{bmatrix} \text{HEAD}[\text{MOD NP}_1]] \\ \text{SLASH } \{\} \\ \text{HD - DTR}[\text{SLASH } \{\text{NP}_1\}] \end{bmatrix}$

This ensures that a relative clause is the top of an unbounded dependency and that it modifies an NP with the same index as the NP value of SLASH. The identical indices ensure that the NP and the gap or resumptive clitic agree in number and gender. The obvious constraints for *def-rel-cl* and *indef-rel-cl* are the following:

(11) a. *def-rel-cl* $\Rightarrow$ [MOD NP[DEF +]]
b. *indef-rel-cl* $\Rightarrow$ [MOD NP[DEF –]]

We also need appropriate lexical descriptions for forms of *ʔallaði* and verbs. The former will need to include the features in (12).

(12)

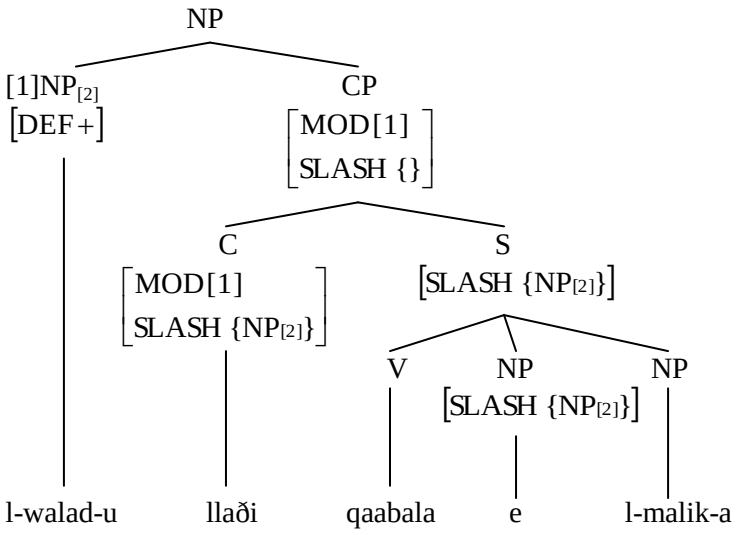
$\begin{bmatrix} \text{MOD NP}[\text{DEF+}] \\ \text{COMP S} < [\text{INV+}] > \end{bmatrix}$

The latter will need to be able to include the feature in (13).

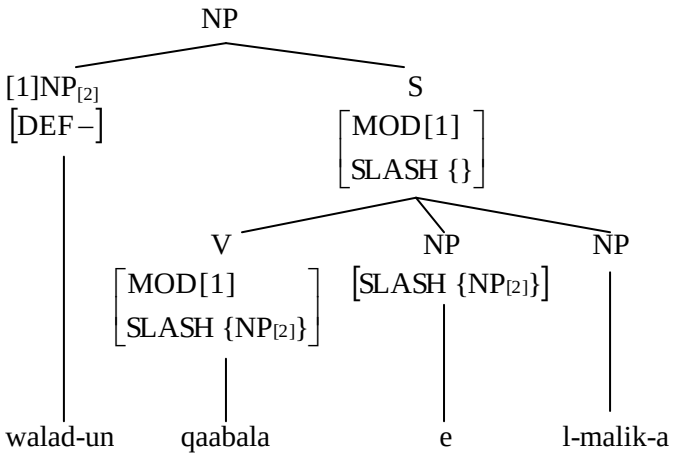
(13) $\left[\text{MOD NP}\left[\text{DEF}-\right]\right]$

This machinery will give the following structures for the complex NPs in (1a) and (2a) (where we assume with Levine and Hukari 2006 that gaps are empty categories):

(14)



(15)



This approach seems quite promising. However, it has two important weaknesses. First it entails that verbs have rather different categories in indefinite relatives from those that they have elsewhere. Since they look just the same as verbs in other contexts, this seems dubious. Secondly, it has two separate stipulations to ensure verb-initial order. On the one hand, it stipulates in (9) that indefinite relatives are head-subject-complement phrases. On the other hand, it stipulates that forms of *ʔallaði* take an [INV +] complement. The analysis misses the generalization that Arabic relative clauses are verb-initial.

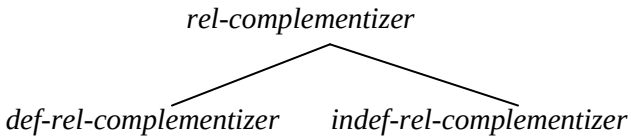
There are also questions about how semantics should be incorporated into this analysis. One might propose that the CONTENT value of the various forms of *ʔallaði* is a restricted index, the restrictions stemming from its complement and the NP it modifies. However as Sag (1997: 474) notes, it is not very plausible to assume that a finite verb has a restricted index as its CONTENT value. It looks, then, as if one would probably have to assume that both definite and indefinite relatives have a proposition as their CONTENT value and to follow Sag in using a special *head-relative-phrase* type to derive a restricted index from this.

It seems, then, this approach has various drawbacks. They suggest that we should look for an alternative approach.

4. An alternative analysis

The obvious alternative to the analysis we have just outlined is one in which indefinite relatives are headed by a phonologically empty counterpart of *ʔallaði*. We can assume the following lexical types:

(16)



rel-complementizer will have the following description:

(17)

$$\left[\begin{array}{l} \textit{rel-complementizer} \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\left[\begin{array}{l} \text{HEAD}\left[\begin{array}{l} c \\ \text{MOD NP}_i \end{array}\right] \\ \text{SUBJ}\ \langle\rangle \\ \text{COMPS}\left\langle\left[\text{LOC}\,|\,\text{CAT}\left[\begin{array}{l} \text{HEAD}\left[\begin{array}{l} \textit{verb} \\ \text{INV}+ \end{array}\right] \\ \text{SUBJ}\ \langle\rangle \\ \text{COMPS}\ \langle\rangle \\ \text{SLASH}\ \{\text{NP}_i\} \end{array}\right]\right]\right\rangle \end{array}\right] \end{array}\right]$$

This will ensure that relative clauses modify an NP and contain a gap or a resumptive clitic with the same index and that they are verb-initial. *def-rel-complementizer* and *indef-rel-complementizer* will have the following descriptions:

(18)

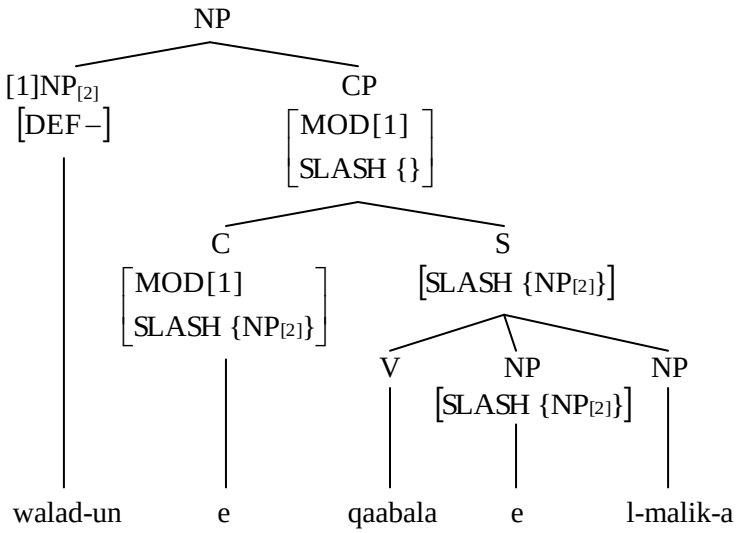
$$\left[\begin{array}{l} \textit{def-rel-complementizer} \\ \text{SS}\,|\,\text{LOCAL}\,|\,\text{CAT}\left[\text{HEAD}\ \left[\text{MOD}\left[\text{DEF}+\right]\right]\right] \end{array}\right]$$

(19)

$$\left[\begin{array}{l} \textit{indef-rel-complementizer} \\ \text{PHON}\ \langle\rangle \\ \text{SS}\,|\,\text{LOCAL}\,|\,\text{CAT}\left[\text{HEAD}\ \left[\text{MOD}\left[\text{DEF}-\right]\right]\right] \end{array}\right]$$

def-rel-complementizer will have various different forms depending on the number and gender and case of the modified NP. *indef-rel-complementizer* is phonologically empty. Within this approach, definite relatives will have the structure in (14). Indefinite relatives will have a similar structure, as in (20).

(20)



In this analysis, finite verbs will have the same category in relative clauses as elsewhere, and we have just one stipulation to ensure verb-initial order. Hence, this analysis does not have the two weaknesses of the first analysis. It will also have a simpler system of phrase types since there is no longer any need for the types of *def-rel-cl* and *indef-rel-cl*.

We might also dispense with the type *rel-cl*. The constraint in (10) ensures that a relative clause modifies an NP with the same index as the NP value of SLASH, but this is also ensured by (17). The only nonredundant feature of (10) is the stipulation that relative clauses are [SLASH {}]. There is an obvious alternative way to ensure this. In most head complement structures if the complement has a non-empty SLASH value, the SLASH Amalgamation Principle requires the head to have the same value. However, there are situations in which the head should not have this value. For example, in (21) the infinitival complement of *easy* is [SLASH {NP}] but *easy* must be [SLASH {}].

(21) Kim is easy to impress.

If SLASH Amalgamation Principle is a default constraint, this can be ensured by a stipulation in the description of *easy* that it is [SLASH {}]. We can take the same approach here. That is, we can replace (17) by (22).

(22)

$$\begin{bmatrix} \textit{rel-complementizer} \\ \text{SS}\left[\begin{array}{l} \text{LOC}\,|\,\text{CAT}\left[\begin{array}{l} \text{HEAD}\left[\begin{array}{l} c \\ \text{MOD NP}_i \end{array}\right] \\ \text{SUBJ}\ \langle\rangle \\ \text{COMPS}\left\langle \left[\text{LOC}\,|\,\text{CAT}\left[\begin{array}{l} \text{HEAD}\left[\begin{array}{l} \textit{verb} \\ \text{INV}+ \end{array}\right] \\ \text{SUBJ}\ \langle\rangle \\ \text{COMPS}\ \langle\rangle \\ \text{SLASH}\ \{\text{NP}_i\} \end{array}\right]\right] \right\rangle \end{array}\right] \\ \text{NONLOCAL}\,|\,\text{SLASH}\ \{\} \end{array}\right] \end{bmatrix}$$

With this revision there is no need for a type *rel-cl* subject to some constraint. Relative clauses are just head-complement structures, whose properties stem from the lexical items that head them.

It is also a simple matter in this approach to deal with semantics. We can simply assume that the CONTENT value of *rel-complementizer* is a restricted index, with the restrictions stemming from its complement and the NP it modifies. This means that there is no need for the special *head-relative-phrase* type which seems necessary in the first analysis.

It seems, then, that there are good reasons for preferring an analysis of Arabic relatives with a phonologically empty head and a simple system of phrase types to an analysis with no phonologically empty heads and a more complex system of phrase types. Thus, whatever may be the case with other languages, with Arabic it looks as though the kind of approach developed in Pollard and Sag (1994) is preferable to the approach developed in Sag (1997).

References

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Levine, R. D. and T. E. Hukari (2006), *The Unity of Unbounded Dependency Constructions*. Stanford, CA: CSLI publications.

Sag, I. A. (1997). English relative clause constructions. *Journal of Linguistics, 33,* 431-484.

Vaillette, N. (2000), Hebrew relative clauses in HPSG, *Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar*, CSLI Publications. pp. 305-324.