Reference patterns in subjunctive complement clauses of Modern Standard Arabic

Tali Arad Greshler
University of Haifa

Livnat Herzig Sheinfux
University of Haifa

Nurit Melnik
The Open University of Israel

Shuly Wintner
University of Haifa

Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar

Polish Academy of Sciences, Warsaw, Poland

Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (Editors)

2016

CSLI Publications

pages 4–22

http://csli-publications.stanford.edu/HeadLex/2016

Keywords: control, coreference, corpus study, Modern Standard Arabic

Arad Greshler, Tali, Herzig Sheinfux, Livnat, Melnik, Nurit, & Wintner, Shuly. (2016). Reference patterns in subjunctive complement clauses of Modern Standard Arabic. In Arnold, Doug, Butt, Miriam, Crysmann, Berthold, King, Tracy Holloway, & Müller, Stefan (Eds.): *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland* (pp. 4–22). Stanford, CA: CSLI Publications.

Abstract

In this paper we investigate the status of control constructions in Modern Standard Arabic (MSA). MSA has several embedded clause constructions, some of which resemble control in English (and other languages). However, these constructions exhibit some notable differences. Chief among them is the fact that the embedded verb carries agreement features that can indicate both coreference and disjoint reference between a matrix argument and the understood subject of the complement clause. We conducted a thorough corpus-based investigation of such constructions, with a special focus on a search for obligatory control in the language. We show that our findings contradict accepted generalizations (and predictions) proposed by state-of-the-art theories of control, as they indicate that there are no "real" control predicates in MSA. We outline an HPSG analysis that accounts for the MSA data.

1 Introduction

Does Modern Standard Arabic (MSA) have control constructions? MSA has several embedded clause constructions, some of which resemble control in English (and other languages). However, these constructions exhibit some notable differences. Chief among them is the fact that the embedded verb carries agreement features that can indicate both coreference and disjoint reference (dis-reference) between a matrix argument (subject or object) and the understood subject of the complement clause.

The first goal of this paper is to investigate whether all verbs in MSA allow for both coreference and dis-reference, or whether there are predicates which enforce coreference between the understood subject of the embedded clause and a matrix argument. Note that in order to consider the phenomenon from a broad theory-neutral perspective we avoid using the term *control* with all its theoretical implications, unless it is specifically mentioned in the proposals we review. Instead, we distinguish between *co-reference predicates*, which enforce coreference between the subject of their complement clause and one of their arguments, and *free-reference predicates*, which do not pose restrictions on the referent of the embedded subject.

In order to determine whether obligatory co-reference predicates exist in the language we conducted a thorough corpus-based search of such constructions. This empirical investigation was informed by previous insights regarding the distinction between co-reference and free-reference predicates specifically in MSA, in other languages, and from a more general cross-linguistic perspective. We show that our findings contradict accepted generalizations (and predictions) proposed by state-of-the-art theories of control.

[†]This research was supported by THE ISRAEL SCIENCE FOUNDATION (grant No. 505/11). We thank Abdelnaser Jabarin for his helpful advice and judgments regarding the MSA data.

The second goal of the paper is to propose an HPSG analysis of the MSA constructions. To achieve this, we build on existing analyses of the MSA clause structure. We consider whether and how they can be extended to account for the phenomena in the focus of this paper.

The structure of the paper is as follows. We begin Section 2 by briefly reviewing some basic properties of MSA that are relevant to the current study and proceed to discuss *ʔan* clauses, which resemble control constructions, in more depth. In Section 3 we review previous proposals that aim to distinguish between co-reference and free-reference predicates. Our corpus findings are presented in Section 4. In Section 5 we outline an analysis that accounts for the MSA data.

2 Background

2.1 Word order and agreement

Modern Standard Arabic is a *pro*-drop language whose unmarked word order is VSO, yet SVO order is also available. The two word orders differ in their agreement patterns. VSO clauses exhibit partial subject–verb agreement, where the verb agrees with its subject in gender and person, yet its number is invariably singular (1a). SVO clauses, on the other hand, exhibit full subject–verb agreement and therefore the verb bears plural agreement when it has a plural subject (1b). The full/partial agreement distinction is only discernable with plural human subjects. Plural inanimate subjects always trigger singular-feminine agreement.

(1) a. *qaraʔat t^{ʕ}-t^{ʕ}aalibaat-u l-kitaab-a*
*read.*3SF *the-students.*PF-NOM *the-book*-ACC
'The female students read the book.'

b. *ʔat^{ʕ}-t^{ʕ}aalibaat-u qaraʔna l-kitaab-a*
*the-students.*PF-NOM *read.*3PF *the-book*-ACC
'The female students read the book.'

Finally, *pro*-dropped subjects trigger full agreement on the verb, as demonstrated in (2).[1]

(2) *qaraʔat l-kitaab-a*
*read.*3SF *the-book*-ACC
'She read the book.' (Not: 'They read the book.')

2.2 Complement clauses

MSA has two types of complement clauses, introduced by two principal particles: *ʔan* and *ʔanna*. Example sentences are given in (3a) and in (3b).[2]

[1] See Section 5.1 for an elaboration.

[2] *ʔanna* is a complementizer. However, the syntactic category of *ʔan* is subject to debate and is identified as a functional head, a marker, or a complementizer. Thus, *ʔanna* is glossed as 'that',

(3) a. *qarrara muħammad-un [ʔan yaktuba*
*decided.*3SM *Muhammad-*NOM(M) *AN write.*3SM-SBJ
r-risaalat-a]
*the-letter-*ACC
'Muhammad decided to write the letter.'

b. *ʕarafa muħammad-un [ʔanna l-walad-a*
*knew.*3SM *Muhammad-*NOM(M) *that the-boy-*ACC
sa-yaktubu r-risaalat-a]
*will-write.*3SM-IND *the-letter-*ACC
'Muhammad knew that the boy would write the letter.'

ʔan clauses and *ʔanna* clauses differ in the following respects. First, the two types of embedded clauses are selected by different predicates. Second, the head of *ʔan* clauses is a verb in the subjunctive mood, while in *ʔanna* clauses it appears in the indicative mood (perfect or imperfect). Third, *ʔan* clauses are verb-initial and when the subject is overt it is marked with nominative case (e.g., (6)), while in *ʔanna* clauses the subject appears clause-initially and bears accusative case. Finally, nothing but negation can intervene between *ʔan* and the subjunctive verb, while in *ʔanna* clauses, as stated, the subject intervenes between *ʔanna* and the verb. In this paper we focus on *ʔan* clauses.

MSA *ʔan* clauses typically appear with no overt subject, yet their unexpressed subject is construed as an argument of the matrix verb. These cases are similar to familiar control constructions in English (and other languages). However, unlike in English, the agreement marking on the subjunctive verb reveals the agreement properties of the intended subject. In (3a) the subjunctive *yaktuba* '*write*' agrees with the matrix subject, Muhammad, which is construed as its understood subject. In (4) the subjunctive *taktuba* '*write*' agrees with the matrix object, Hind, which is construed as its understood subject.

(4) *ʔaqnaʕa muħammad-un hind-an ʔan taktuba*
*convinced.*3SM *Muhammad-*NOM(M) *Hind-*ACC(F) *AN write.*3SF.SBJ
r-risaalat-a
*the-letter-*ACC
'Muhammad convinced Hind to write the letter.'

The MSA construction differs from control in English in another respect. The understood subject of the *ʔan* clause and the matrix argument (subject or object) do not necessarily share a reference. Thus, (3a) is actually ambiguous, as the understood subject of the embedded clause can refer to someone other than Muhammad, resulting in the additional reading illustrated in (5).

---

while *ʔan* is glossed as 'AN', without committing to a particular analysis. See Habib (2009) for a discussion of the syntactic category of *ʔan*.

(5) *qarrara* *muħammad-un* *[ʔan yaktuba* *r-risaalat-a]*
*decided.*3SM *Muhammad-*NOM*(*M*)* *AN* *write.*3SM-SBJ *the-letter-*ACC
'Muhammad$_i$ decided that he$_j$ would write the letter.'

In addition, since the understood subject of *ʔan* clauses does not necessarily corefer with the matrix subject (or another argument), the subjunctive verb may exhibit agreement properties which are distinct from those of the matrix verb. As an example, consider (6). The embedded verb bears third-person-singular-feminine (3SF) agreement and can optionally have an overt agreeing subject. This is another property in which MSA *ʔan* clauses differ from control constructions: the embedded clause can have an overt subject.

(6) *qarrara* *muħammad-un* *ʔan taktuba* *(hind-un)*
*decided.*3SM *Muhammad-*NOM*(*M*)* *AN write.*3SF-SBJ *(Hind-*NOM*(*F*))*
r-risaalat-a
*the-letter-*ACC
'Muhammad decided that Hind/she would write the letter.'

An additional configuration, which we will refer to here as a 'backward pattern', is one where only the embedded subject is overt (7). In this case, similarly to (3a), its 'forward pattern' counterpart, when the embedded verb and the matrix verb agree the sentence is ambiguous: the unexpressed matrix subject can either be construed as Muhammad, the subject of the embedded verb *yaktuba* '*write*', or as someone else, resulting in the second interpretation presented in (7).

(7) *qarrara*$_{i/j}$ *[ʔan yaktuba*$_i$ *muħammad-un*$_i$ *r-risaalat-a]*
*decided.*3SM *AN* *write.*3SM.SBJ *Muhammad-*NOM*(*M*)* *the-letter-*ACC
'Muhammad$_i$ decided that he$_i$ would write the letter.'
'He$_j$ decided that Muhammad$_i$ would write the letter.'

To conclude, MSA *ʔan* clauses differ from control constructions in other languages in four principle respects: (1) Arabic *ʔan* clauses contain a finite subjunctive verb form; (2) The subjunctive bears agreement features; (3) The subject of the *ʔan* clause does not necessarily corefer with an argument of the matrix predicate; (4) The *ʔan* clause can involve an overt embedded subject. Note that (3) and (4) are independent of each other; there can be an embedded subject in the *ʔan* clause or not, and this subject can corefer with the matrix subject or not.

3 Distinguishing co-reference and free-reference predicates

Examples similar to the introductory examples in (3a) and (4)-(7) are found in reference grammars of MSA (Badawi et al., 2004; Cantarino, 1976; Ryding, 2005). Yet in none of these sources do the authors explicitly distinguish between co-reference and free-reference predicates. Nevertheless, this question is addressed

from a functionalist perspective by Persson (2002) and from a generative linguistics perspective in a study by Habib (2009).

Persson (2002), in her corpus-based study of sentential complements in MSA, distinguishes between *ʔan* clauses with an overt embedded subject (and no coreference), and *ʔan* clauses which she describes as clauses in which the embedded subject is deleted under coreference. She argues that the semantic properties of embedding verbs determine their preference for either construction; *manipulative* predicates (*force*, *allow*) prefer coreference, while *cognitive* predicates (desiderative, commentative, fearing, e.g, *want*, *wish*) prefer dis-reference. Persson excludes modality predicates from her study, due to her assumption that they obligatorily require the complement clause subject to be co-referent with the matrix subject. Habib (2009), on the other hand, assumes that there are no "real" control predicates in MSA; all *ʔan* clauses allow for both coreference and dis-reference.

The literature does not seem to have a conclusive answer to the question of whether "real" control exists in MSA. Nevertheless, an interesting parallel is found in Modern Greek (MG), a language which shares a number of syntactic properties with MSA.

Roussou (2009) discusses control and non-control constructions in MG. Like MSA, MG is a *pro*-drop language. MG has two types of complement clauses: *oti* clauses (8), and *na* clauses (9).

(8) *O Yannis pistevi* ***[oti*** *to sipiti ine/itan oreo]*
The Yanis.NOM.S *believes*.S *that the house*.NOM.3S *is/was*.3S *beautiful*
'Yannis believes that the house is/was beautiful.'

(9) *O Kostas matheni* ***[na*** *odhiji]*
the Kostas learn.3S *PRT drive*.3S
'Kostas is learning (how) to drive.'

The distinctions between the two types of complement clauses are reminiscent of those between *ʔan* and *ʔanna* clauses in MSA. The mood of *oti*-complements is always indicative, and their tense is variable. *Na*-complements, on the other hand, have subjunctive mood and invariable present tense. Furthermore, *oti*-complements can be separated from the verb by different elements; *na* must be adjacent to its selecting verb, with only the possibility of a negative element intervening.

Roussou (2009) shows that some MG predicates (e.g., *matheno* '*learn*') require the understood subject to be co-referential with the matrix subject, while others (e.g., *thelo* '*want*') allow for both a co-referential and a non-coreferential interpretation. As examples consider (9) and (10). In (9), Kostas must be the understood subject of *drive*, while in (10), the understood subject of *drive* can be Kostas or someone else.

(10) *O Kostas theli [na odhiji]*
the Kostas want.3S *PRT drive*.3S
'Kostas wants (him) to drive.'

With regard to MG, Roussou (2009, p. 1828) suggests that "there seems to be a continuum, which has aspectuals and then modals on the one end and volitionals (and epistemics) on the other". In between, there are predicates which may be closer to either end, and are subject to individual speakers' preferences. Roussou's proposed continuum is shown in Figure 1.

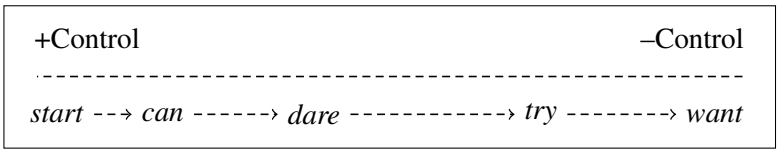

Figure 1: The control continuum (Roussou, 2009)

The distinction between obligatory control (OC) predicates and no control (NC) predicates is discussed by Landau (2013) in his comprehensive study of control.[3] Landau proposes a categorical bifurcation between two types of predicates, based on the semantic (in)dependence of the tense of their complement clauses ([T]), as well as their manifestation of overt morphological agreement ([Agr]). The tense specification of complement clauses depends on whether or not their tense is anaphoric to the tense of the matrix clause. Thus, when the complement clause is *tensed* the matrix and embedded events can be temporally mismatched (11a), but when the complement clause is *untensed* they must match (11b).

(11) a. Yesterday, John hoped to solve the problem tomorrow. → infinitive is [+T]

b. *Yesterday, John managed to solve the problem tomorrow. → infinitive is [-T]

Based on this characterization, Landau categorizes the types of predicates which select tensed or untensed complement clauses.

(12) Predicates which select untensed [-T] complements

a. Implicatives (dare, manage, remember,...)

b. Aspectuals (start, stop,...)

c. Modals (have, need, may,...)

d. Evaluative adjectives (rude, silly,...)

(13) Predicates which select tensed [+T] complements

a. Factives (glad, sad, like,...)

[3]Landau distinguishes between OC, NC, and non-obligatory control (NOC). OC and NC occur in complement clauses, while NOC occurs in subject and adjunct clauses. OC and NOC clauses host a PRO subject, and NC clauses host a *pro*/DP subject. PRO in OC is interpreted as a bound variable, which is co-indexed with a co-dependent of the matrix clause. PRO in NOC is logophoric or topic-bound.

b. Propositional (believe, think, claim,...)

c. Desideratives (want, prefer, hope,...)

d. Interrogatives (wonder, ask, find out,...)

The combination of the tense [T] and agreement [Agr] parameters produces four different options, which interact with control. According to Landau's *finiteness rule for Obligatory Control* if a complement clause has slots for both T and Agr, and they are both positive, then no-control (NC) obtains. Otherwise, control is obligatory (OC).

	+T	**-T**
+Agr	NC	OC
-Agr	OC	OC

Table 1: The finiteness rule for Obligatory Control (Landau, 2013)

Landau's (2013) rule implies that if a complement clause is semantically untensed it will enforce obligatory control. Thus, Landau (2013) predicts that "[t]here cannot be a language where modal, aspectual and implicative verbs or evaluative adjectives allow an uncontrolled complement subject" (p. 106).

The picture that emerges from the studies presented so far is that the distinction between co-reference and free-reference predicates is directly linked to their semantic properties. Building on these studies we form predictions regarding the types of predicates associated with each construction. While Habib (2009) does not acknowledge the existence of obligatory coreference in MSA, Persson (2002), Roussou (2009) and Landau (2013) all identify modals as typically coreference-enforcing predicates. Applying Roussou's (2009) continuum to MSA we predict that predicates that are closer to the left end of her continuum would be more likely to enforce coreference. According to Landau's (2013) analysis, given [+Agr], as is the case in MSA *ʔan*-clauses, which are headed by a subjunctive form that shows overt morphological agreement, [+T] implies NC, and [-T] implies OC. Thus, the prediction is that MSA predicates which select an untensed complement clause would be the ones to enforce coreference. These predictions are put to the test in the following section, where we present corpus findings regarding the reference patterns of a representative set of *ʔan*-clause-taking predicates.

4 A corpus study of co-reference and free-reference predicates in MSA

In order to determine whether co-reference predicates exist in MSA we conducted a corpus-based investigation of *ʔan* clauses in contemporary MSA. Our searches focused on representative predicates from Roussou's (2009) continuum and Landau's (2013) classification, in addition to a set of tri-valent manipulative predicates,

which were demonstrated by Persson (2002) to prefer coreference and which are typically object control predicates.

The corpus that we used is the 115-million-token sample of the *arTenTen* corpus of Arabic (Arts et al., 2014). This sample has been tokenized, lemmatized and part-of-speech tagged with MADA (Habash & Rambow, 2005; Habash et al., 2009) and installed in the Sketch Engine (Kilgarriff et al., 2004). The morphological tagging of the corpus provides a way of defining queries which target particular person, number and gender features, as well as case and mood. Consequently, we were able to retrieve instances where the matrix predicate and the embedded predicate match in their gender and person agreement, as well as those where there is a mismatch. Furthermore, we could control for the existence or lack of a possible subject (i.e., agreeing nominative noun) following the predicates.[4]

Nevertheless, the search results are not exhaustive. There are numerous instances of erroneous morphological tags, which contributed to false positive results as well as false negatives. Moreover, we decided to favor precision over recall, and limited the distance between the predicates. Consequently, instances with longer NP subjects or intervening adverbials were not retrieved. These limitations notwithstanding, in what follows we provide examples of coreference and dis-reference for a representative set of predicates. Due to the non-exhaustivity of the searches we do not present quantitative data with regard to the distribution of coreference and dis-reference. We do, however, note whether we found dozens of similar examples or whether there were only several examples of disjoint reference.

The corpus search revealed evidence for both coreference and dis-reference with representatives of verbs on Roussou's (2009) continuum, and in Landau's (2013) categories, listed in increasing order by their likelihood to enforce coreference, according to Roussou: the volitional verb *ʔaraada* '*want*', the implicatives *ħaawala* '*try*' and *ʒaruʔa* '*dare*', and the dynamic modal *ʔistatˤaaʕa* '*be able*'. In addition, we found dis-reference examples with the manipulative predicates *ʔaqnaʕa* '*convince*' and *samaħa* '*allow*'. In what follows we present corpus-based examples of coreference and dis-reference with each of the aforementioned predicates.

Volitionals We start at the right end of Roussou's (2009) continuum. Volitionals are predicted by Roussou (2009) and by Landau (2013) to allow free-reference. Consider the volitional *ʔaraada* '*want*' in (14).

[4]For example, to retrieve instances of forward coreference with the verb *ʔaraada* '*want*', we queried for cases where the lemma *ʔaraada* '*want*' is optionally followed by non-verbal material (i.e., the subject, adverbs), then *ʔan* and an adjacent subjunctive verb, which in turn is followed, not necessarily immediately, by a non-nominative noun (i.e., not the subject). Moreover, we restricted the two verbs to share their person and gender properties. The corresponding CQL (Contextual Query Language) query that we constructed was: 1:[tag="verb" & lemma="ʔaraada"] [tag!="verb"] {0,3} [word="ʔan"] 2:[tag="verb" & modus="s"] []{0,2} [tag="noun" & case!="n"] & 1.gender=2.gender & 1.person=2.person.

(14) a. *ʔaraada [ʔan yaʕmala diraasat-an]*
*wanted.*3SM *AN do.*3SM.SBJ *study-*ACC
'He wanted to conduct a study.'

b. *ʔaraada [ʔan yakuuna r-radd-u wat^{ʕ}aniyy-an]*
*wanted.*3SM *AN be.*3SM.SBJ *the-reaction-*NOM(M) *national-*ACC
'He wanted the reaction to be national.'

In (14a), the subject of the embedded predicate corefers with the subject of the matrix predicate; the same person is both the 'wanter' and the 'conductor' of the study. In (14b), on the other hand, the embedded clause involves an overt subject, 'reaction', whose reference is distinct from that of the matrix subject. Our corpus searches revealed dozens of examples of disjoint reference with the predicate *ʔaraada* '*want*'.

Implicatives Moving left on Roussou's (2009) continuum, we found dozens of examples of disjoint reference with the predicate *ħaawala* '*try*', indicating that it is indeed a free-reference predicate. While the matrix and embedded verbs share a subject in (15a), in (15b) the matrix verb bears 1P agreement while the embedded verb bears 3SM agreement and has an overt subject. Clearly, the two subjects do not share a reference.[5]

(15) a. *ħaawala r-raʒul-u [ʔan yatakallama maʕa-na]*
*tried.*3SM *the-man-*NOM *AN speak.*3SM.SBJ *with-us*
'The man tried to speak with us.'

b. *ʔinna-na nuħaawilu [ʔan yataħaddaθa s^{ʕ}amt-u-na]*
*indeed-we try.*1P *AN speak.*3SM.SBJ *silence-*NOM*-our(*M*)*
'We are trying to make our silence speak.'

The implicative 'dare' is closer to the left end of Roussou's (2009) continuum and is classified in Landau's (2013) categorization as an untensed predicate. Thus, the prediction is that it will enforce coreference, or in other words, be an OC predicate. However, as (16b) shows, this is not the case. MSA *ʒaruʔa* '*dare*' allows free-reference between the embedded subject and its subject; the verb 'be' in (16b) has its own overt subject, 'her opinion', and does not match in agreement with the matrix verb, 'dare'. Admittedly, the disjoint reference example presented here is the only one we were able to find with this predicate. Note, however, that *ʒaruʔa* '*dare*' in itself is an infrequent verb (12.93 per million instances), with substantially fewer attestations of it followed by an *ʔan* clause (1.36 per million).

[5]Interestingly, the disjoint reference examples (15b) and (16b) involve a possessive pronominal clitic on the embedded subject, which refers back to the matrix subject. Such "indirect" coreference with a matrix argument is frequent in dis-reference examples, and can also be expressed as an object clitic on the embedded verb, but it is not obligatory. This coreference creates cohesion between the two events denoted by the two clauses.

(16) a. *laa yaʒruʔu raʒul-un [ʔan yaquula l-ħaqiiqat-a fi*
*not dare.*3SM *man-*NOM *AN say.*3SM.SBJ *the-truth-*ACC *in*
l-zawaaʒ-i]
*the-marriage-*GEN
'No man dares to say the truth in the marriage.'

b. *lan taʒruʔa [ʔan yakuuna raʔy-u-haa ɣayr-a*
*never dare.*3SF *AN be.*3SM.SBJ *opinion-*NOM*-her(*M*) not-*ACC
musaanid-in li-lmaɣrib-i]
*supportive-*GEN *to-Morocco-*GEN
'She will never dare that her opinion would be non-supportive of Morocco.'

Manipulatives Tri-valent manipulatives do not appear in Roussou's (2009) continuum, yet Persson (2002) identifies them as the ones which generally impose a coreference restriction. Obtaining exhaustive results with predicates from this class was even more complex than obtaining them with 'subject-control' predicates. However, here too we find evidence of both types of reference relations, with several instances of disjoint reference. (17b) is a disjoint reference example of the predicate *ʔaqnaʕa* '*convince*', and (18b) is a similar example of the predicate *samaħa* '*allow*'.

(17) **Convince**

a. *wa-fi l-masaaʔ-i kaanat malaak qad ʔaqnaʕat*
*and-in the-evening-*GEN *was.*3SF *Malak(*F*) already convinced.*3SF
waalid-a-haa [ʔan yaʔmura saaʔiq-a-hu
*father-*ACC*-her AN order.*3SM.SBJ *driver-*ACC*-his(*M*)*
l-xaasˁsˁ-a bi-ʔiisˁaal-i buuʒaa ʔila qaryat-i-hi]
*the-private-*ACC *in-delivering Buja to village-*GEN*-his*

'And in the evening, Malak had already convinced her father to order his private driver to deliver Buja to his village.'

b. *ʔaqnaʕnaa-hum [ʔan yuʕayyina huwa*
*convinced.*1P*-them AN appoint.*3SM.SBJ *he.*NOM
l-ħukuumat-a]
*the-government-*ACC
'We convinced them that he would appoint the government.'

(18) **Allow**

a. *iðaa lam nasmaħu li-l-ʔameriikaan-i [ʔan yamurruu min*
*if not allow.*1P *to-the-Americans-*GEN *AN pass.*3PM.SBJ *from*
ʔaraadˁii t-turkiyya]
territory the-Turkish

'If we don't allow the Americans to pass from Turkish territory'

b. *fa-mawqiʕ-u-hu l-ʔiʒtimaaʕiyy-u laa yasmaħu lahu [ʔan*
and-status-NOM-*his the-social*-NOM *not allow.*3SM *to.him AN*
yakuuna bnu-hu fii haaða l-makaan-i]
*be.*3SM.SBJ *son*-NOM-*his in this the-place*-GEN
'And his social status does not allow him that his son will be in this place.'

Modals Modals like 'can' are close to the left (OC) end of Roussou's (2009) continuum and are classified as untensed by Landau (2013). The prediction is therefore that they would enforce coreference. This prediction, however, does not hold. We found dozens of instances of the predicate *ʔistatˤaaʕa* '*be able*' in which the embedded subject does not corefer with the matrix subject. One such case is (19b), in which the predicate is used as a dynamic modal expressing intention or willingness. The matrix subject is a *pro*-dropped first-person-plural subject while the embedded subject is the third-person-singular-feminine 'government'.

(19) a. *lam ʔastatˤiʕ [ʔan ʔasmaʕa sˤawt-a-hu ʔaw ʔaraa-hu]*
*not be.able.*1S *AN hear.*1S.SBJ *voice*-ACC-*his or see.*1S.SBJ-*him*
'I couldn't hear his voice or see him.'

b. *lan nastatˤiiʕa [ʔan tataħammala l-ħukuumat-u*
*never be.able.*1P.SBJ *AN carry.*3SF.SBJ *the-government*-NOM(F)
kaamil-a n-nafaqaat-i]
all-ACC *the-expenses*-GEN
'We will never be able (to accept the fact that) the government will carry all the expenses.'

All the *ʔan*-clause selecting predicates that were investigated in our corpus study turned out to be free-reference predicates, as instances of disjoint reference with them were attested. Importantly, we found disjoint reference examples of modals, which were predicted to enforce coreference. Consequently, we tentatively conclude that MSA does not have predicates which enforce coreference.

5 Towards an analysis

5.1 An HPSG analysis of VSO and SVO clauses in MSA (Alotaibi & Borsley, 2013)

The syntactic structure of VSO and SVO Arabic clauses has been thoroughly discussed in the literature (Fassi Fehri, 1993; Mohammad, 2000; Aoun et al., 2010; Alotaibi & Borsley, 2013, among others). The main challenge is the subject–verb agreement asymmetries between SVO and VSO clauses described in Section 2.1. The analysis put forth by Aoun et al. (2010) and elaborated and cast in HPSG by

Alotaibi & Borsley (2013) proposes that clause structure in MSA is invariantly VSO, where number agreement is suppressed (see Figure 2).

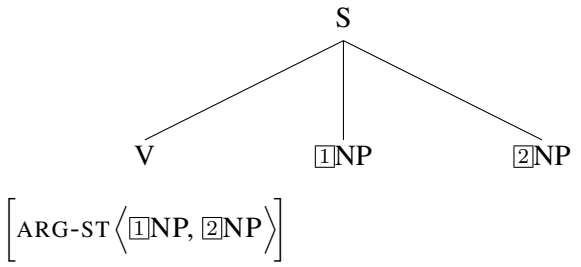

Figure 2: VSO

In the two constructions where full agreement on the verb is found, namely SVO structures and *pro*-drop, the manifestation of full person–number–gender agreement is triggered by the existence of a clitic, referred to as *pro*, which realizes an unexpressed subject. This account, proposed by Alotaibi & Borsley (2013), echoes the analysis proposed by traditional Arab grammarians. In SVO structures what looks like a preverbal subject is in fact a topic which is associated with *pro* subject resumptive pronoun (see Figure 3). This analysis is supported by the fact that subject arguments in SVO clauses are required to be definite.

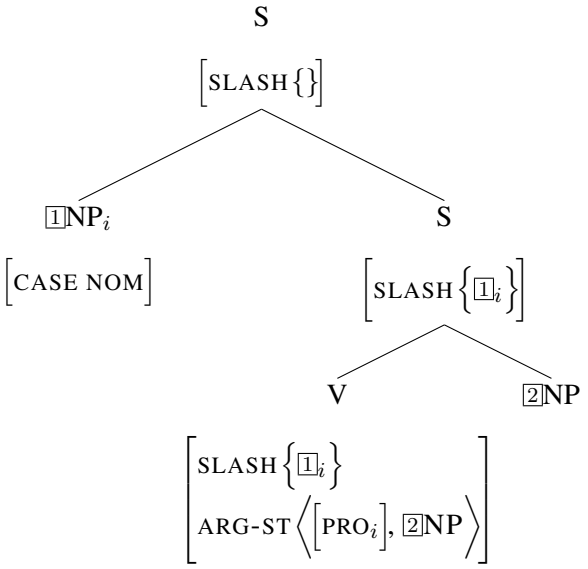

Figure 3: SVO

What is crucial for the current discussion is the idea, which originated in traditional grammar and was adopted and formalized by Alotaibi & Borsley (2013), that *pro*-dropped subjects always trigger full agreement. As we show in the next section, extending this notion to the analysis of MSA coreference constructions

can account for the agreement patterns observed in forward constructions but not for the ones observed in backward constructions.

5.2 A *pro*-drop analysis of co-reference and free-reference in MSA

For languages like Modern Greek (and other languages), which have co-reference predicates and free-reference predicates, it is natural to assume that each is associated with a distinct syntactic structure. With regard to MSA, however, if it is indeed the case that it does not have predicates which enforce coreference (i.e., OC predicates), a straightforward analysis would be to assume one structure for both coreference and dis-reference, namely, a free-reference structure.

Consider the schematic representation in Figure 4. Constructions with *ʔan* complement clauses are simply structures with two independent subjects. The matrix verb combines with its subject (lexical NP or *pro*) and with its *ʔan*-clause complement.[6] This complement clause is preceded by a complementizer/marker *ʔan*. The clause itself is in a VSO configuration and is headed by a subjunctive verb. Its subject is either a lexical NP or *pro*.[7]

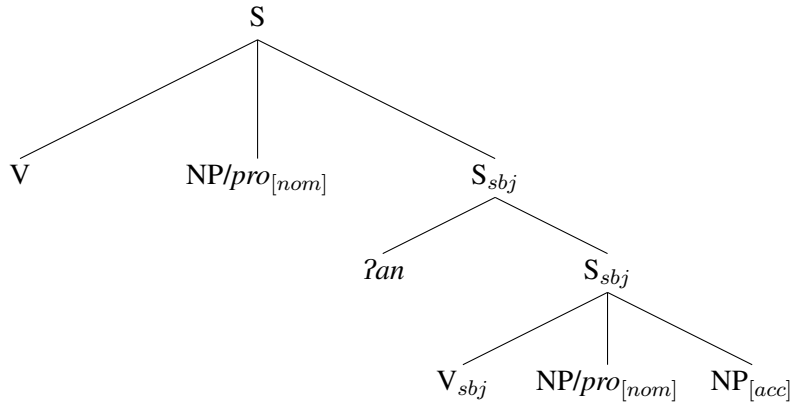

Figure 4: *ʔan* clause complement – coreference/dis-reference

This kind of analysis is based on the *pro*-drop property of MSA; each of the clauses, the matrix clause and the embedded clause, can independently either have an overt subject or a *pro*-dropped subject. There are no constraints on the agreement relations between the two predicates, and therefore they do not need to match. Consequently, what can be construed as subject control is in actuality just a case where the two subjects have identical agreement features, and one of them, either the matrix subject in the backward pattern, or the embedded subject in the forward pattern (or both) is *pro*-dropped. As such, the proposed *pro*-drop analysis does not

[6]Note that the NP/$pro_{[nom]}$ node is an abbreviated notation to indicate the possibility of either using a lexical NP or *pro*-dropped subject and does not imply the existence of empty categories in syntax.

[7]When both the matrix and the embedded subjects are overt NPs, coreference is impossible.

assume a special structure for control and builds on the grammar of MSA to predict the possible constructions of *ʔan*-clause-taking predicates. It is similar in spirit to the analysis proposed by Habib (2009) for all *ʔan* clauses in MSA, and by Roussou (2009) for no-control in Modern Greek, which is also a *pro*-drop language.

Let us first focus on the proposed structure for the forward pattern. In what follows are examples of *ʔan* clauses with plural human subjects, for which the full/partial agreement distinction is observable.

(20) a. *qarrarat* *tˤ-tˤaalibaat-u* *[ʔan taktubna* *r-risaalat-a]*
*decided.*3SF *the-students.*PF-NOM *AN* *write.*3PF-SBJ *the-letter-*ACC

b. *ʔatˤ-tˤaalibaat-u* *qarrarna* *[ʔan taktubna* *r-risaalat-a]*
*the-students.*PF-NOM *decided.*3PF *AN* *write.*3PF-SBJ *the-letter-*ACC
'The female students decided to write the letter.'

In (20a), the plural human subject follows the matrix verb, thus triggering partial agreement, as exhibited by the 3SF agreement on the verb. The embedded verb, on the other hand, exhibits plural agreement, since it involves a *pro*-dropped subject, which according to the grammar of MSA triggers full agreement (see (2)). In (20b), on the other hand, the plural human subject precedes the matrix verb, resulting in full (3PF) agreement on the matrix verb. The embedded verb exhibits full agreement due to its *pro*-dropped subject, just like it does in (20a).

Figures 5 and 6 illustrate the syntactic structure of the forward pattern, with its two variations. The proposed *pro*-drop analysis along with the assumption that *pro*-dropped subjects trigger full agreement (Alotaibi & Borsley, 2013) correctly predict the agreement variations observed in the forward pattern.

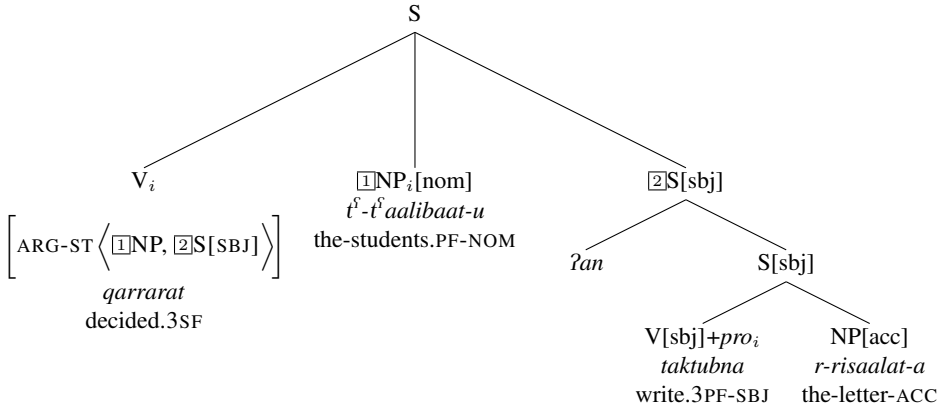

Figure 5: VSO forward pattern - coreference

Moving on to the backward pattern, consider the structure given in Figure 7. According to the *pro*-drop analysis, the matrix verb in this case is *pro*-dropped and the embedded subject is overt. Assuming that *pro*-dropped subjects always trigger

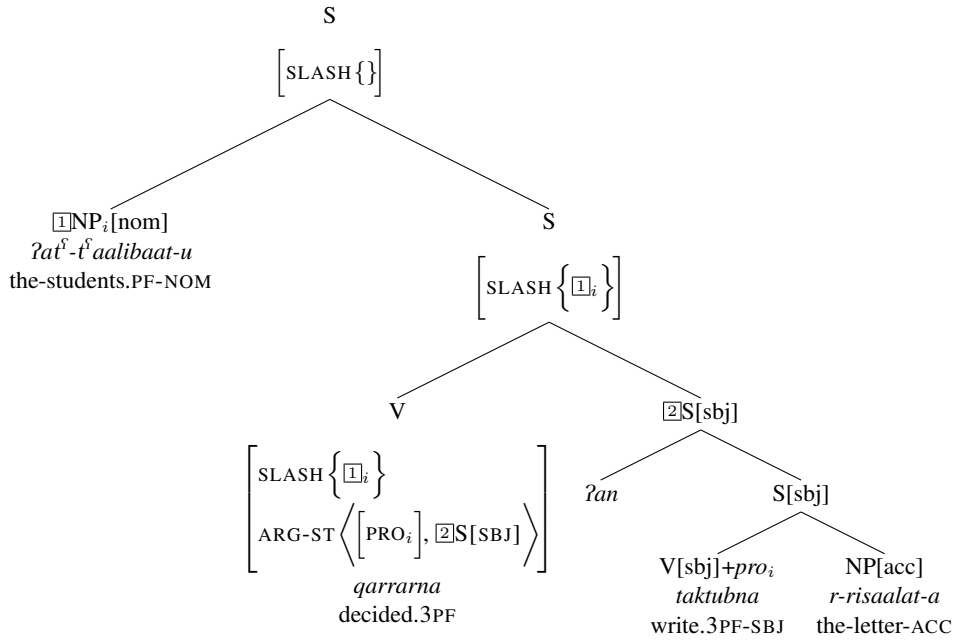

Figure 6: SVO forward pattern - coreference

full agreement, we expect the matrix verb in the backward coreference pattern to exhibit full agreement with the embedded subject.

This, however, is not what is revealed by corpus data. Searches for backward patterns with overt embedded subjects which are both human and plural retrieved only instances in which the matrix verb exhibits partial agreement with the subject. Consider (21), a corpus example of a backward pattern involving a human plural embedded subject.

(21) *takaaliif-u l-ʕilaʒ-i l-baahiðˤat-u llati [laa*
costs-NOM *the-treatment*-GEN *the-exaggerated*-NOM *that not*
yastatˤiiʕu [ʔan yataħammalu-ha l-fuqaraaʔ-u]]
be.able.3SM *AN bear*.3SM.SBJ-*it the-poor*.PM
'The exaggerated costs of the treatment that the poor are not able to bear'

The embedded subject *l-fuqaraaʔ-u* '*the poor*' follows the embedded verb *yataħammalu-ha* '*bear*', triggering partial agreement on it, as predicted by the grammar of MSA. However, the singular agreement on the matrix verb *yastatˤiiʕu* '*be able*' is not predicted by the *pro*-drop analysis and the assumption that *pro* subjects trigger full agreement. Thus, although the *pro*-drop analysis predicts the agreement variations attested in the forward pattern, it makes the wrong prediction with regard to the backward pattern.

Wurmbrand & Haddad (2016) explore backward raising patterns among *ʔafʕaal ʔalmuqaaraba* '*verbs of appropinquation*' in Standard Arabic (SA). This verb class

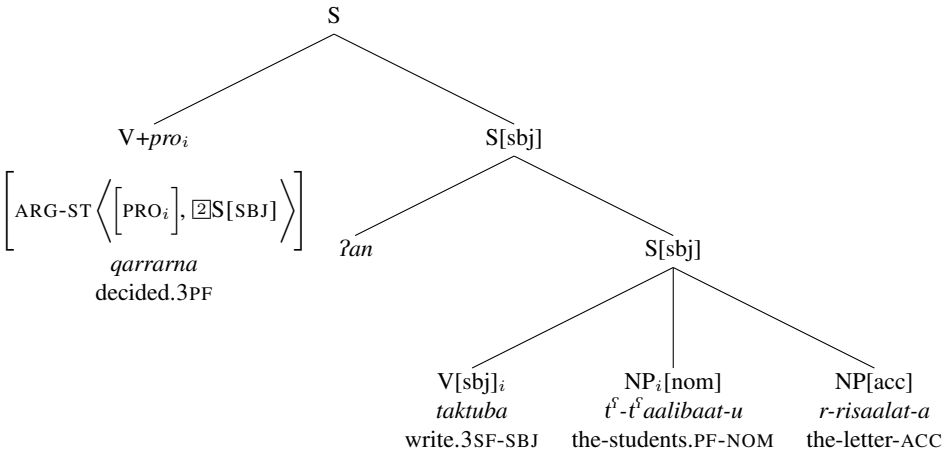

Figure 7: Backward pattern - coreference

encompasses three semantic types: verbs of proximity, verbs of hope, and verbs of inception (Wright, 2007). Wurmbrand & Haddad argue that these verbs can appear in one of two backward raising patterns: one where the matrix verb exhibits full agreement with the overt subject in the embedded clause, and one where agreement is only partial (excluding number).[8] The two patterns are demonstrated in (22).

(22) a. *ʔawʃakna* *[(ʔan)* *tanʒaħa* *t^{ʕ}-t^{ʕ}aalibaat-u]*
were.about.to.3PF *(AN)* *succeed*.3SF.SBJ *the-students*.PF-NOM

b. *ʔawʃakat* *[(ʔan)* *tanʒaħa* *t^{ʕ}-t^{ʕ}aalibaat-u]*
were.about.to.3SF *(AN)* *succeed*.3SF.SBJ *the-students*.PF-NOM

'The female students were about to succeed.'

Wurmbrand & Haddad propose that this pattern is unique only to Standard Arabic verbs of appropinquation, and is not found with other raising predicates or control predicates. They attribute the agreement alternation to the different positions of the unpronounced raised copy (pre-verbal and post-verbal).[9] However, the authors acknowledge that although the full agreement case is the one that conforms with prescriptive grammar, they were not able to find naturalistic instances of this structure in contemporary newspapers. They did find instances of the second pattern, where the matrix verb exhibits partial agreement with the embedded subject. Nevertheless, they assume that both orders are available in Standard Arabic. Our corpus study reveals, contrary to Wurmbrand & Haddad, that the backward pattern

[8]Wurmbrand & Haddad also discuss an impersonal backward pattern, where the matrix verb exhibits default 3SM agreement. This pattern is not relevant to the current discussion.

[9]The alternating agreement is a crucial factor in their analysis since it provides evidence for the structural effects of the deleted higher copy of the subject. This, according to Polinsky & Potsdam (2006), is a necessary condition for "real" backward raising, as opposed to cases of long-distance agreement between the matrix and the embedded predicates.

also exists with "control" predicates. Like them, we were not able to find instances of the full agreement pattern in the corpus.

Corpus-based usage data contradict prescriptive grammar and the internal logic of the grammar for "control" predicates, just like it does for verbs of appropinquation. We believe that this discrepancy can be ascribed to the special circumstances of MSA, which is a language that is not spoken natively by any of its speakers.[10] Thus, we propose that the use of partial agreement in the backward pattern is motivated by analogy to the partial subject–verb agreement found in simple VSO clauses. This type of reasoning may explain why although the partial agreement pattern conflicts with the internal logic of MSA grammar, it is the pattern which speakers choose to use. Certainly, more work is needed for a complete analysis that accounts for the full range of patterns found with *ʔan* complement clauses. This remains an open issue for future work.

6 Conclusion

We showed that there is no evidence for the existence of obligatory coreference (or control) with *ʔan*-clause-taking predicates in MSA. A one-structure *pro*-drop analysis with no specific assumptions regarding control accounts for most of the data, but does not align with the agreement pattern attested in the backward construction. The integration of these data into the theory requires some additional assumptions, which seems to involve extra-grammatical factors, related to the non-native status of MSA.

References

Alotaibi, Mansour & Robert D. Borsley. 2013. Gaps and resumptive pronouns in Modern Standard Arabic. In Stefan Müller (ed.), *Proceedings of the 20th international conference on HPSG*, 6–26. Stanford: CSLI Publications.

Aoun, Joseph E., Elabbas Benmamoun & Lina Choueiri. 2010. *The syntax of Arabic*. Cambridge University Press.

Arts, Tressy, Yonatan Belinkov, Nizar Habash, Adam Kilgarriff & Vit Suchomel. 2014. arTenTen: Arabic corpus and word sketches. *Journal of King Saud University - Computer and Information Sciences* 26(4). 357 – 371. doi: http://dx.doi.org/10.1016/j.jksuci.2014.06.009. http://www.sciencedirect.com/science/article/pii/S1319157814000330.

Badawi, El Said, Michael Carter & Adrian Gully. 2004. *Modern written Arabic: A comprehensive grammar*. Routledge.

[10]MSA is the literary standard of the Arab world, but it is acquired in school. The mother tongue of its speakers is some regional dialect of Arabic.

Cantarino, Vicente. 1976. *Syntax of modern Arabic prose*, vol. 3. Bloomington: Indiana University Press.

Fassi Fehri, Abdelkader. 1993. *Issues in the structure of Arabic clauses and words*. Dordrecht: Kluwer.

Habash, Nizar & Owen Rambow. 2005. Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop. In *Proceedings of the 43rd annual meeting of the association for computational linguistics (acl'05)*, 573–580. Ann Arbor, Michigan: Association for Computational Linguistics. http://www.aclweb.org/anthology/P/P05/P05-1071.

Habash, Nizar, Owen Rambow & Ryan Roth. 2009. MADA+TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization. In Khalid Choukri & Bente Maegaard (eds.), *Proceedings of the second international conference on Arabic language resources and tools*, The MEDAR Consortium.

Habib, Rania. 2009. The syntax of the Standard Arabic *ʔan* and *ʔanna*. In Kleanthes K. Grohmann & Phoevos Panagiotidis (eds.), *Selected papers from the 2006 Cyprus syntaxfest*, 159–194.

Kilgarriff, Adam, Pavel Rychly, Pavel Smrz & David Tugwell. 2004. The Sketch Engine. In *Proceedings of EURALEX*, 105–116.

Landau, Idan. 2013. *Control in generative grammar: A research companion*. Cambridge University Press.

Mohammad, Mohammad A. 2000. *Word order, agreement, and pronominalization in Standard and Palestinian Arabic*. Amsterdam: John Benjamins.

Persson, Maria. 2002. *Sentential object complements in Modern Standard Arabic*, vol. 2. Lund University.

Polinsky, Maria & Eric Potsdam. 2006. Expanding the scope of control and raising. *Syntax* 9(2). 171–192.

Roussou, Anna. 2009. In the mood for control. *Lingua* 119(12). 1811–1836.

Ryding, Karin C. 2005. *A reference grammar of Modern Standard Arabic*. Cambridge University Press.

Wright, William. 2007. *A grammar of the Arabic language*. New York: Simon Wallenberg.

Wurmbrand, Susi & Youssef A Haddad. 2016. Cyclic spell-out derived agreement in Arabic raising constructions. In Youssef A. Haddad & Eric Potsdam (eds.), *Perspectives on Arabic linguistics XXVIII: Papers from the annual symposium on Arabic linguistics, Gainesville, Florida, 2014*, 193–228. Amsterdam/Philadelphia: John Benjamins Publishing Company.