

Seeking control in Modern Standard Arabic

Tali Arad Greshler, Livnat Herzig Sheinflux, Nurit Melnik and Shuly Wintner

1 Introduction

Does Modern Standard Arabic (MSA) have control constructions? MSA has several embedded clause constructions, some of which resemble subject 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 co-reference and dis-reference between the matrix subject 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. We then propose a formal HPSG analysis, which builds on previous accounts of the syntax of MSA.

2 Background

Modern Standard Arabic (MSA) is a *pro*-drop language whose unmarked word order is VSO, yet SVO order is also available. MSA has two types of complement clauses, introduced by two principal particles: *?an* and *?anna*. Example sentences are given in (1a) and (1b).

- (1) a. *qarrara muhammad-un [?an yaktub-a r-risa:lat-a].*
decided.3SM Muhammad-NOM AN write.3SM-SBJ the-letter-ACC
'Muhammad decided to write the letter.'
- b. *?arafa muhammad-un [?anna l-walad-a sa-yaktub-u r-risa:lat-a].*
knew.3SM Muhammad-NOM that the-boy-ACC will-write.3SM-IND the-letter-ACC
'Muhammad knew that the boy would write the letter.'

The two types of embedded clauses are selected by different verbs. The head of *?an* clauses is a verb in the subjunctive mood, while in *?anna* clauses it appears in the indicative mood (perfect or imperfect). Moreover, *?an* clauses are verb-initial and when the subject is overt it is marked with nominative case (e.g., (2)), while in *?anna* clauses the subject appears clause-initially and bears accusative Case.

In this paper we focus on *?an* clauses, which typically occur with no overt subject, and bear resemblance 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. For example, in (1a) the subjunctive *yaktub-a* 'write' agrees with the matrix subject, Muhammad, which can be construed as its understood subject. This, however, is not the only reading; the understood subject can also refer to someone else (i.e., 'Muhammad_i decided that he_j will write the letter.').

In addition, since the understood subject of *?an* clauses does not necessarily co-refer with a matrix subject (or another argument), the subjunctive verb may exhibit distinct agreement properties. As an example, consider (2).

- (2) *qarrara muhammad-un ?an taktub-a (raniat-u) r-risa:lat-a.*
decided.3SM Muhammad-NOM AN write.3SF-SBJ (Rania-NOM) the-letter-ACC
'Muhammad decided that Rania/she will write the letter.'

Naturally, with two different overt subjects there is dis-reference. When the embedded subject is omitted the agreement properties of the embedded verb reveal properties of the understood subject, as in other cases of *pro*-drop. In light of these facts we ask two questions: (1) Are there any cases of "real" control in MSA? and (2) What is the syntactic structure of *?an* clauses in this language?

3 Seeking control

As illustrated by examples (1a) and (2), the verb *qarrara* ‘decide’ allows for both co-reference and dis-reference between its matrix subject and the understood subject of the embedded *?an* clause. Similar examples are given in various reference grammars of MSA (Badawi et al., 2004; Cantarino, 1976; Ryding, 2005). In none of these sources do the authors explicitly distinguish between co-referring and dis-referring predicates. In what follows we present a number of theoretical studies which consider this issue, specifically with regards to the MSA phenomenon, and as a more general phenomenon.

Persson (2002), in her corpus-based study of sentential complements in MSA, distinguishes between *?an* clauses with an overt embedded subject (and no co-reference), and *?an* clauses which she describes as clauses in which the embedded subject is deleted under co-reference. She argues that the semantic properties of embedding verbs determine their preference for either construction. She 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 *?an* clauses allow for both co-reference 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. Like MSA, MG is a *pro*-drop language. The subjunctive *na*-clauses found in MG are remarkably similar to MSA *?an* clauses. Some 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. Roussou (2009, p. 1828) argues with regard to MG 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.

The distinction between obligatory control (OC) predicates and no control (NC) predicates is discussed by Landau (2013) in his comprehensive study of control. Landau proposes a categorical bifurcation between two types of predicates, based on the semantic (in)dependence of the tense of their complement clauses, as well as their manifestation of morphological agreement. He predicts that “[t]here cannot be a language where modal, aspectual and implicative verbs or evaluative adjectives allow an uncontrolled complement subject” (Landau, 2013, p. 106).

In the first part of our study we challenge the generalizations and predictions described above, by conducting a corpus-based investigation of *?an* clauses in contemporary MSA. The search is based on 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 corpus search revealed evidence for both control and NC with representatives of verbs on Roussou’s continuum, and in Landau’s categories: the volitional verbs *?araxda* ‘want’, *ha:wala* ‘try’, *zaru?a* ‘dare’, and the modal *?istat^ʕa:ʕa* ‘be able’. Examples of control and NC with *?istat^ʕa:ʕa* ‘be able’ are given in (3a) and (3b), respectively.

- (3) a. *lam ?astat^ʕi:ʕu [?an ?asmaʕa s^ʕawt-a-hu ?aw ?ara:hu]*
neg.PAST 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 nastat^ʕi:ʕu [?an tataħammala l-ħuku:mat-u ka:mil-a l-nafaqat-i]*
neg.FUT be.able.1P.IND AN carry.3SF.SBJ the-gov’t.SF-NOM all-ACC the-expenses-GEN
 ‘It will never be possible for us that the government will carry all the expenses.’

In addition we find predicates in MSA whose complement clauses are untensed, yet which are compatible with NC, contrary to Landau’s prediction. To date, we have not found any evidence of a predicate which enforces co-reference.

4 Analysis

Under the assumption that there is no obligatory control in MSA, a straightforward account is to propose one structure for all cases, modulo *pro*-drop (see the schematic representation in Figure 1). Clauses with *?an* complement clauses are simply structures with two independent subjects. There are no constraints on the agreement relations between the two predicates, and consequently they do not need to match. What can be considered subject control is in actuality just a case where the two subjects have identical agreement features, and either the matrix subject or the embedded subject (or both), is *pro*-dropped. This is similar in spirit to the analysis proposed by Habib (2009) for all MSA *?an* clauses, and by Roussou (2009) for no-control in Modern Greek.

One problem with this analysis is that it predicts only one of the two backward control patterns which are argued to occur in MSA. We cannot elaborate on this issue for space reasons, but a discussion appears in the full version of the paper. A second issue relates to the question of whether the ambiguity of sentences such as (1a) is associated with two distinct syntactic structures, or whether it is due to semantico-pragmatic constraints. A similar issue is raised with respect to MG by Terzi (1992), and in a more general cross-linguistic context by Landau (2004). In Landau's framework OC predicates select complement clauses with PRO subjects, while NC predicates select clauses with *pro* subjects. Landau shows that the two null elements can be teased apart by several empirical tests. One such test builds on the observation that PRO is an anaphoric element which supports only a sloppy reading under VP-ellipsis, while lexical pronouns support both sloppy and strict readings. These tests can be used to determine whether the co-reference version of (1a) is really an instance of control.

Naturally, evidence for a distinction between a sloppy or strict reading in this context are difficult if not impossible to come by using a corpus. Moreover, due to the diglossic situation of Arabic, there are no native speakers of MSA, so speakers' intuitions are more suspect than usual. Nevertheless, we consulted some highly proficient speakers with regards to the possible interpretation of sentences such as the following.

- (4) *qarrara muhammad-un [?an yaktub-a r-risa:lat-a] wa laisa ?ahmad-un.*
decided.3SM Muhammad-NOM AN write.3SM-SBJ the-letter-ACC and NEG Ahmad-NOM
'Muhammad decided to write the letter but not Ahmad.'

According to our informants, the only interpretation that is possible is the one where Ahmad decided that he himself will not write the letter (similarly to English), in other words – a sloppy reading. This suggests that the unexpressed subject of the VP in the first conjunct is not a lexical pronoun which has free reference, but a PRO (or bound variable).

The empirical findings reported above are captured by our HPSG-based analysis. We propose that sentences such as (1a) are associated with two structures: *no-control* and *control* (Figures 1 and 2, respectively). The analysis of the control version builds on Alotaibi & Borsley's (2013) analysis of gaps and resumptive pronouns in MSA, and in particular *?anna* complement clauses (see the abbreviated Figure 3). Similarly to the sentential complement of *?anna*, we propose that the subject of controlled *?an* clauses is SLASHED. Control is achieved by the structure-sharing of the INDEX values of the SLASHED subject of the *?an* clauses with the INDEX of its controller – the matrix subject (or object, not shown here). However, unlike *?anna*, *?an* is a subjunctive *marker*, and not a head.

The proposed analysis accounts for the agreement patterns exhibited by the matrix and embedded predicates, which vary according to the relative position of the subject (i.e., full agreement in (S)VO; partial agreement in VSO). Moreover, it provides a way of capturing the control-NC alternation in predicates which allow both structures, and it can also accommodate cases of OC, namely predicates which only allow the control structure, if they are eventually found to exist in the language.

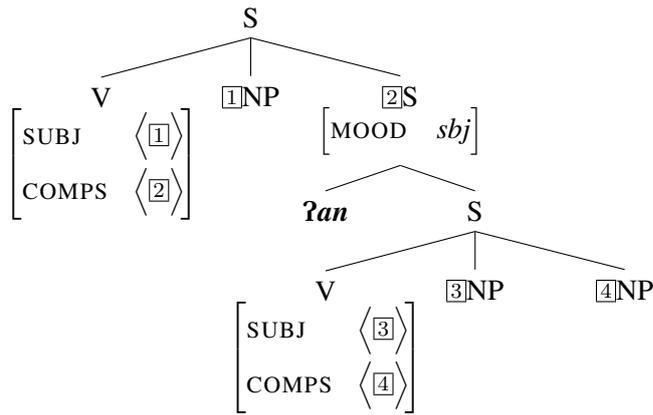


Figure 1: *?an* clause – no control

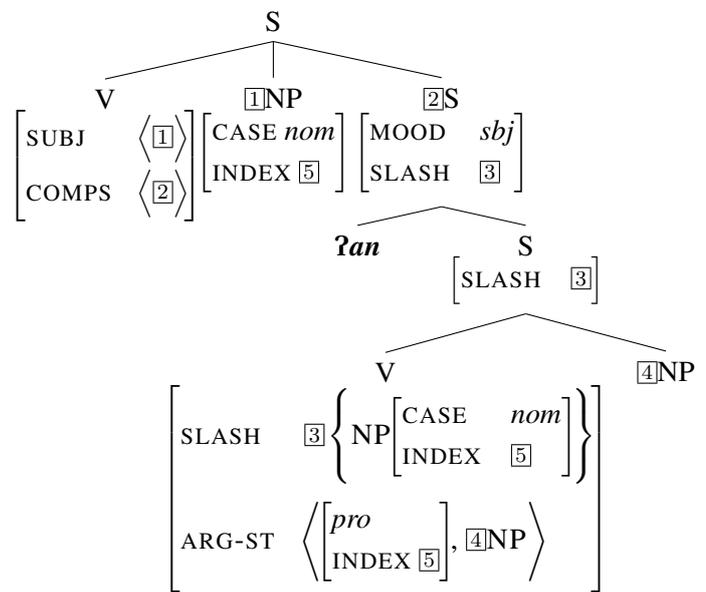


Figure 2: *?an* clause – subject control

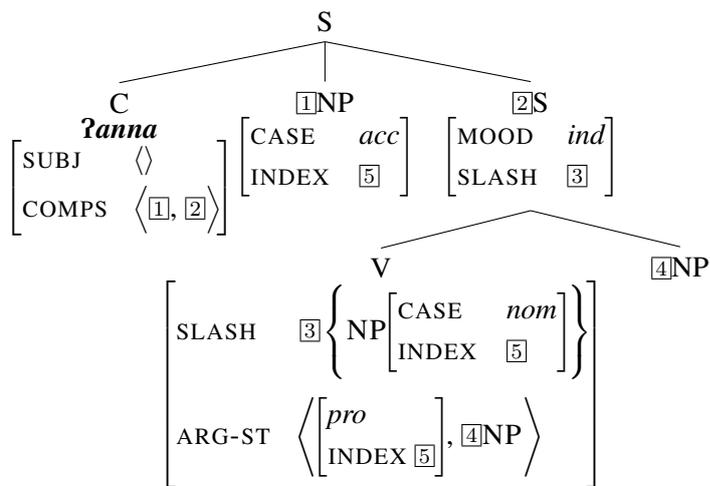


Figure 3: *?anna* clauses (Alotaibi & Borsley, 2013)

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.
- Arts, Tressy, Yonatan Belinkov, Nizar Habash, Adam Kilgarriff & Vit Suchomel. 2014. arTen-Ten: 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.
- Cantarino, Vicente. 1976. *Syntax of modern Arabic prose*, vol. 3. Bloomington: Indiana University Press.
- 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. 2004. The scale of finiteness and the calculus of control. *Natural Language and Linguistic Theory* 22(4). 811–877.
- Landau, Idan. 2013. *Control in generative grammar: A research companion*. Cambridge University Press.
- Persson, Maria. 2002. *Sentential object complements in Modern Standard Arabic*, vol. 2. Lund University.
- 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.
- Terzi, Arhonto. 1992. *PRO in finite clauses: A study of the inflectional heads of the Balkan languages*: City University of New York dissertation.