Seeking control in Modern Standard Arabic

Tali Arad Greshler¹, Livnat Herzig Sheinfux¹, Nurit Melnik² and Shuly Wintner¹

¹Department of Computer Science,
   University of Haifa, Israel

²Department of Literature, Language and the Arts,
   The Open University of Israel

HEADLEX 2016
Overview

1. Background
   - Modern Standard Arabic
   - ?an clauses

2. The study
   - Research questions
   - Co-referring and dis-referring predicates
   - Predictions
   - A corpus study
   - Proposed analysis

3. Conclusions
**Word Order and Agreement**

**VSO**: Unmarked, partial agreement

1. \( qara\, t^f a\, liba\, t-u \quad l-kita\, b-a \)
   \( \text{read.3SF the-students.PF-NOM book-ACC} \)
   ‘The female students read the book.’

**SVO**: Marked, full agreement

2. \( t^f a\, liba\, t-u \quad qara\, na \quad l-kita\, b-a \)
   \( \text{the-students.PF-NOM read.3PF book-ACC} \)
   ‘The female students read the book.’

**pro-drop**: Full agreement

3. \( qara\, at \quad l-kita\, b-a \)
   \( \text{read.3SF book-ACC} \)
   ‘She read the book.’ (Not: ‘They read the book.’)
Complement clauses

\textit{？an}na clauses

(4) ？arafa mu？ammad-un [？an？a l-walad-a sa-yaktubu
knew.3SM Muhammad-NOM that the-boy-ACC will-write.3SM.IND
r-risa:lat-a].
the-letter-ACC

‘Muhammad knew that the boy would write the letter.’

？an clauses

(5) qarrara mu？ammad-un [？an yaktuba r-risa:lat-a].
decided.3SM Muhammad-NOM AN write.3SM.SBJ the-letter-ACC

‘Muhammad decided to write the letter.’
**?AN CLAUSES**

- Verb initial
- Subjunctive form
- The embedded verb carries agreement features
- Ambiguous between co-reference and dis-reference with matrix argument

(6) qarrara muhammad-un [?an yaktuba r-risa\text{\textless }\text{\textgreater }lat-a].
    decided.3SM Muhammad-NOM AN write.3SM.SBJ the-letter-ACC

‘Muhammad decided that he would write the letter.’
**?an clauses: Patterns**

(7) a. qarrara$_i$ muḥammad-un$_i$ [?an yaktuba$_i$/j r-risa:lat-a]. decided.3SM Muḥammad-NOM AN write.3SM.SBJ the-letter-ACC
   ‘Muḥammad$_i$ decided that he$_i$/j would write the letter.’

b. qarrara$_i$/j [?an yaktuba$_i$ muḥammad-un$_i$ r-risa:lat-a]. decided.3SM AN write.3SM.SBJ Muḥammad-NOM the-letter-ACC
   ‘Muḥammad$_i$ decided that he$_j$ would write the letter.’
   ‘He$_j$ decided that Muḥammad$_i$ would write the letter.’

c. qarrara$_i$ muḥammad-un$_i$ [?an taktuba$_j$ (raniat-u$_j$) decided.3SM Muḥammad-NOM AN write.3SF.SBJ (Rania-NOM) r-risa:lat-a].
   the-letter-ACC
   ‘Muḥammad decided that Rania/she would write the letter.’
**?AN clauses: Patterns**

- **Matrix subject**
  - $V1 \gg S \gg [AN \gg V2 \gg O]$
  - $S \gg V1 \gg [AN \gg V2 \gg O]$

- **Embedded subject**
  - $V1 \gg [AN \gg V2 \gg S \gg O]$

- **Two different subjects**
  - $V1 \gg S1 \gg [AN \gg V2 \gg S2 \gg O]$

- **No subjects**
  - $V1 \gg [AN \gg V2 \gg O]$
Research Questions

• Do all verbs in Modern Standard Arabic (MSA) allow for both co-reference and dis-reference in ?an clauses?
• Is there obligatory control in MSA?
• What is the syntactic structure of the construction(s)?
The study: Co-referring and dis-referring predicates

**Sentential complements in MSA**
(Persson, 2002)

- A corpus-based study
- *?an* clauses with an overt embedded subject vs. *?an* clauses in which the subject is deleted under co-reference
- The semantics of the embedding verb determines the preference for dis-reference or co-reference:
  - **Manipulative** predicates prefer co-reference (*force*, *allow*).
  - **Cognitive** predicates (desiderative, commentative, fearing) prefer dis-reference (*want*, *wish*).
  - **Modal** predicates were not included under the assumption that they always require co-reference.
Standard Arabic ?an and ?anna

(Habib, 2009)

- There are no real “control” predicates in MSA.
- All ?an-clauses allow for both co-reference and dis-reference.
Modern Greek Complement Clauses
(Roussou, 2009)

**?oti clauses**

(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.’

**na clauses**

(9) O Kostas matheni [na odhiji]. the Kostas learn.3s PRT drive.3s Kostas is learning (how) to drive.’
Modern Greek na-clauses
(Roussou, 2009)

Obligatory co-reference

(10) O Kostas matheni [na odhiji].
    the Kostas learn.3s PRT drive.3s
    Kostas is learning (how) to drive.’

Co-reference/dis-reference

(11) O Kostas theli [na odhiji].
    the Kostas want.3s PRT drive.3s
    Kostas wants (him) to drive.’
THE CONTROL CONTINUUM
(Roussou, 2009)

+Control

\[ \text{start} \rightarrow \text{can} \rightarrow \text{dare} \rightarrow \text{try} \rightarrow \text{want} \]

–Control
OBLIGATORY CONTROL VS. NO CONTROL
(LANDAU, 2013)

Predicates which select **tensed** complements

- Factives (glad, sad, like,...)
- Propositional (believe, think, claim,...)
- Desideratives (want, prefer, hope,...)
- Interrogatives adjectives (wonder, ask, find out,...)

  (12) Yesterday, John hoped to solve the problem tomorrow.

Predicates which select **untensed** complements

- Implicatives (dare, manage, remember,...)
- Aspectuals (start, stop,...)
- Modals (have, need, may,...)
- Evaluative adjectives (rude, silly,...)

  (13) *Yesterday, John managed to solve the problem tomorrow.*
OBLIGATORY CONTROL VS. NO CONTROL

(LANDAU, 2013)

Landau’s finiteness rule for Obligatory Control:

<table>
<thead>
<tr>
<th></th>
<th>T+</th>
<th>T-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agr+</td>
<td>NC</td>
<td>OC</td>
</tr>
<tr>
<td>Agr-</td>
<td>OC</td>
<td>OC</td>
</tr>
</tbody>
</table>

- If a complement clause is untensed it will enforce obligatory control.
- “There cannot be a language where modal, aspectional and implicative verbs or evaluative adjectives allow an uncontrolled complement subject.” (p. 106)
Seeking control in MSA: Predictions

- Persson (2002): Obligatory co-reference in MSA with modals
- Habib (2009): No obligatory co-reference in MSA
- Roussou (2009): A continuum (aspectuals — desideratives)
- Landau (2013): Control is obligatory when the complement clause is untensed
METHOD

- Corpus based search:
  The 115-million token sample of *arTenTen* corpus of Arabic (Arts et al., 2014), installed in the Sketch Engine (Kilgarriff et al., 2004).
- Representative predicates from Roussou’s continuum and Landau’s classification
  - ?ara:da ‘want’ (desiderative)
  - ?a:wala ‘try’ (implicative)
  - ?ar[u?a ‘dare’ (implicative)
  - ?aqna?a ‘convince’ (manipulative)
  - sama?ha ‘allow’ (manipulative)
  - ?istt?a:qa ‘be able’ (modal)
?araːda ‘want’ (desiderative)

(14) a. ?araːda [ʔan yaʃmala diraːsat-an]
wanted.3SM AN do.3SM.SBJ study-ACC
‘He wanted to conduct a study.’
b. ?araːda [ʔan yakuna r-radd-u watʕaniyy-an]
wanted.3SM AN be.3SM.SBJ the-reaction-NOM national-ACC
‘He wanted the reaction to be national.’
ha:wala ‘try’ (implicative)

(15) a. ha:wala r-ra3ul-u [ʔan yatakallama maʔa-na] tried.3SM the-man-NOM AN speak.3SM.SBJ with-us ‘The man tried to speak with us.’

b. ?inna-na nuha:wilu [ʔan yatahaddaθa sˤamt-u-na] indeed-we try.1P.IND AN speak.3SM.SBJ silence-NOM-our ‘We will try that our silence will speak.’
3aruʔa ‘dare’ (implicative)

(16) a. la: yaʔruʔu raʔul-un [ʔan yaquːla l-ḥaqiːqat-a fi: not dare.3SM man-NOM AN say.3SM.SBJ the-truth-ACC in l-zawaːʒ-i] the-marriage-GEN
‘No man dares to say the truth in the marriage.’

b. lan taʔruʔu [ʔan yakuna raʔy-u-ḥa yair-a never dare.3SF AN be.3SM.SBJ opinion-NOM-her not-ACC musaːnid-in li-l-maʃrab-i] supportive-GEN to-Morocco-GEN
‘She will never dare that her opinion would be non-supportive of Morocco.’
The study

A corpus study

؟اقناة ‘convince’ (manipulative)

(17) a. malaːk qad ʔaqnaʕat waːlid-a-ha ʔan yaʔmura Manak already convinced.3SF father-ACC-her AN order.3SM.SBJ
saːʔiq-a-hu...] driver-ACC-his...

‘Malak had already convinced her father to order his driver...’

b. ʔaqnaʕaː hum [AN yuʕayyina huwa convinced.1P-them AN appoint.3SM.SBJ he.NOM l-ḥukuːmat-a]
the-government-ACC

‘We convinced them that he will appoint the government.’
samaха ‘allow’ (manipulative)

(18) a. iḍā: lam nasmaḥu li-l?ameri:ka:n-i [ʔan yamurru: min if NEG allow.1P to-the-Americans-GEN AN pass.3PM.SBJ from ʔara:dī: t-turkiya ɣada:n] territory Turkish tomorrow
   ‘If we don’t allow the Americans to pass from Turkish territory tomorrow.’

   ‘And his social status does not allow him that his son will be in this place.’


The study

A corpus study

?isttə’a ‘be able’ (modal)

(19) a. lam ?astatʕaʔu [ʔan ?asmaʕa sʕawt-a-hu ?aw ?arar-hu] not be.able.1s AN hear.1s.SBJ voice-ACC-his or see.1s.SBJ-him 'I can’t hear his voice or see him.'

b. lan nastatʕaʔu ʔan tataʔammala l-ḥuku:mat-u never be.able.1P.IND AN bear.3SF.SBJ the-government-NOM ka:mi:l-a l-nafaqa:t-i all-ACC the-expenses-GEN 'We will never be able that the government will bear all the expenses.'
**Interim Summary**

- Instances of dis-reference were found across all semantic categories.
- Although Landau’s tensed-untensed classification holds for MSA, predicates which select untensed complements allowed dis-reference.

<table>
<thead>
<tr>
<th></th>
<th>?an clause complements</th>
<th>Co-reference</th>
<th>Dis-reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>All predicates</td>
<td>249,190</td>
<td>100,763</td>
<td>66,946</td>
</tr>
<tr>
<td>?ara:da ‘want’</td>
<td>20,590</td>
<td>11,034</td>
<td>4,137</td>
</tr>
<tr>
<td>?istt’a:fa ‘be able’</td>
<td>14,568</td>
<td>11,884</td>
<td>299</td>
</tr>
</tbody>
</table>
POSSIBLE STRUCTURE

Co-reference/dis-reference:

Dis-reference:
**Backward “control”**

- The proposed analysis predicts **full agreement** when only the embedded subject is realized:

```
S
  \( V_{\text{pro}} \)
  S_{\text{subj}}
    \(? an \)
      \( S_{\text{subj}} \)
        \( V_{\text{subj}} \)
          \( \text{NP}_{i[\text{nom}]} \)
            \( \text{NP}_{[\text{acc}] \text{P]} } \)
```

- But with co-reference, only cases with **partial agreement** were found.

(20) \textit{taka\text{"u}li\text{"u} l-\text{"u}ila\text{"u}z-i l-ba\text{"u}hi\text{"u}d\text{"u}at-i llati [la costs-NOM the-treatment-GEN the-expensive-GEN that neg yastat\text{"u}i\text{"u} [?an yata\text{"u}hammilu-ha l-fuqara:] ] be.able.3SM AN bear.3SM.SBJ-it the-poor.PM}

‘The costs of the expensive treatment that the poor will not be able to bear’
Only raising verbs of “appropinquation” (proximity, hope, inception) can occur in a backward raising configuration:

\[
\begin{align*}
\text{?awʃakat}/\text{?awʃakna} & \quad [(?\text{an}) \text{ tanẓaḥa} \\
\text{were.about.to.3SF}/\text{3PF (AN)} & \quad \text{succeed.3SF.SBJ} \\
\text{tʃ-ʃa:liba:t-u}] & \quad \text{the-students.PF-NOM} \\
\text{‘The female students were about to succeed.’}
\end{align*}
\]

Prescriptively, partial agreement on the matrix predicate is considered ungrammatical.

But no conclusive evidence of full agreement was found in contemporary texts.

Both patterns are available in SA (traditional and contemporary).
Backward patterns with “control” predicates

- Contrary to Wurmbrand & Haddad (2016), the backward pattern was found also with “control” predicates.
- Corpus-based usage data contradict prescriptive grammar and the internal logic of the grammar.
- We propose that the use of partial agreement in the backward pattern is motivated by analogy to the partial agreement of simple VSO clauses.
- We suspect that this has to do with the fact that MSA is not spoken natively by any speakers.
- This phenomenon remains an open issue.
Conclusions

- Predicates which select *?an* clause complements vary with respect to their “preference” for co-reference vs. dis-reference.
- This variance seems to align with Roussou’s (2009) semantic-based continuum and with Landau’s (2013) tensed/untensed distinction.
- Nevertheless, there is no evidence for the existence of obligatory co-reference or control in MSA *?an* clauses.
Bibliography


