Valency and Information Structure:
A quantitative approach to *from – to* juxtaposition in Arabic

Jiří Milička
Charles University, Prague

Abstract

In Arabic, mutual order of prepositional phrases syntactically dependent on one head is neither fixed nor random. This paper explores the factors affecting the order of prepositions *from* and *to*. Many factors related to syntax, morphology and phonology are taken into account and analysed with a corpus driven approach.

Introduction

Managing valency frames is a basic prerequisite for successful communication and a great effort was put in its description either for purposes of language acquisition or processing. But a sensitive language user feels that the issue of valencies is more complicated than traditional or generative grammar expects. For example one can have the impression that when a verb has two arguments, their order is not random, but neither is it fixed, and that it is to some extent determined by the intention of the speaker or by the information flow or that it may be influenced by phonological patterns of the surrounding text etc. The main purpose of this paper is to convert these subjective impressions into falsifiable hypotheses and to show which of them are proven false. A suitable way to achieve that is to rely on quantification and to "trust the text".

Prepositions *from* (من) and *to* (إلى) were chosen (because of their high frequency) to explore their mutual position in a sentence. In Arabic texts, we can find them in both positions:

From – to construction:	*To – from* construction:
ذهبت من بيتي إلى بيته.	ذهبت إلى بيته من بيتي.
I went *from* my house *to* his house.	I went *to* his house *from* my house.

But classic grammars (e.g. Bakr), modern grammars (e.g. Wright 1859) and recent grammars (e.g. Gulli – Carter – Badawi 2004) do not tell us what factors influence upon the word order of these prepositions.

A monography treating the information structure of Czech texts (Sgall – Buráňová – Hajičová 1980) suggests that the order of arguments is influenced by the information structure, but believing that this is not the only factor which plays an important role, we will not test only this hypothesis, but many others.

The Hypotheses and the Methodology

A random sample of clauses containing *from* and *to* was automatically taken from corpus CLARA. All of the occurrences where *from* and *to* were not syntactically dependent on the same word, were manually eliminated and each clause was manually annotated to describe various qualities related to syntax, morphology and phonology (using a program designed especially for this purpose).

Figure 1: The printscreen of the annotation program

Then each factor was tested by Fisher's exact test. The significance level was determined at 0.05.

Results

General Observation

First, it is important to say that the *from – to* construction is significantly more frequent than *to – from*.

From – to   175
To – from   26

The probability of both constructions having the same frequency is lower than 0.0001 according to the binomial distribution.

Some structuralists would do with the statement that *from – to* is the so called unmarked word order and *to – from* is marked inversion, but we are concerned with searching for factors that affect distribution of these two constructions or at least correlate with it.

1. Information Structure

The first oposition is related to the information structure. The question is whether the *from – to / to – from* construction is in the topic part of the sentence or in the focus part. Here is an example[1]:

ذهبت من بيتي إلى بيته.	من بيتي ذهبت إلى بيته.
I went from my house to his house.	From my house I went to his house.

According to the traditional grammar, the second sentence would be agrammatical. Within our data set (201 occurrences) a *from – to / to – from* construction never occurred in the topic part of the sentence and thus seems to be really ungrammatical.

Another question related to the information structure is whether the dependent is known from the previous context. Example:

الآن أنا في بيتي ، سأذهب من بيتي إلى بيته.	الآن أنا في بيته ، ذهبت إلى بيته من بيتي.
I am in my house. I will go from my house to his house.	I am in his house. I went to his house from my house.

Here are the data:

From (known) to (known)	26
From (known) to (unknown)	21
From (unknown) to (known)	0
From (unknown) to (unknown)	127
To (known) from (known)	8
To (known) from (unknown)	11
To (unknown) from (known)	2
To (unknown) from (unknown)	5

From which we choose two pairs to make a contingency table:

	From+known to + unknown	From+unknown to+known	Total
From - to	21	0	21
To - from	2	11	13
Total	23	11	34

Fisher's test: $p < 0.0001$

---

[1] The examples are not taken from the corpus and they are here just for illustration.

There is a significant tendency that if the *from* argument is known from the previous context and *to* argument is unknown, then the *from* argument would be the first one and vice versa.

2. Phonological Structure

The prepositions can be preceded by a word ending with a consonant or with a vowel. The question is whether *from – to / to – from* construction is affected by this phonological quality.

<div dir="rtl">

ذهبتُ من بيتي إلى بيته.   ذهبَت من بيتي إلى بيته.

</div>

I went from my house to his house.   She went from my house to his house.

The contingency table:

	Cons. + from Voc. + to	Voc + from Cons. + to	Total
From - to	13	36	49
To - from	3	2	5
Total	16	38	54

Fisher's Test: $p = 0.15$

There is not a significant relationship between mutual order of prepositions *from – to* and the fact that the preceding word ends with a consonant / vowel.

A noun determined by the definite article *al-* can syntactically depend on the prepositions. The question is, whether the *from – to / to – from* construction is affected by this article (which changes the phonological quality of the preposition *to* ($il\bar{a} + al- \rightarrow ilal-$).

<div dir="rtl">

ذهبت من السوق إلى بيته.   ذهبت من بيته إلى السوق.

</div>

I went from the market to his house.   I went from his house to the market.

	From + al- to + not (al-)	From + not (al-) to + al-	Total
From - to	16	22	38
To - from	2	6	8
Total	18	28	46

Fisher's Test: $p = 0.45$

There is not a significant relationship between mutual order of prepositions *from – to* and the fact that the following word begins with a definite article or not.

3. Grammatical Categories of Dependent Members

An Arabic noun can be determined not only by the definite article, as was mentioned above, but also by a possessive pronoun or by a genitive in apposition. These alternatives do not affect the prepositions phonologically, thus this phenomena can be counted to syntax.

ذهبت من سوق إلى بيتي.　　　　　　ذهبت من بيتي إلى سوق.

I went from a market to my house.　　I went from my house to a market.

	From + al- to + not (al-)	From + not (al-) to + al-	Total
From - to	8	8	16
To - from	4	2	6
Total	12	10	22

Fisher's Test: $p = 0.65$

We have very few data – in Arabic it is much more common that both of the arguments are determined or both undetermined. Maybe the lack of data is the reason why Fisher's exact test shows that there is not a significant relationship between the mutual order of prepositions *from – to* and the fact that the dependent are determined or not.

4. Morphological Category of Head

Nouns are syntactically very similar to verbs in Arabic, especially the deverbative ones called *mas)dars* and participia.

ذهبت من السوق إلى بيتي.　　　　　انا ذاهب من بيتي إلى السوق.

I went from the market to my　I am going from my house to the market.
house.

	Verb	Noun	Total
From - to	110	65	175
To - from	19	7	26
Total	129	72	201

Fisher's Test: $p = 0.38$

It seems that it does not matter, whether the *from – to* construction depends on a verb or on a noun.

5. Grammatical Categories of Head

Many grammatical categories were explored – tense, aspect and person for verbs, case for nouns and gender and negation for both. Also presence / absence of a subject was examined.

But none of these categories have a significant relationship with order of the examined prepositions.

6. Meaning of Dependent Members

The dependents can have temporal meaning:
From Sunday to Friday   من الأحد إلى الجمعة

Or they can contain some other quantitative information:
From 9 to 10 kg   من 9 إلى 10 كيلو

Or they can be related to a place:
From France to Egypt   من فرنسا إلى مصر

Or they can be related to a person:
From Muhammad to Ali   من محمد إلى علي

Fisher's test did not show any significant results, but it seems that temporal and quantitative meanings tend to be described in the *from – to* word order.

Sometimes we can easily compare which argument is more general. For example a large territory can be considered as a more general term than a little one, a group of people than an individual.

انتقل من دمشق إلى أوربا.   انتقلوا إلى دمشق من كل أنحاء العالم.
He moved from Damascus to Europe.   They moved to Damascus from all over the world.

	From + more general	To + more general	Total
From - to	9	26	35
To - from	16	4	20
Total	25	30	22

Fisher's Test: $p = 0.0002$

There is a significant tendency that if the *from* argument is a more general term than the *to* argument, then the *to* argument would be the first one and vice versa.

If an Arabic pronoun depends on a preposition, it merges with the preposition into one word. For example *to* (إلى) + *he* (هو) → *to him* (إليه). Does the occurrence of these pronouns affect the order of the examined prepositions?

ذهبت من بيتي إليه.   ذهبت إليه من بيتي.
I went from my house to him.   I went to him from my house.

	From + pron. to + non pron.	From + non pron. To + pronoun	Total
From - to	9	0	9
To - from	0	16	16
Total	9	16	25

Fisher's Test: $p < 0.0001$

There is a significant tendency that if the *from* argument consists of a pronoun and the *to* argument does not, then the *from* argument would be the first one and vice versa.

Since pronouns are short and the less general terms tend to be shorter than the more general ones, we can express the previous two hypotheses more generally: *The second argument has significantly more syllables and more words than the first one.*
To test this hypothesis we use Student's test (it is not ideal, but easy to interpret).

	Average length			
	Words		Syllables	
	1	2	1	2
From - to	1.41	1.99	4.38	5.95
To - from	0.42	2.08	1.73	6.58
Together	4.10	6.13	1.30	2.03
p (t-test)	<0.00001		<0.00001	

When speaking about the pair of prepositions *from – to*, the first argument tends to be shorter than the second one, according to paired two tailed Student's test.

7. Meaning of Head

The amount of data does not allow us to find out precisely which words tends to be the head of the *from – to* construction and which one tends to be the head of the opposite one, but we can visualize multisets of these words. *Wordle* is a suitable tool for such visualisations.

Figure 2: *Wordle* visualisation of the multiset of words occurring as a head of *from – to* constructions.
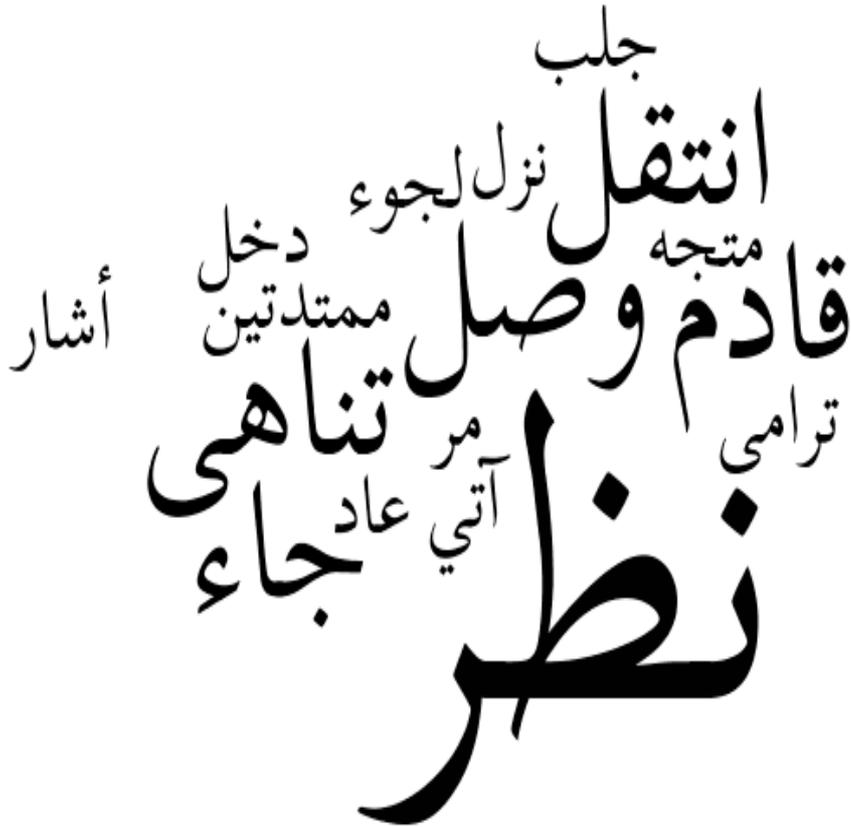

Figure 3: *Wordle* visualisation of the multiset of words occurring as a head of t*o – from* constructions.

We can notice, that *from – to* constructions tend to be dependent on words of نقل root (words with meaning of movement are derived from this root). Also words like *go (out)* and *jump* are striking. *To – from* construction tends to be dependent on words *look, come, coming* etc.

8. Other

Some phrases have a fixed word order. For example:
From the cradle to the grave من المهد إلى اللحد
From the Gulf to the ocean من الخليج إلى المحيط
But testing hypotheses like these needs another methodology and more data.

Of course all the qualities that possibly influence the word order cannot be explored.

Conclusion

During the research we successfully tested the following hypotheses: The *from – to* construction is (in Arabic text) more frequent than the *to – from* construction. Variants *From – to / to – from* are not distributed randomly. The first argument tends to be shorter, less general, known from the previous text and consisting of a pronoun.

This conclusion sounds coherent, because phenomena known from the previous text are (according to our intuition) likely to be expressed by a pronoun and shortly.

The next step in the research should be exploring more prepositions and more languages – hoping it will lead us to some general linguistic hypothesis.

References

Gulli, A. – Carter, M. – Badawi, M. A (2004): Modern written Arabic: Comprehensive Grammar. Routlegdge. 812 p.
Sgall, P. – Hajičová, E. – Buráňová, E. (1980).: Topic – Comment in Czech. Prague.
Bakr, Y. (s. d.): Nusūs fī ´n-nah*wi ´l-'Arabīyi 1, 2 (Texts on Arabic Grammar, 8th – 14th century). Beyrut. 600 + 566 p.
Wright, W. (1859): A Grammar of The Arabic Language. Edinburgh. XVI + 298 + 317 + 450 p.

Corpora

CLARA (Corpus Linguae Arabicae): Corpus of Modern Standard Arabic (about 40,000,000 word tokens)
CLAUDIA (Corpus Linguae Arabicae Universalis Diachronis): Corpus of old Arabic texts (about 300,000,000 word tokens)