

How much structure is needed? The case of the Persian VP

Pegah Faghiri & Pollet Samvelian

Université Sorbonne Nouvelle & CNRS

{pegah.faghiri,pollet.samvelian}@univ-paris3.fr

HeadLex 2016, Warsaw Poland

Goals and Background

Persian syntax: an overview

The Two Object Position Hypothesis (TOPH)

Assessing the TOPH

A flat structure for the Persian VP

Conclusions

Goals and Background

- ▶ We present a series of quantitative studies, including corpus-based and experimental studies, to tease apart between available views of the VP in Persian.
- ▶ Persian is an SOV language with mixed head direction (e.g. head-initial in NP, PP and CP), flexible word order and null pronouns.
- ▶ The prevailing view of the Persian VP initially suggested in generative studies assumes a hierarchical structure with two object positions, mainly motivated by the existence of Differential Object Marking in Persian.
- ▶ Our data do not support this hierarchical view, while they are compatible with a flat structure view of the VP.

Outline

Goals and Background

Persian syntax: an overview

- Word order, etc.

- Nominal determination

- Differential Object Marking

- Different realizations of the DO

The Two Object Position Hypothesis (TOPH)

- The Two Object Position Hypothesis

- Arguments in favor of the hierarchical view

Assessing the TOPH

- Canonical word order in ditransitive constructions

- Semantic Fusion with the Verb

 - Scope Ambiguity

 - Binding Relations

 - Licensing Parasitic Gaps

- Coordination

- Conclusion on the TOPH

A flat structure for the Persian VP

Conclusions

An overview of Persian syntactic properties

- ▶ SOV with free word order in the clausal domain :

- (1)
- a. Puyān Sepide=rā did
Puyan Sepideh=DOM see.PST.3SG
'Puyan saw Sepideh.'
 - b. Sepide=rā Puyān did (OSV)
 - c. Puyān did Sepide=rā (SVO)
 - d. Sepide=rā did Puyān (OVS)
 - e. did Puyān Sepide=rā (VSO)
 - f. did Sepide=rā Puyān (VOS)

- ▶ Null arguments

- (2)
- a. Puyān Sepide=rā did ?
'Did Puyan see Sepideh?'
 - b. na na-did
No NEG-voir.PST.3SG
'No, he didn't see her'

- ▶ No overt marker for definiteness
ex. (*in*) *ketāb* '(This)/the book'
- ▶ Indefiniteness is overtly marked by:
 - ▶ the enclitic =*i*, ex. *ketāb=i* 'a book'
 - ▶ the cardinal *yek*, ex. *yek ketāb* 'a book'
 - ▶ both, ex. *yek ketāb=i* 'a book'
- ▶ A (singular) noun carrying no (formal) determination or quantification can either correspond to a definite NP or to a bare noun (N.B. in the object position, only the latter is possible).
- ▶ Bare nouns are not specified for number and can have a mass reading, ex. *ketāb* 'a book/some books'; they can be either generic/kind-level or existential.

Persian displays Differential Object Marking (DOM) realized with the enclitic =*rā* (*colloquial* =(r)o) :

- ▶ Definite DOs are always marked:

(3) Sara xarguš*(=*rā*) did
Sara rabbit=DOM saw
'Sara saw the rabbit.'

- ▶ However, definiteness is not the only feature triggering DOM (e.g. specificity, topicality, etc.)

- ▶ DOM is considered as a complex phenomenon and cannot be captured by a binary feature (e.g. Lazard, 1982; Meunier and Samvelian, 1997; Ghomeshi, 1997; Lazard et al., 2006)
- ▶ Yet, in most of the works discussed here DOM is claimed to be triggered by a binary [\pm specific] feature (e.g. Karimi, 2003, 2005)
 - ▶ specific DO \rightarrow marked with $=r\bar{a}$
 - ▶ non-specific DO \rightarrow unmarked

Different realizations of the DO

Bare DOs

Indefinite or quantified (unmarked) DOs

Marked DOs

Bare DOs with ou without modifiers

- (4) (man) xarguš did-am
I rabbit saw-1 SG
'I saw a rabbit/rabbits.'
- (5) xarguš=e sefid did-am
rabbit=EZ white saw-1 SG
'I saw a white rabbit/white rabbits.'

Indefinite or quantified (unmarked) DOs

- (6) (yek) xarguš=i did-am
(a) rabbit=INDEF saw-1 SG
'I saw a rabbit.'
- (7) yek xarguš did-am
a rabbit=INDEF saw-1 SG
'I saw a rabbit.'
- (8) čand xarguš did-am
some rabbit=INDEF saw-1 SG
'I saw a few rabbits.'

Marked DOs

- (9) xarguš=rā did-am
rabbit=DOM saw-1 SG
'I saw the rabbit.'
- (10) (yek) xarguš=i=rā did-am ...
(a) rabbit=INDEF=DOM saw-1 SG
'I saw a (particular) rabbit...'

Outline

Goals and Background

Persian syntax: an overview

Word order, etc.

Nominal determination

Differential Object Marking

Different realizations of the DO

The Two Object Position Hypothesis (TOPH)

The Two Object Position Hypothesis

Arguments in favor of the hierarchical view

Assessing the TOPH

Canonical word order in ditransitive constructions

Semantic Fusion with the Verb

Scope Ambiguity

Binding Relations

Licensing Parasitic Gaps

Coordination

Conclusion on the TOPH

A flat structure for the Persian VP

Conclusions

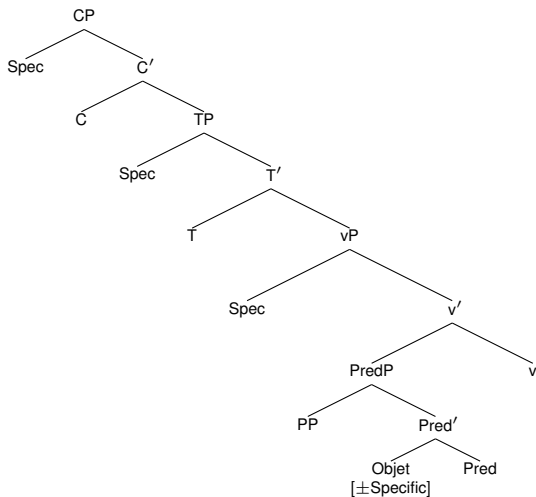
“Two Object Position Hypothesis” (TOPH) I

- ▶ Unmarked DOs have been assumed to be VP internal while *rā*-marked DOs are VP external (cf. Diesing, 1992)
- ▶ Marked (definite or indefinite) and unmarked (bare or indefinite) DOs occur in two different syntactic positions (at spell out), whether base-generated, ex. (11), or as a result of a movement, ex. (12).
(Karimi, 1990; Browning and Karimi, 1994; Ghomeshi, 1997; Karimi, 2005; Ganjavi, 2007; Modarresi, 2014)

- (11) a. [VP DP_[+Specific] [V' PP V]]
b. [VP [V' PP [V' DP_[-Specific] V]]] Karimi (2003, p. 105)

“Two Object Position Hypothesis” (TOPH) II

(12)



Karimi (2005, p. 108)

Arguments in favor of the Two Object Position Hypothesis (TOPH)

Rā-marked and unmarked DOs are claimed to display several syntactic and semantic asymmetries. These asymmetries involve:

- ▶ The relative order with respect to the IO
- ▶ Semantic fusion with the verb
- ▶ Scope ambiguity
- ▶ Binding relations
- ▶ Licensing parasitic gaps
- ▶ Coordinate structures

Arguments in favor of the Two Object Position Hypothesis (TOPH)

Claim 1: Canonical Word order

Theoretical studies and (some) grammars have assumed that in ditransitive constructions, *rā*-marked DOs precede while unmarked DOs (bare or indefinite) follow the IO (Karimi, 1994; Browning and Karimi, 1994; Mahootian, 1997; Rasekhmahand, 2004; Ganjavi, 2007; Windfuhr and Perry, 2009; Roberts et al., 2009, among others):

- (13) a. (S) OD=*rā* OI V
 b. (S) OI OD V

Arguments in favor of the Two Object Position Hypothesis (TOPH)

Claim 2: Semantic fusion with the verb

- ▶ *Rā*-marked DOs are considered as (independent) participants of the event described by the verb and hence semantically autonomous.
- ▶ Unmarked DOs are assumed to be a part of the predicate, and semantically non-autonomous.

According to Karimi (2003) this explains why unmarked DOs, contrary to marked DOs, cannot:

1. take wide scope (and hence cannot trigger scope ambiguity)
2. enter binding relations
3. license parasitic gaps

Arguments in favor of the Two Object Position Hypothesis (TOPH)

Claim 3: Coordinate constructions

Marked and unmarked DOs cannot appear together in a coordination.

Goals and Background

Persian syntax: an overview

- Word order, etc.

- Nominal determination

- Differential Object Marking

- Different realizations of the DO

The Two Object Position Hypothesis (TOPH)

- The Two Object Position Hypothesis

- Arguments in favor of the hierarchical view

Assessing the TOPH

- Canonical word order in ditransitive constructions

- Semantic Fusion with the Verb

- Coordination

- Conclusion on the TOPH

A flat structure for the Persian VP

Conclusions

Claim 1: Canonical Word order

In the neutral/canonical word order *rā*-marked DOs precede while unmarked DOs follow the IO in ditransitive constructions:

- (14) a. Kimea aqlab [_{IO} barā mā] [_{DO} (ye) še'r] mi-xun-e
Kimea often for us a poem IPFV-read-3SG
'It is often the case that Kimea reads poems/a poem for us.'
- b. Kimea aqlab [_{DO} ye še'r=o] [_{IO} barā mā]
Kimea often a poem=DOM for us
mi-xun-e
IPFV-read-3SG
'It is often the case that Kimea reads a (particular) poem for us.'

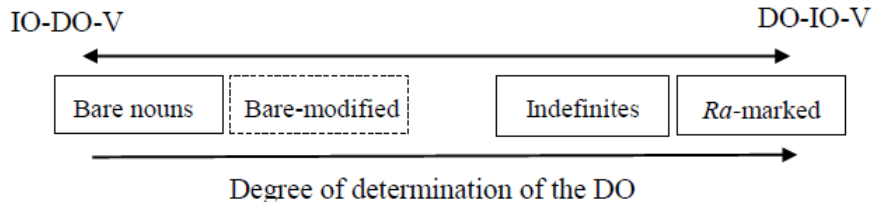
(Karimi, 2003)

Word order Preferences in ditransitive sentences

Objection

Our recent corpus-based and experimental data have invalidated this generalization (Faghiri & Samvelian, 2014; Faghiri et al., 2014; Faghiri et al., forthcoming).

Word order preferences in ditransitive sentences follow a cline rather than being dichotomous



Word order Preferences in ditransitive sentences

- ▶ While marked DOs do have a strong preference for the DO-IO-V order, only bare single-word unmarked DOs have a comparable preference for the reverse order.
- ▶ Indefinite (unmarked) DOs group with marked DOs in preferring the DO-IO-V order overall (but they show a less stronger preference for this position)
- ▶ Bare modified DOs, i.e., bare DOs carrying modifiers, have a significantly less stronger preference for the IO-DO-V.

Word order Preferences in ditransitive sentences

- ▶ Functional factors such as relative length and humanness (or semantic role) are shown to play a significant role in determining the relative word order between the DO and the IO following the “long-before-short” and “animate-before-inanimate” tendencies.
- ▶ Moreover, comparative data suggest that the relative order between the IO and the DO displays a substantial amount of variation comparing to the relative order between the subject and the IO (cf. Faghiri et al., forthcoming)

In conclusion

- ▶ Word order in ditransitive constructions is not a matter of grammatical/strong constraints (i.e. positional syntax) but a matter of preferences and can be accounted for in terms of the interaction of functional factors.
- ▶ Different tendencies observed in our data converge into the general cross-linguistically established tendency to produce more (conceptual) accessible constituents earlier in the sentence.

Claim 2: Semantic fusion with the verb

Unmarked DOs are part of the predicate and form a semantic (and syntactic) unit with the verb (Karimi, 2003).

- (15) Kimea har šab (ye) sib mi-xor-e
Kimea every night (a) apple IPFV-eat.PRS-3SG
'Kimea eats apples (= **does (an) apple eating**) every night.'

An appropriate answer to the question "What does Kimea do every night?"

Claim 2: Semantic fusion with the verb

Unmarked DOs are part of the predicate and form a semantic (and syntactic) unit with the verb (Karimi, 2003).

→ Sentences containing unmarked DOs can only receive an activity/process reading, ex. (16-a), while those containing a marked DO have an eventive reading, ex. (16-b).

- (16) a. (man) *dar do daqiqe / barāye yek sāat sib
I in two minute / for one hour apple
xord-am
eat.PST-1SG
'I ate apples for one hour.'
- b. (man) dar do daqiqe / *barāye yek sāat sib=rā
I in two minute / for one hour apple
xord-am
eat.PST-1SG
'I ate the apple in two minutes.'

Objection

These generalizations only hold for bare unmarked DOs.

- ▶ Bare DOs are highly cohesive with the verb, leading some scholars to consider them as semantically incorporated to the verb.
- ▶ Indefinite unmarked DOs are inarguably referential NPs and can be construed as (independent) entities undergoing the event described by the verb.

Objection

These generalizations hold only for bare unmarked DOs.

- ▶ The “durative adverbial test” argument initially mentioned by Ghomeshi and Massam (1994) applies only to bare DOs.
- ▶ The authors claim that bare DOs are non-referential and as such cannot delimit the event described by the verb and hence are only compatible with adverbials denoting a process.

Objection

These generalizations hold only for bare unmarked DOs.

- ▶ The massive (or cumulative) reading implied by bare DOs is indeed incompatible with a telic reading (cf. e.g. Krifka, 1989, 1992)
- ▶ But, indefinite unmarked DOs are quantized and as such are compatible with a telic reading.

(17) Maryam dar do daqiqe / *barāye yek sāat se=tā
Maryam in two minute / for one hour three=CLF
sib(=rā) xord
apple(=DOM) eat.PST.3SG
'Maryam ate three apples in two minutes.'

Claim 2.1: Scope ambiguity

a. Only *rā*-marked DOs can trigger scope ambiguity when scrambled to the left periphery (Karimi, 2003).

- (18) a. [**har** dānešju=i] [**ye** še'r=ro] bāyād
every student=INDF a poem=DOM must
be-xun-e
SUBJ-read.PRS-3SG
'Every student has to read one poem (out of a specific set).' ($\forall > \exists$)
- b. [**ye** še'r=ro_i] [**har** dānešju=i] t_i bāyād be-xun-e ($\forall > \exists$;
 $\exists > \forall$)

Claim 2.1: Scope ambiguity

b. Unmarked DOs as part of the predicate can never take wide scope over the IO (Karimi, 2003).

- (19) a. [**har** dānešju=i] [**ye** še'r] bāyād
every student=INDF a poem must
be-xun-e
SUBJ-read.PRS-3SG
'Every student must read a poem.' ($\forall > \exists$)
- b. [**ye** še'r_i] [**har** dānešju=i] t_i bāyād be-xun-e ($\forall > \exists$)

Objection

- ▶ These judgments and/or interpretations are not straightforward.
- ▶ Not all studies accept the claim that unmarked DOs cannot take wide scope (Ghomeshi, 1997; Modarresi and Simonenko, 2007; Modarresi, 2014)

(20) hame film=i did-and
everybody movie=INDF watch.PST-3SG
'Everybody watched a movie.' ($\forall > \exists$; $\exists > \forall$)

→ We believe that solid experimental data are needed in order to make any generalization.

Claim 2.2: Binding Relations

Only *rā*-marked DOs can bind an anaphora in the IO position.

- (21) a. man [se=tā bačče-hā=ro]_i [be hamdige_i]
I three=CLF child-PL=DOM to each other
mo'arrefi kardam
introduction do.PST-1 SG
'I introduced three children to each other.'
- b. *man [se=tā bačče]_i [be hamdige_i] mo'arrefi
kardam (Karimi, 2003)

Objection

Contrary to Karimi's claim, unmarked DOs do bind anaphora in the IO position, as shown by the following attested examples.

- (22) [čand varaq kâqaz]_i [be hamdige_i] mangane
some sheet paper to each other staple
mi-kon-e
IPFV-do.PRS-1 SG
'She staples a few sheets of paper together (lit. to each other).'

Semantic Fusion with the Verb II

Binding Relations

- (23) Lidya yeki=ro mi-šnās-e ke [doxtar
Lidya someone=DOM IPFV-know.PRS-3SG that girl
pesar]_i [be ham_i] mo'arrefi mi-kon-e
boy to each other introduction IPFV-do.PRS-3SG
'Lidya knows someone who introduces boys and girls to each
other.'

Claim 2.3: Licensing Parasitic Gaps

Only *rā*-marked DOs can license parasitic gaps

- (24) a. Kimea in ketāb= o_i [qablaz in-ke $-_i$
Kimea this book=DOM before that
be-xun-e] t_i be man dād
SUBJ-read.PRS-3SG to me give.PST.3SG
'Kimea gave me this book before reading (it).'
- b. *Kimea ketāb $_i$ [qablaz in-ke $-_i$ be-xun-e] be man t_i
dād

Karimi (2003)

Objection

We claim on the contrary that unmarked DOs can license parasitic gaps in favorable contexts, e.g. where the DO is discursively prominent.

Our claim is supported by an acceptability ratings experiment:

- ▶ Likert scale from 1 (absolutely unacceptable) to 7 (completely acceptable)
- ▶ 25 participants

Objection

Below an example of our rated sentences:

- (25) man **bastani**_i [bā-inke xeyli -_i dust
I ice-cream even-though very like
dār-am] ba'd-az šām -_i
have.PRS-1SG after dinner
ne-mi-xor-am
NEG-IPFV-eat.PRS-1SG
'I don't eat ice cream after dinner even though I like (ice
cream) very much.'

Mean rate : 6.54

Claim 3: Coordinate structures

Marked and unmarked DOs cannot appear together in a coordination

- (26)
- a. man diruz [in aks=ro] va [in ketāb=ro]
I yesterday this picture=DOM and that book=DOM
xarid-am
buy.PST-1 SG
'Yesterday, I bought this picture and that book.'
 - b. man diruz [aks] va [ketāb] xarid-am
I yesterday picture and book buy.PST-1 SG
'Yesterday, I bought pictures and books.'
 - c. *man diruz [in aks=ro] va [ketāb]
I yesterday this picture=DOM and book
xarid-am
buy.PST-1 SG

Objection

Our acceptability rating experiment clearly contradicts this claim

Details:

- ▶ Acceptability rating from 1 (absolutely unacceptable) to 7 (completely acceptable)
- ▶ Conducted via a web-based questionnaire (on Ibex-Farm): 46 participants
- ▶ Latin Square Design:
 - ▶ Control condition: only a marked DO (DO1)
 - ▶ Coordination of DO1 with an unmarked DO (DO2):
 1. Unmarked-marked (DO2 and DO1) order
 2. Marked-unmarked (DO1 and DO2) order
 - ▶ Postposition (of DO1 or DO2)
 1. Unmarked-marked (DO2 V and DO1) order
 2. Marked-unmarked (DO1 V and DO2) order
- ▶ 20 target items, combined with 40 fillers and 5 practice items

Example of an item:

DO1 : form=e takmil-šode=rā 'the completed form'

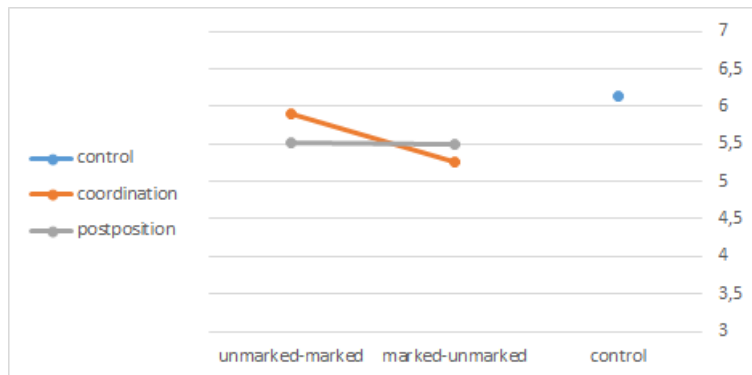
DO2 : yek qat'e aks 'a photo'

(27) barāye sabtenām kāfi ast [form=e takmil-šode=rā]
for registration enough is form=EZ completed=DOM
barāye mā ersāl kon-id
for us send do.PRS-2PL
'To register you only need to send us the completed form.'

(28) a. ... [yek qat'e aks va form=e takmil-šode=rā]
... a piece photo and form=EZ completed=DOM
b. ... [form=e takmil-šode=rā va yek qat'e aks]
... form=EZ completed=DOM and a piece photo

(29) a. ... [DO2] barāye mā ersāl kon-id va [DO1]
... DO2 for us send do-2PL and DO1
b. ... [DO1] barāye mā ersāl kon-id and [DO2]
... DO1 for us send do-2PL and DO2

Results:

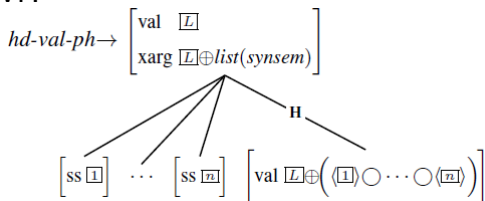


No conclusive arguments in favor of the TOPH

- ▶ Overall our data shows that there is no conclusive empirical evidence in favor of the TOPH.
- ▶ This hypothesis yields erroneous predications with respect to word order preferences, which constitute a cline rather than being dichotomous and can be explained via a set of interacting universal functional principles.
- ▶ If a hierarchical analysis is to be maintained, it should either posit more than two positions, or it should be based on bareness instead of markedness.
- ▶ None of these solutions is satisfactory, given, among other things, that different types of DOs can be coordinated.

The “Two Object Position Hypothesis” has no empirical ground

- ▶ Accordingly, in line with Samvelian (2001); Bonami and Samvelian (2015), we posit a flat structure for the Persian VP.



- ▶ In a flat structure multiple dependents of the verb are realized in the same local tree without constraining their relative order.

- ▶ Word order preferences for different DO types, can be can be accounted for in terms of cross-linguistically valid interacting factors, such as discourse accessibility, definiteness, length (or grammatical weight) and animacy, and stated in terms of the principle of “prominent-first”, pointed out for other SOV languages, such as Japanese (Yamashita and Chang, 2001).

Conclusion

- ▶ We showed that the behavior of the DOs in Persian cannot be accounted for in terms of a hierarchical phrase structure, since the differences between different types of DOs are a matter of cline rather than a dichotomous opposition.
- ▶ Trying to account for these empirical facts by adding more structure, as theoretically appealing as it may seem, not only does not provide an appropriate modeling of data but also makes erroneous predictions.
- ▶ On the contrary, a simplified structure accompanied by few functional principles constitutes a more satisfying option to explore.

- Bonami, O. and Samvelian, P. (2015). The diversity of inflectional periphrasis in Persian. *Journal of Linguistics*, 51(02):327–382.
- Browning, M. and Karimi, E. (1994). Scrambling to object position in Persian. *Studies in scrambling*, pages 61–100.
- Diesing, M. (1992). *Indefinites*. Linguistic Inquiry Monographs. Cambridge, Mass: MIT Press.
- Faghiri, P. and Samvelian, P. (2014). Constituent ordering in Persian and the weight factor. In Pinon, C., editor, *Empirical Issues in Syntax and Semantics 10 (EISS 10)*, pages 215–232. CNRS.
- Faghiri, P., Samvelian, P., and Hemforth, B. (2014). Accessibility and word order: The case of ditransitive constructions in Persian. In Müller, S., editor, *Proceedings of HPSG 2014*, pages 217–237. Stanford: CSLI Publications.
- Faghiri, P., Samvelian, P., and Hemforth, B. (2017). Is there a canonical order in persian ditransitive constructions? corpus based and experimental studies. *Folia Linguistica*, page Forthcoming.
- Ganjavi, S. (2007). *Direct Objects in Persian*. University of Southern California.
- Ghomeshi, J. (1997). Topics in Persian VPs. *Lingua*, 102(2):133 – 167.
- Ghomeshi, J. and Massam, D. (1994). Lexical/syntactic relations without projection. *Linguistic Analysis*, 24(3-4):175–217.
- Karimi, S. (1990). Obliqueness, specificity, and discourse functions: Ra in Persian. *Linguistic Analysis*, 20:139–191.
- Karimi, S. (1994). Word-order variations in contemporary spoken Persian. *Persian Studies in North America*, pages 43–73.
- Karimi, S. (2003). On object positions, specificity, and scrambling in Persian. In Karimi, S., editor, *Word Order and Scrambling*, pages 91–124. Blackwell Publishing Ltd.
- Karimi, S. (2005). *A Minimalist Approach to Scrambling: Evidence from Persian*. Studies in Generative Grammar. Mouton De Gruyter.
- Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event semantics. In Renate Bartsch, J. v. B. and van Emde Boas, P., editors, *Nominal Reference, Temporal Constitution and Quantification Semantics and Contextual Expressions*, pages 75 –115.

Bibliographie II

- Krifka, M. (1992). Thematic relations as links between nominal reference and temporal constitution. In Sag, I. A. and Szabolcsi, A., editors, *Lexical matters*, pages 29–53. Stanford University.
- Lazard, G. (1982). Le morphème *rā* en persan et les relations actanciennes. *Bulletin de la Société de Linguistique de Paris Paris*, 77(1):177–208.
- Lazard, G., Richard, Y., Hechmati, R., and Samvelian, P. (2006). *Grammaire du persan contemporain*. Bibliothèque iranienne. Institut français de recherche en Iran.
- Mahootian, S. (1997). *Persian*. New York: Routledge.
- Meunier, A. and Samvelian, P. (1997). La postposition *rā* en persan : son rôle dans la détermination et sa fonction discursive. *Cahiers de Grammaire*, 25:187–232.
- Modarresi, F. (2014). *Bare nouns in Persian: Interpretation, Grammar and Prosody*. PhD thesis, Humboldt-Universität zu Berlin.
- Modarresi, F. and Simonenko, A. (2007). Quasi noun incorporation in Persian. In *LingO*, pages 181–189.
- Rasekhmahand, M. (2004). Jāyghā maf'ul mostaqim dar fārsi [the position of the direct object in Persian]. *nāme farhangestān*, 6(4):56–66.
- Roberts, J. R., Barjasteh Delforooz, B., and Jahani, C. (2009). *A study of Persian discourse structure*. Studia Iranica Upsaliensia, 12. Acta Universitatis Upsaliensis.
- Samvelian, P. (2001). Le statut syntaxique des objets nus en persan. *Bulletin de la Société de Linguistique de Paris*, 96(1):349–388.
- Windfuhr, G. L. and Perry, J. R. (2009). Persian and Tajik. In Windfuhr, G. L., editor, *The Iranian languages*, Routledge Language Family Series, pages 416–544. London: Routledge.
- Yamashita, H. and Chang, F. (2001). “Long before short” preference in the production of a head-final language. *Cognition*, 81(2):B45–B55.