



Phonological Phrasing Affects Vocabulary Insertion: Evidence from the Persian copula

SAMAN MEIHAMI

MOHSEN MAHDAVI MAZDEH

UNIVERSITY OF ARIZONA

NACIL 4, MAY 2025

PRELIMINARIES

- The Persian present indicative copula has two forms:

(1)		Short form	Full form
	1SG	=am	hast=am
	2SG	=i	hast=i
	3SG	=e	hast=∅
	1PL	=im	hast=im
	2PL	=in/=id	hast=in
	3PL	=an	hast=an

- Following previous work (Béjar & Kahnemuyipour 2017; Okubo & Nomoto 2023), we take the short form to be agreement inflection and assume that it is preceded by a null copula.
 - The only potential exception is 3SG. But even there, we may still assume =e is the regular present tense agreement inflection.
- In other words, *hast* and ∅ are allomorphs.

PRELIMINARIES

- In most contexts, the two forms are both grammatical.

(2) a. xošhâl hast=am.

happy be=1SG

“I’m happy.”

b. xošhâl=ø=am.

happy=be=1SG

“I’m happy.”

- In the special case of the third person singular, the full form is sometimes ungrammatical or highly marked.

(3) a. maryam mariz=ø=e.

Maryam sick=be=3SG

“Maryam is sick.”

b. ??Maryam mariz hast=ø.

Maryam sick be-3SG

“Maryam is sick.”

We do not discuss the special 3SG case today.

- Finally, there are contexts where the short form is ungrammatical.

(4) a. fardâ hast=i?

tomorrow be=2SG

“Are you [going to be] there tomorrow?”

b. *fardâ=ø=yi?

tomorrow=be=2SG

Intended: “Are you [going to be] there tomorrow?”

This is the focus of our attention.

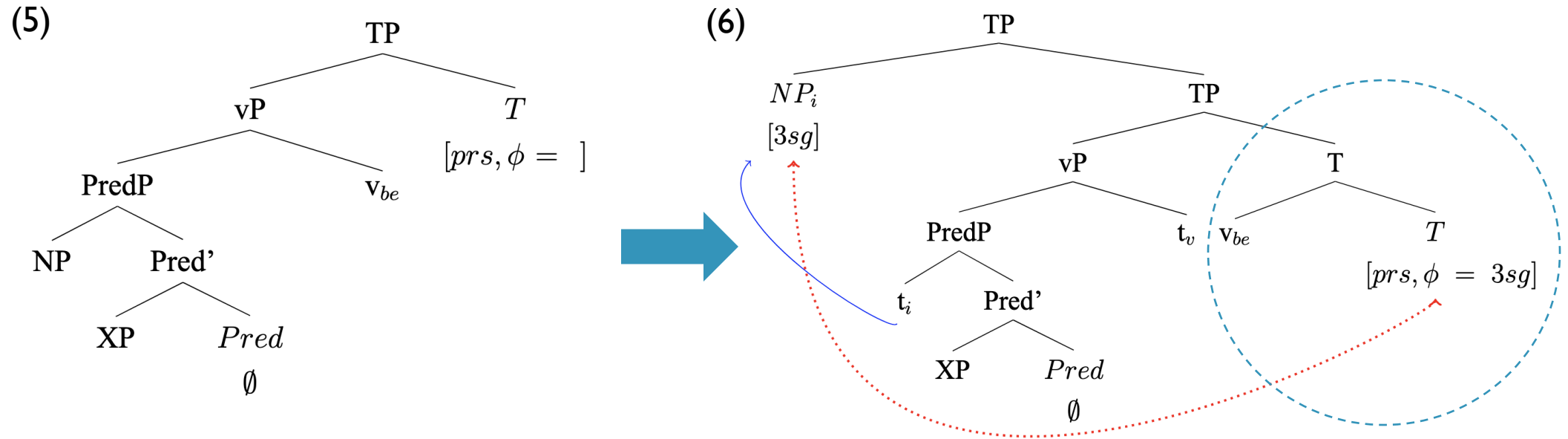
THE RESEARCH QUESTION

- What governs the distribution of the full vs. short forms of the Persian “to be” verb?
- Outline:
 - The distribution is affected by whether there is a predicate in the vP (as noted by Béjar & Kahnemuyipour 2017).
 - The distribution is also affected by focus (as noted by Okubo & Nomoto 2023).
 - The distribution cannot be accounted for through stress (contrary to Okubo & Nomoto 2023).
 - The decisive factor is whether the copula must constitute a Prosodic Word or not.
 - The Persian copula facts suggest that Vocabulary Insertion must have access to prosodic constituency information (contra Embick & Noyer 2001, in line with Ackema & Neeleman 2003, Chung 2003, Henderson 2012).

THE SYNTACTIC STRUCTURE

Assumptions to begin with:

- Minimalist Program (Chomsky 1995)
- Distributed Morphology (Halle and Marantz 1993; Embick & Noyer 2001)
- Bottom-Up Vocabulary Insertion (Bobaljik 2017)



(Ghomeshi 2025; Okubo & Nomoto 2023)

(Béjar & Kahnemuyipour 2017)

THE DATA

- At a more descriptive level, when is the short form disallowed?
 - 1. When there is no predicate in the entire sentence
 - Existence and presence (Béjar & Kahnemuyipour 2017)
 - Ellipsis (Okubo & Nomoto 2023)
 - 2. When the predicate follows the copula
 - 3. When the predicate has moved above other elements
 - 4. When the copula has narrow focus (Okubo & Nomoto 2023)
- In all but one of the cases, no predicate precedes the copula in the vP.

CASES WHERE THE SHORT FORM IS DISALLOWED

- 1. When there is no predicate in the sentence, including:

- Cases denoting existence:

(7) a. edâlat [_{VP} hast=∅]
justice be=3SG
“Justice exists.”

b. *edâlat=∅=e
justice=be=3SG
intended: “Justice exists”

- Cases denoting presence:

(8) a. fardâ [_{VP} hast=i?]
tomorrow be=2SG
“Are you [going to be] there tomorrow?”

b. *fardâ=∅=yi?
tomorrow=be=2SG
intended: “Are you [going to be] there tomorrow?”

- Cases involving ellipsis:

(9) [Are they angry?]
a. agar hast=an man xabar na-dâr-am.
if be=3PL I news NEG-have-1SG
“If they are, I’m not aware of it.”

b. *agar=∅=an man xabar na-dâr-am.
if=be=3PL I news NEG-have-1SG
intended: “If they are is, I’m not aware of it.”

CASES WHERE THE SHORT FORM IS DISALLOWED

- 2. When the predicate follows the copula, including:
 - Cases where the predicate is long or appears last for information-structure reasons

(10) esm=e ketâb hast-ø “dar jostoju=ye zaman=e azdastrafte”
name-EZ book be-3SG in search=EZ time-EZ lost

“The name of the book is *In search of lost time.*”

(11) *esm=e ketâb=ø=e “dar jostoju=ye zaman=e azdastrafte”
name-EZ book=be=3SG in search=EZ time=EZ lost

“The name of the book is *In search of lost time.*”

- Poetic language

(12) a. hast=ø šab
be=3SG night

“It’s night time.” (*Nima Yushij*)

b* .ø=e šab
be=3SG night

Intended: “It’s night time.”

CASES WHERE THE SHORT FORM IS DISALLOWED

- 3. When the predicate has moved above other elements.

(13) a. mariz agar [_{VP} hast-∅] be-gu na-yâ-d.
sick if be=3SG IMP-say NEG-come-3SG
“If s/he is sick, tell her/him not to come.”

b. *mariz agar=∅=e be-gu na-yâ-d.
sick if=3SG IMP-say NEG-come-3SG
intended: “If s/he is sick, tell her/him not to come.”

(14) a. mariz šâyad [_{VP} hast=i.]
sick maybe be=2SG
“Maybe you’re sick.”

b. *mariz šâyad=∅=i.
sick maybe=2SG
intended: “Maybe you’re sick.”

(15) a. puldâr nist=am vali sâlem ke [_{VP} hast=am]
rich NEG.be=1SG but healthy FOC be=1SG
“I am not rich, but I am healthy.”

b. *puldâr nist=am vali sâlem ke ∅=am
rich NEG.be=1SG but healthy FOC be=1SG
intended: “I am not rich, but I am healthy.”

(With changes, based on Okubo & Nomoto 2023)

CASES WHERE THE SHORT FORM IS DISALLOWED

- 4. When the copula has narrow focus.

(The double underline indicates focus)

(16) a. negarân [_{VP} hast-am], vali kam.
worried be=3SG but little
“I AM worried, but [only] a little.”

b. *negarân=~~ø~~=am, vali kam.
worried=1SG but little
intended: “I AM worried, but [only] a little.”

UNIFYING THE CASES

- Overall, the short form is disallowed in two environments:
 1. When no predicate precedes the copula in the vP.
 2. When the copula has narrow focus.
- What unifies these cases?

THE STRESS-BASED ANALYSIS

- Okubo & Nomoto 2023: The short form is disallowed when the copula is stressed.
 - When the copula has narrow focus, it necessarily receives stress. Thus, the full form is required.

(17) negarân hast-am, vali kam.
 worried be=3SG but little
 “I AM worried, but [only] a little.”

- They argue that when the copula is the first element in the vP, it receives phrasal stress (relying on Kahnemuyipour 2003). Thus, the full form is required.

(18) ()_φ ()_φ
 ()_ω ()_ω ()_ω ()_ω
 man [vP xošhâl =∅ =am].
 I happy=∅=1SG
 “I’m happy.”

(19) ()_φ ()_φ ()_φ
 ()_ω ()_ω ()_ω ()_ω
 man fardâ [vP **hast=am**].
 I tomorrow be=1SG
 “I’m [going to be] there tomorrow.”

THE STRESS-BASED ANALYSIS

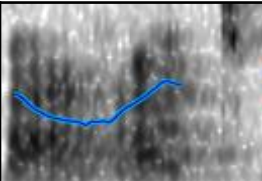
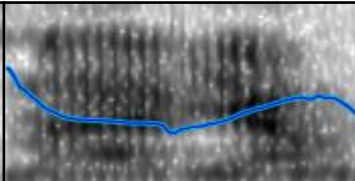
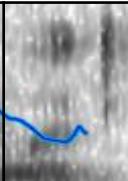

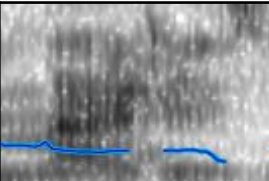
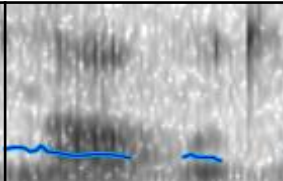
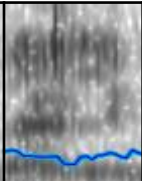
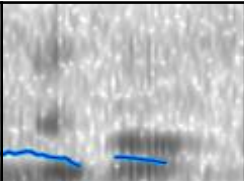
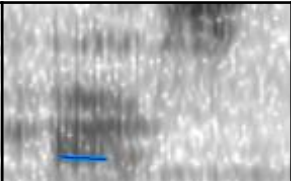
- There are two concerns regarding Okubo & Nomoto's (2023) analysis;
 - **Theoretical concern:** In many approaches to morphology-phonology interface, Vocabulary Insertion precedes stress assignment.
 - This is not a concern if we assume OT-like parallel morphology-phonology machinery.
 - Neither is this a concern if we assume that phonology acts like a filter on syntax (e.g., Anttila 2016)
 - **Empirical concern:** The prohibition of the short form does not always coincide with stress.
 - Sometimes there is prohibition, but no stress.
 - Sometimes there is stress, but no prohibition.

THE EMPIRICAL CONCERN

- Sometimes, the short form is prohibited but the copula is unstressed:
 - 1. Post-focal deaccenting (for more on deaccenting after focus, see Rahmani et al. 2018)

(20) a. faqat maryam bud ke mariz na-bud vali mi-goft- \emptyset hast= \emptyset .
 only Maryam was that sick NEG-was but DUR-say.PST-3SG be=3SG
 “It was only Maryam who wasn’t sick but said that she was.”

b. *faqat maryam bud ke mariz na-bud vali mi-goft- \emptyset =e.
 only Maryam was that sick NEG-was but DUR-say.PST-3SG 3SG
 “It was only Maryam who wasn’t sick but said that she was.”

								
faqat	maryam	bud	ke	mariz	nabud	vali	migoft	hast
LH* H-	LH* L- L%							

THE EMPIRICAL CONCERN

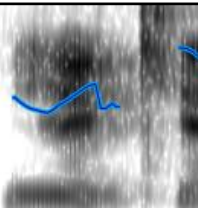
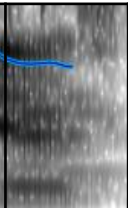
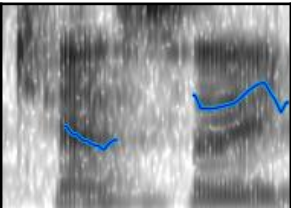
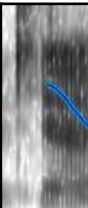
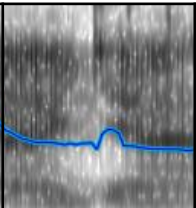
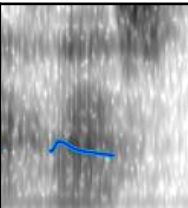
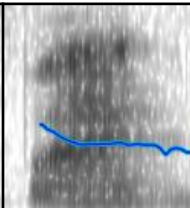
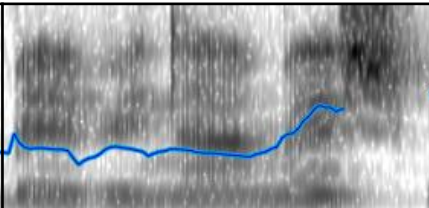
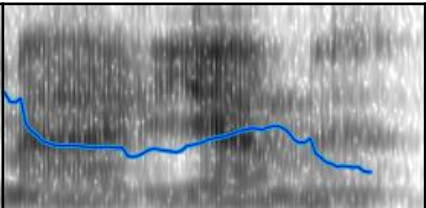
- Sometimes, the short form is prohibited but the copula is unstressed:
 - 1. Post-focal deaccenting
 - 2. Deaccenting in relative clauses (for more on RC deaccenting, see Sadat-Tehrani 2007)

(21) a. yek=i az kes-â=yi ke agar hast=∅ bâyard be-bin-am=eš maryam=∅=e.
 one=INDF from person-PL=INDF that if be=3SG must SBJV-see-1SG=3SG Maryam=be=3SG

“One of the people whom I should meet if they’re around is Maryam.”

b. *yek=i az kes-â=yi ke agar=∅=e bâyard be-bin-am=eš maryam=∅=e..
 one=INDF from person-PL=INDF that if=be=3SG must SBJV-see-1SG=3SG Maryam=be=3SG

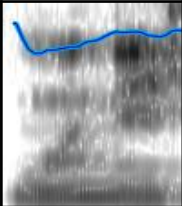
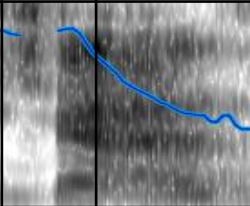
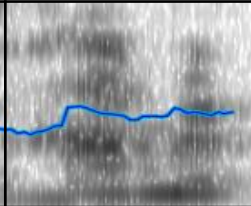
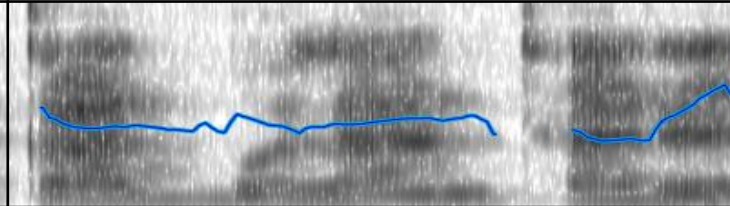
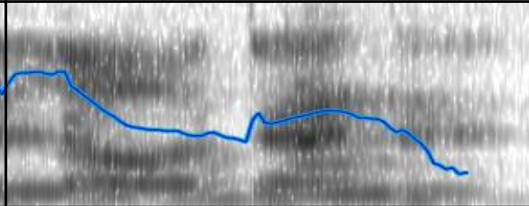
“One of the people whom I should meet if they’re around is Maryam.”

								
yeki	az	kas-â=yi	ke	agar	hast-∅	bâyard	bebinam=eš	maryam=e
LH* H-	LH* H-	LH* H-					LH* L- L%	

THE EMPIRICAL CONCERN

- There are also cases where there is stress even though the short form is allowed.
 - Again, this involves relative clauses.

(22) un=i ke az hame qadboland-tar=ø=e lâdan=e.
 that=INDF that than everyone tall-CMP=ø=INDF Ladan=3SG
 “The one that’s taller is Ladan.”

								
un=i	ke	az	hame	qadboland-tar=ø=e			lâdan=e	
H* H-		LH* H-					LH* L- L%	

- In such cases, even though the syllable containing the copula is stressed, the copula is part of a bigger unit whose stresses are flattened and behaves like a giant word for stress purposes.

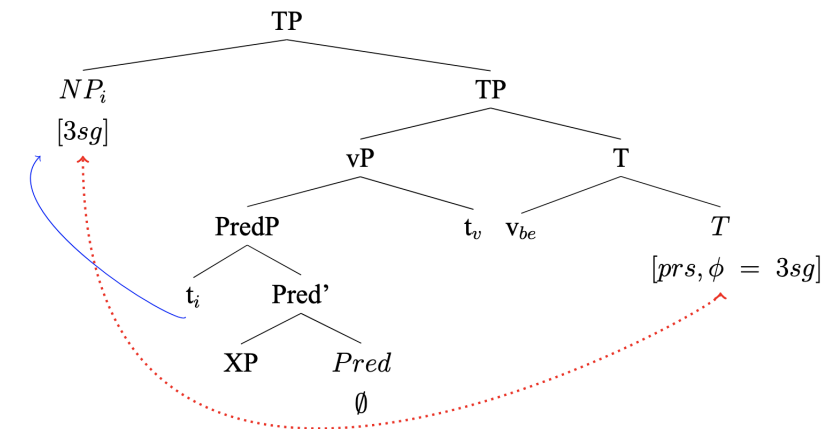
THE EMPIRICAL CONCERN

- We conclude that while Okubo and Nomoto's (2023) intuition is on the right track, stress per se cannot account for the distribution of the short form of the Persian copula.
- Instead, we argue that the decisive factor is whether the copula is forced to mark the beginning of a Prosodic Word.

OUR PROPOSAL

- The Persian inflectional agreement suffixes are function words and inherently unable to act as independent Prosodic Words (Hosseini 2012, based on Selkirk 1995).
- The left edge of the vP marks the beginning of a Prosodic Phrase (and thus a new Prosodic Word).
 - This occurs with natural assumptions under all major theories of prosodic mapping.
 - Edge-based mapping; left edges of XPs (Kahnemuyipour 2003)
 - Edge-based mapping; right edges of XPs
 - Match theory (Selkirk 2011)

(23)



OUR PROPOSAL

- When the copula is at the beginning of a Prosodic Word, the short form is prohibited.

(24) ()_φ ()_φ
 ()_ω ()_ω ()_ω
 *sobh=e fardâ [_{VP} ø=am]
 tomorrow=EZ morning be=1SG
 “Tomorrow morning, I’ll be there.”

IMPLEMENTATION

- How do we implement the idea that allomorph selection between \emptyset and *hast* relies on Prosodic Wordhood?
- In a global parallel implementation of the morphology-phonology interface (e.g., OT), things are straightforward.
 - Allomorphy selection and the creation of prosodic boundaries occur in parallel, and the globally optimal solution is selected.
- What happens under local derivational accounts (e.g., Distributed Morphology)?

IMPLEMENTATION

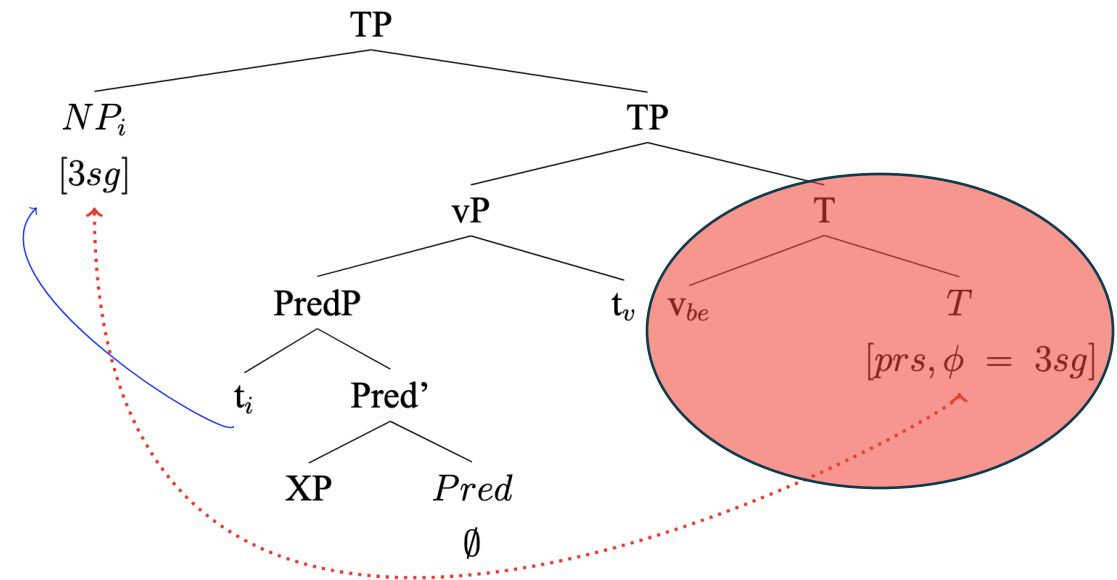
- According to Embick and Noyer (2001), Vocabulary Insertion precedes the formation of prosodic structure.
- The Persian copula data do not support this.
- Instead, the facts are compatible with the idea that some prosodic structure is formed before Vocabulary Insertion (Ackema & Neeleman 2003; Chung 2003; also see Henderson 2012).

$$(25) \quad \begin{array}{ccc} \phi(\quad)_{\phi} & \phi(\quad)_{\phi} & \\ \omega(\quad)_{\omega} & \omega(\quad)_{\omega} & \\ [NP \ N] & [_{vP} \ v_{be} \quad T] & \end{array}$$

- The insertion rules are then as follows:

$$(26) \quad \begin{array}{l} \bullet \ v_{be} \leftrightarrow \text{hast} / \omega(\quad) \\ \bullet \ v_{be} \leftrightarrow s / \omega(\dots[+\text{syl}, +\text{low}] \quad) \\ \bullet \ v_{be} \leftrightarrow \emptyset / \omega(A \quad) \\ \bullet \ T [+PRS, \phi = 3sg] \leftrightarrow \emptyset / v_{\text{hast}} \quad \\ \bullet \ T [+PRS, \phi = 3sg] \leftrightarrow \emptyset / v_s \quad \\ \bullet \ T [+PRS, \phi = 3sg] \leftrightarrow e \end{array}$$

(27)



IMPLEMENTATION

- The optionality in normal cases follows from the optionality of PWord boundary formation.
 - Cf. Ackema & Neeleman's (2003) optional rule for bundling pronouns and words preceding them into a single Pword.

FOCUS

- Recall that the short form is prohibited in focus positions as well.

(The double underline indicates focus)

(28) a. negarân [_{VP} hast-am], vali kam.
worried be=3SG but little
“I AM worried, but [only] a little.”

b. *negarân=~~ø~~=am, vali kam.
worried=1SG but little
intended: “I AM worried, but [only] a little.”

- How do we unify the focus case with the other cases?
 - We suggest that during the formation of prosodic structure, the left edge of a focused elements must be aligned with the left edge of a Prosodic Word.
 - The same phenomenon is seen in Armenian (Doalatian 2022).
 - Generally, the alignment of focused elements with prosodic constituents is well-documented (see Truckenbrodt 1999; Féry 2013)

CONCLUSION

- The present indicative Persian copula has an allomorphy between \emptyset and *hast*.
- We showed that allomorph selection relies on Prosodic Wordhood.
- Assuming a DM model, this supports the line of research that argues that Prosodic Wordhood affects Vocabulary Insertion.
- Moreover, under our account, all four cases where \emptyset is not allowed are unified.

REFERENCES

- Ackema, P., & Neeleman, A. (2003). Context-sensitive spell-out. *Natural Language & Linguistic Theory*, 21(4), 681–735.
- Béjar, S. and Kahnemuyipour, A., 2017. Non-canonical agreement in copular clauses¹. *Journal of linguistics*, 53(3), pp.463–499.
- Bobaljik, J.D., 2017. Distributed morphology. In *Oxford research encyclopedia of linguistics*.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chung, Sandra. 2003. The syntax and prosody of weak pronouns in Chamorro. *Linguistic Inquiry* 34: 547–599.
- Dolatian, H. (2022). An apparent case of outwardly-sensitive allomorphy in the Armenian definite. *Glossa: a journal of general linguistics*, 7(1).
- Embick, D. and Noyer, R., 2001. Movement operations after syntax. *Linguistic inquiry*, 32(4), 555–595.
- Féry, C. (2013). Focus as prosodic alignment. *Natural Language & Linguistic Theory*, 31, 683–734.
- Ghomeshi, J. (2025, January). The features of the Iranian copula: a synchronic view [Keynote address]. Tenth International Conference on Iranian Linguistics (ICIL10), Viterbo, Italy.
- Henderson, R. (2012). Morphological alternations at the intonational phrase edge: The case of K'ichee'. *Natural Language & Linguistic Theory*, 30, 741–787.
- Hosseini, A. (2012). The prosodization of function words in Persian. *Orientalia Suecana*, 61, 126–137.
- Kahnemuyipour, A. (2003). Syntactic categories and Persian stress. *Natural Language & Linguistic Theory*, 21(2), 333–379.
- Okubo, W. and Nomoto, H., 2023. A null stem analysis of Persian copular verbs. In *Advances in Iranian Linguistics II* (pp. 231–262). John Benjamins Publishing Company.
- Rahmani, H., Rietveld, A. C. M., & Gussenhoven, C. (2018). Post-focal and factive deaccentuation in Persian.
- Sadat-Tehrani, N. (2007). *The intonational grammar of Persian*. PhD dissertation. University of Manitoba.
- Selkirk, Elisabeth (1995). “The Prosodic Structure of Function Words.” In: Beckman, Jill, Dickey, Laura W., and Urbanczyk, Suzanne (eds), *Papers in Optimality Theory*. University of Massachusetts Occasional Papers 18. Amherst, Mass.: GLSA, pp. 439–469.
- Selkirk, E. (2011). The syntax-phonology interface. *The handbook of phonological theory*, 435–484.
- Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic inquiry*, 30(2), 219–255.