



Corpus-Based Typology and Word Order in Persian

Mohammad Rasekh-Mahand¹ 

Professor, Department of Linguistics, Bu-Ali
Sina University, Hamedan, Iran

Abstract

This study examines word order typology in Persian through the lens of corpus-based typology, challenging traditional grammar-based classifications. While grammar-based studies have characterized Persian as a ‘mixed’ language—exhibiting both verb-final and verb-initial features—they fail to capture the gradient nature of word order variation. Reviewing recent quantitative corpus studies across multiple syntactic parameters, this research demonstrates that Persian shows considerable disorder in head-dependent ordering but remarkable stability in subject-object ordering (over 90% subject-before-object). The presence of differential object marking (*rā*) does not enable flexible subject-object order, contrary to expectations. Furthermore, Persian places nominal objects before verbs in approximately 82% of cases, pronominal objects before verbs in 92% of cases, and adjectives after nouns in 95% of cases. The study advances areal explanations for these mixed properties, situating Persian within the Western Asia Transition Zone, where centuries of contact between Turkic (verb-final) and Semitic (verb-initial) languages have shaped Iranian languages. This regional perspective explains why Persian exhibits apparently contradictory typological features without necessarily being in transition toward a different type. The findings highlight the superiority of corpus-based approaches for revealing systematic patterns in word order variation and emphasize the need for expanded Persian corpora.

Keywords: Areal Typology, Corpus, Persian, Verb-Final, Verb-Initial, Word Order

1. rasekh@basu.ac.ir (Corresponding Author)

How to cite: Rasekh-Mahand, M. (2025). Corpus-Based Typology and Word Order in Persian. *Language and Linguistics*, 21(42), 1- 60. doi: [10.30465/lsi.2026.54514.1844](https://doi.org/10.30465/lsi.2026.54514.1844)

1. Introduction

This article presents a comprehensive investigation into the typology of word order in the Persian language, with a specific focus on the emerging paradigm of corpus-based typology. Traditional grammar-based typological studies, following the foundational work of Greenberg (1963) and Dryer (1992), have typically categorized languages into discrete typological classes based on the relative order of core syntactic elements, particularly the positioning of the verb with respect to subject and object. Persian has posed a persistent challenge to such categorical classifications, as it exhibits mixed characteristics that do not align neatly with either verb-initial or verb-final language types.

Research Question(s)

Does Persian conform to a consistent word order typology, and if not, what factors explain its mixed nature? Through a systematic review of corpus-based typological studies, this article demonstrates that Persian's word order properties are more complex and gradient than previously recognized, showing high disorder in some syntactic domains and remarkable stability in others.

2. Literature Review

Greenberg's (1963) seminal work established implicational universals that link word order patterns across languages. Dryer's (1992) subsequent elaboration identified strong correlations between verb-object positioning and other syntactic properties, allowing languages to be classified as either verb-initial/head-initial or verb-final/head-final. Persian was found to present an inconsistent profile when evaluated against these correlation pairs.

Building on Dabir Moghaddam's (2013, 2018) comprehensive application of Dryer's parameters to Persian, the article shows that Persian demonstrates: Verb-final features in only four parameters, including the placement of adpositional phrases, manner adverbs, the copula, and wh-words. Verb-initial features in nine parameters, including relative clauses, genitive constructions, complementizers, and interrogative particles. Intermediate or mixed behavior in four additional parameters (comparative constructions, auxiliary placement, tense/mood markers, and object positioning). This distribution reveals that Persian cannot be confidently assigned to either typological category. Dabir-Moghaddam (2013) proposed two competing hypotheses: Persian is either undergoing a slow typological shift from verb-

3 | Rasekh-Mahand

final to verb-initial, or it has inherently free word order. Examination of historical data from Old and Middle Persian supports the shift hypothesis, showing a gradual but consistent movement toward verb-initial patterns across historical periods.

3. Methodology

The last two decades have witnessed a paradigm shift in typological research, driven by the proliferation of linguistic corpora. This development has given rise to what is now termed corpus-based typology, token-based typology, or gradient typology, in contrast to traditional feature-based or grammar-based approaches (Haig et al., 2022; Levshina, 2019, 2021; Levshina et al., 2023). Key innovations of this approach include: direct access to primary data, moving beyond second-hand judgments; context-sensitive analysis grounded in actual language use; quantitative measurements that capture the gradient nature of linguistic features; Discovery of new cross-linguistic generalizations and re-evaluation of established universals; enhanced explanatory power for cognitive, functional, diachronic, and areal accounts.

Several corpus types have proven valuable for typological investigations, including parallel corpora (e.g., Universal Declaration of Human Rights translations, Europarl, film subtitle corpora), comparable corpora (e.g., Multi-CAST), and corpora with unified annotation (e.g., Universal Dependencies). These resources enable systematic cross-linguistic comparison and quantitative analysis of word order patterns.

4. Results

Head-Dependent Ordering: Futrell et al. (2015) examined dependency length minimization across 34 languages, measuring the disorder (entropy) in head-dependent ordering. Persian occupies a position in the lower half of their entropy scale, indicating considerable inconsistency in whether it is consistently head-initial or head-final across different construction types. While the language places verbs sentence-finally, it cannot be characterized as consistently head-final. Notably, both Turkish (head-final) and Arabic (head-initial) show approximately half the disorder of Persian, demonstrating greater typological consistency.

Subject-Object Ordering: Remarkably, despite its mixed head-dependent ordering, Persian exhibits high stability in subject-object ordering. Futrell et al. (2015) and Levshina et al. (2023) demonstrate that Persian places subjects before objects in over 90% of cases, placing it among languages with rigid

subject-object order. This finding is particularly significant given Persian's differential object marking (DOM) system employing the marker *rā* (Rasekh-Mahand & Parizadeh, 2024). Contrary to expectations that case marking enables freer word order, Persian's DOM does not result in flexible subject-object ordering, aligning with the cross-linguistic pattern that languages with robust case systems may nonetheless maintain fixed word order.

Interaction between Case Marking and Word Order: Levshina (2019) and Levshina et al. (2023) establish a nuanced relationship between case marking and word order flexibility. Languages with comprehensive case systems (e.g., Lithuanian, Hungarian, Latin) tend to show greater word order flexibility, while languages with reduced or differential marking exhibit more rigid order. Persian's DOM system places it in an intermediate position: it has some case marking but maintains relatively fixed word order. This pattern confirms functional-adaptive explanations for word order properties: languages tend to maintain sufficient disambiguation strategies, either through case marking or through fixed word order, but these strategies are not necessarily mutually enabling (Levshina, 2019, p. 547).

Agreement and Word Order: Spirgath (2025) examines the interaction between verb agreement and word order entropy, showing a negative correlation: languages with stronger agreement tend to show more rigid word order. Persian, which has robust agreement marking, confirms this pattern, showing relatively low word order disorder (approximately 20% flexibility) despite its rich agreement system. This finding distinguishes agreement from case marking, as the two systems serve complementary functions in information management.

Multiple Word Order Parameters: Gerdes et al. (2021) provide detailed quantitative analysis of several word order parameters: Object-Verb Order: Persian places both nominal and pronominal objects before the verb in over 90% of cases, showing remarkable consistency comparable to Turkish and contrasting sharply with Arabic and English. Adjective-Noun Order: Persian overwhelmingly (approximately 95%) places adjectives after nouns, confirming Greenberg's Universal 19 and showing greater consistency than expected in a mixed typology language. Adverb Placement: Persian places lexical adverbs before the verb in over 96% of cases, but adverbial clauses appear postverbally in approximately 17% of cases, revealing variable behavior depending on adverbial type. Adposition Type: Persian is consistently prepositional, aligning with Arabic and contrasting with Turkish, demonstrating one of the most stable parameters in the language.

5 | Rasekh-Mahand

Post-Predicate Elements: Focusing on post-predicate constituents, Rasekh-Mahand et al. (2024) and Haig et al. (2024a) demonstrate that goal and caused goal arguments show the highest frequency of postverbal placement, with Persian placing 84% of goal arguments after the verb. This behavior is not random but systematically conditioned by semantic role, with animacy, style, and constituent weight playing secondary roles. Most importantly, this phenomenon cannot be attributed to information structure but reflects a structural property of colloquial Persian.

5. Discussion

The article advances areal-typological explanations for Persian's mixed word order properties, drawing on the concept of the Western Asia Transition Zone (WATZ) proposed by Haig (2017) and Haig and Khan (2018). This region, encompassing parts of Iran, eastern Turkey, and northern Iraq, has historically been a contact zone between Turkic (verb-final, postpositional), Semitic (verb-initial, prepositional), Indo-European (primarily Iranian), and Kartvelian language families.

Stilo (2005, 2009) pioneered the analysis of how Persian and other Iranian languages have been shaped by their position between these typologically distinct families. His research demonstrates that Iranian languages in the north and east (adjacent to Turkic and Caucasian languages) show more verb-final characteristics, while those in the west and south (adjacent to Semitic languages) show more verb-initial features. Haig et al. (2025) provide corpus-based confirmation of these patterns across 35 language varieties in the WATZ. Key findings include:

Object Position: While Iranian and Turkic languages historically maintain object-before-verb order in over 90% of instances, Semitic languages in the region show variable behavior. Some Neo-Aramaic varieties, surrounded by Iranian languages, have adopted object-before-verb order, demonstrating contact-induced change.

Postverbal Goal Arguments: The widespread tendency to place goal arguments after the verb extends across language families in the WATZ, with Iranian and Turkic languages showing rates of 60-100% postverbal goal placement. This pattern is strongest in western Iranian languages (closer to Semitic contact) and weaker in eastern varieties (further from Semitic contact).

Adposition Type: Iranian languages show a gradient from predominantly postpositional in the north and east (closer to Turkic) to predominantly

prepositional in the west and south (closer to Semitic), with many languages exhibiting both.

These areal patterns challenge purely historical explanations for Persian's mixed typology. Rather than representing an inherent property or a unidirectional diachronic shift, Persian's mixed characteristics reflect centuries of contact with typologically diverse language families. This perspective explains why Persian can be simultaneously verb-final (in object-verb order) and prepositional (patterned with verb-initial languages) without indicating transitional status or free word order.

6. Conclusion

This review of corpus-based typological research reveals several crucial insights about Persian word order. Persian's typological properties are not binary but exist on a continuum, showing high disorder in some domains (head-dependent ordering, postverbal argument placement) and high stability in others (subject-object ordering, adjective-noun order, adposition type). The interaction between case marking, agreement, and word order supports functional explanations: languages maintain sufficient disambiguation mechanisms without unnecessary redundancy. Persian's mixed properties cannot be attributed solely to internal historical development but reflect prolonged contact with neighboring language families in the Western Asia Transition Zone. Corpus-based approaches provide more accurate and nuanced descriptions than grammar-based typology, revealing systematic patterns in apparent disorder.

Bibliography:

- Csató, É. Á., Isaksson, B., & Jahani, C. (Eds.) (2005). *Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic and Turkic*. Routledge.
- Dabir Moghaddam, M. (2018). Typological approaches and dialects. In A. Sedighi & P. Shabani-Jadidi (Eds.), *The Oxford handbook of Persian linguistics* (pp. 52–88). Oxford University Press.
- Dryer, M. S. (1992). The Greenbergian word order correlations. *Language*, 68, 81–138.
- Dryer, M. S. (1997). On the six-way word order typology. *Studies in Language*, 21(2), pp. 69–103.
- Dryer, M. S. (2013). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *WALS Online* (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591>. Retrieved from <http://wals.info/chapter/81>.

7 | Rasekh-Mahand

- Dryer, M. S., & Haspelmath, M. (Eds.) (2013). *WALS Online* (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591>. Retrieved from <https://wals.info>.
- Frommer, P. (1981). *Post-verbal phenomena in colloquial Persian syntax* (Doctoral dissertation), University of Southern California.
- Futrell, R., Mahowald, K., Edward Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *PNAS*, 112(33), pp. 10336–10341.
- Gerdes, K., Kahane S., and Chen, X. (2021). Typometrics: From Implicational to Quantitative Universals in Word Order Typology. *Glossa: a journal of general linguistics* 6(1): 17. 1–31. DOI: <https://doi.org/10.5334/gjgl.764>
- Goldhahn, D., Eckart, Th., Quasthoff, U. (2012). Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In: Calzolari, N., Choukri, Kh., Declerck, Th. et al. (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pp. 759–765. Istanbul: ELRA.
- Greenberg, J. (1963/1966). Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. *Universals of Language*, London: MIT Press, pp. 110-113.
- Haig, G. (2017). Western Asia: East Anatolia as a transition zone. In Hickey, Raymond (ed.), *The Cambridge Handbook of Areal Linguistics*. Cambridge: CUP. pp. 396-423.
- Haig, G. & Khan, G. (2018). Introduction. In: Haig, Geoffrey & Geoffrey Khan (eds.) *The languages and linguistics of Western Asia. An areal perspective*. Berlin: Mouton de Gruyter. Pp. 1-29.
- Haig, G. & Rasekh-Mahand, M. (2019). Post-predicate constituents in Iranian and neighboring languages. Position paper for the project Post-predicate constituents in Iranian and neighboring languages. https://www.uni-bamberg.de/fileadmin/aspra/05_Events/2019_Post
- Haig, G., & Rasekh-Mahand, M. (Eds.). (2022). *HamBam: The Hamedan-Bamberg corpus of contemporary spoken Persian*. University of Bamberg. <http://multicast.aspra.uni-bamberg.de/resources/hambam/>
- Haig, G., & Schnell, S. (2014). *Annotations using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and guidelines for annotators* (Version 7.0). <https://multicast.aspra.uni-bamberg.de/>
- Haig, G., & Schnell, S. (2015). *Multi-CAST: Multilingual corpus of annotated spoken texts*. University of Bamberg. <https://multicast.aspra.uni-bamberg.de/>
- Haig, G., & Schnell, S. (Eds.). (2023). *Multi-CAST: Multilingual corpus of annotated spoken texts* (Version 2311). University of Bamberg. <http://multicast.aspra.uni-bamberg.de/>
- Haig, G., Schnell, S., & Seifart, F. (Eds.). (2022). *Doing corpus-based typology with spoken language corpora: State of the art* (Language Documentation &

- Conservation Special Publication No. 25). University of Hawai'i Press. <https://doi.org/10.5167/uzh-217096>
- Haig, G., & Rasekh-Mahand, M. (Eds.). (2022). *HamBam: The Hamedan-Bamberg corpus of contemporary spoken Persian*. University of Bamberg. <http://multicast.aspra.uni-bamberg.de/resources/hambam/>
- Haig, G., & Schnell, S. (2014). *Annotations using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and guidelines for annotators* (Version 7.0). <https://multicast.aspra.uni-bamberg.de/>
- Haig, G., & Schnell, S. (2015). *Multi-CAST: Multilingual corpus of annotated spoken texts*. University of Bamberg. <https://multicast.aspra.uni-bamberg.de/>
- Haig, G., & Schnell, S. (Eds.). (2023). *Multi-CAST: Multilingual corpus of annotated spoken texts* (Version 2311). University of Bamberg. <http://multicast.aspra.uni-bamberg.de/>
- Haig, G., Schnell, S., & Schiborr, N. N. (2021). Universals of reference in discourse and grammar: Evidence from the Multi-CAST collection of spoken corpora. In G. Haig, S. Schnell, & F. Seifart (Eds.), *Doing corpus-based typology with spoken language corpora* (Language Documentation & Conservation Special Publication No. 25). University of Hawai'i Press.
- Haig, G., Vollmer, M., & Thiele, H. (2019). Multi-CAST Northern Kurdish. In G. Haig & S. Schnell (Eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts* (Version 2311). University of Bamberg. <http://multicast.aspra.uni-bamberg.de/#nkurd>
- Haig, G., Stilo, D., Schiborr, N. N., & Dogan, M. (Eds.). (2022). *WOWA: Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation*. University of Bamberg. <http://multicast.aspra.uni-bamberg.de/resources/wowa/>
- Haig, G., Rasekh-Mahand, M., Stilo, D., Schreiber, L., & Schiborr, N. N. (Eds.). (2024a). *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology* (Contact and Multilingualism 8). Language Science Press. <https://langsci-press.org/catalog/book/395>
- Haig, G., Rasekh-Mahand, M., Stilo, D., Schreiber, L., & Schiborr, N. (2024b). Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In G. Haig, M. Rasekh-Mahand, D. Stilo, L. Schreiber, & N. Schiborr (Eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology* (pp. 1–48). Language Science Press. <https://langsci-press.org/catalog/book/395>
- Haig, G., Noorlander, P., & Schiborr, N. (2025). Which word order features are stable in a contact setting? Corpus-based evidence from the Western Asian Transition Zone. In J. Darquennes, J. C. Salmons, & W. Vandebussche (Eds.), *Language contact, an international handbook* (Vol. 2, pp. 159–184). De Gruyter Mouton. <https://doi.org/10.1515/9783110443011-010>

9 | Rasekh-Mahand

- Levshina, N. (2015). European analytic causatives as a comparative concept: Evidence from a parallel corpus of film subtitles. *Folia Linguistica*, 49(2), 487–520.
- Levshina, N. (2019). Token-based typology and word order entropy. *Linguistic Typology*, 23(3), 533–572. <https://doi.org/10.1515/lingty-2019-0025>
- Levshina, N. (2022a). Corpus-based typology: Applications, challenges and some solutions. *Linguistic Typology*, 26(1), 129–160.
- Levshina, N. (2022b). *Communicative efficiency: Language structure and use*. Cambridge University Press.
- Levshina, N., Nambodiripad, S., Allasonnière-Tang, M., Kramer, M., Talamo, L., Verkerk, A., Wilmoth, S., Garrido Rodriguez, G., Gupton, T. M., Kidd, E., Liu, Z., Naccarato, C., Nordlinger, R., Panova, A., & Stoyanova, N. (2023). Why we need a gradient approach to word order. *Linguistics*, 61(4), 825–883. <https://doi.org/10.1515/ling-2021-0098>
- Parizadeh, M., & Rasekh-Mahand, M. (2024). Post-predicate elements in Early New Persian. In G. Haig, M. Rasekh-Mahand, D. Stilo, L. Schreiber, & N. Schiborr (Eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology* (pp. 184–196). Language Science Press. <https://langsci-press.org/catalog/book/395>
- Rasekh-Mahand, M., & Parizadeh, M. (2024). Different functions of ‘rā’ in New Persian: A semantic map analysis. *Journal of Historical Linguistics*, 14(1), 31–57. <https://doi.org/10.1075/jhl.21056.ras>
- Rasekh-Mahand, M., Izadi, E., Parizadeh, M., Haig, G., & Schiborr, N. (2024). Post-predicate elements in colloquial Persian: A multi-variate analysis. In G. Haig, M. Rasekh-Mahand, D. Stilo, L. Schreiber, & N. Schiborr (Eds.), *Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology* (pp. 151–182). Language Science Press. <https://langsci-press.org/catalog/book/395>
- Schnell, S., & Schiborr, N. N. (2022). Crosslinguistic corpus studies in linguistic typology. *Annual Review of Linguistics*, 8, 171–191. <https://doi.org/10.1146/annurev-linguistics-031120-104629>
- Spirgath, A. (2025). The interaction of word order entropy and verb agreement: A token-based approach. *Linguistics in Amsterdam*, 16(1), 30–50.
- Stilo, D. (1994). Phonological systems in contact in Iran and Transcaucasia. In M. Marashi (Ed.), *Persian studies in North America: Studies in honor of Mohammad Ali Jazayery*. Iranbooks.
- Stilo, D. (2005). Iranian as buffer zone between the universal typologies of Turkic and Semitic. In É. Á. Csató, B. Isaksson, & C. Jahani (Eds.), *Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic and Turkic* (pp. 35–63). Routledge.
- Stilo, D. (2009a). Circumpositions as an areal response: The case study of the Iranian zone. *Turkic Languages*, 13, 3–33.

- Stilo, D. (2009b). Case in Iranian: From reduction and loss to innovation and renewal. In A. L. Malchukov & A. Spencer (Eds.), *The Oxford handbook of case*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199206476.013.0049>
- Stilo, D. (2012). Intersection zones, overlapping isoglosses, and ‘Fade-out/Fade-in’ phenomena in Central Iran. In B. Aghaei & M. R. Ghanoonparvar (Eds.), *Iranian languages and culture: Essays in honor of Gernot Ludwig Windfuhr* (pp. 3–33). Mazda Publishers.
- Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)* (pp. 2214–2218). European Language Resources Association (ELRA).
- Vatani, T., Väyrynen, J. J., & Virpioja, S. (2010). Language identification of short text segments with n-gram models. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)* (pp. 3423–3430). European Language Resources Association (ELRA).
- Verkerk, A. (2014). *The evolutionary dynamics of motion event encoding* [Doctoral dissertation, Radboud University Nijmegen].
- Zeman, D., et al. (2024). *Universal Dependencies 2.15* [Data set]. LINDAT/CLARIAH-CZ digital library, Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <http://hdl.handle.net/11234/1-5787>



رده‌شناسی پیکره‌بنیاد و توالی کلمات در زبان فارسی

استاد، گروه زبان‌شناسی، دانشگاه بوعلی سینا، همدان، ایران

محمد راسخ‌مهند 

چکیده

این مقاله به بررسی رده‌شناسی توالی کلمات در زبان فارسی با تمرکز بر رویکرد رده‌شناسی پیکره‌بنیاد اختصاص دارد. مطالعه رده‌شناختی دستوربنیاد توالی کلمات در زبان فارسی نشان داده است که این زبان، با این‌که گرایشی غالب ولی غیرقطعی به سمت الگوهای فعل‌آغازی دارد، در طبقه‌بندی ساده‌شده به عنوان زبان «فعل‌آغاز» یا «فعل‌پایان» قرار نمی‌گیرد، بلکه زبانی آمیخته است. در مورد چرایی این رده آمیخته نیز فرضیه‌هایی مطرح شده است، از جمله این‌که زبان فارسی در حال گذار به رده فعل‌آغاز است. در این مقاله نخست با مرور مطالعات رده‌شناسی پیکره‌بنیاد با استفاده از داده‌های زبانی واقعی و کمی، تصویر کامل‌تر و مدرجی از بی‌نظمی توالی کلمات در زبان‌های جهان و به تبع آن وضعیت فارسی ارائه شده است. مطالعات پیکره‌ای منتخب نشان می‌دهند که بی‌نظمی توالی کلمات در فارسی یکسان نیست. برای مثال، در توالی هسته و وابسته، فارسی بی‌نظمی نسبتاً بالایی دارد، اما در توالی فاعل و مفعول، توالی بسیار ثابتی (عمدتاً فاعل قبل از مفعول) دارد. همچنین، اگرچه فارسی مفعول‌نمایی افتراقی با استفاده از نشانه «را» دارد، این امر منجر به آزادی زیاد در ترتیب فاعل و مفعول نشده است. در نهایت، سعی در تبیین ویژگی‌های توالی کلمات زبان فارسی بر اساس رده‌شناسی منطقه‌ای شده است. به عبارتی سعی شده نشان داده شود بی‌نظمی‌های موجود در توالی کلمات این زبان و تفاوت آن با زبان‌های فعل‌آغاز و فعل‌پایان، می‌تواند حاصل تماس زبانی بلندمدت و تاثیرپذیری از زبان‌های همسایه، به‌ویژه زبان‌های سامی و ترکی، در «منطقه انتقالی غرب آسیا» باشد. در نهایت بر اهمیت پژوهش‌های پیکره‌بنیاد و ساخت پیکره‌های بیشتر برای زبان فارسی و سایر زبان‌های ایرانی تأکید شده است.

کلیدواژه: توالی کلمات، رده‌شناسی، پیکره، تماس زبانی، فارسی.

۱- رده‌شناسی و توالی کلمات

رده‌شناسی نوین (کرافت^۱، ۲۰۰۳) با بررسی توالی کلمات^۲ یا ترتیب واژه شروع شد (گرینبرگ^۳، ۱۹۶۳) و از آن زمان نیز همیشه توالی کلمات یکی از مباحث اساسی در این حوزه بوده است. از زمان گرینبرگ تا حدود دو دهه گذشته، داده‌های مورد بررسی رده‌شناسان عمدتاً از منابع نوشته شده، مانند دستورهای توصیفی یا مقالات مختلف، و در برخی موارد نیز یافته‌های میدانی بوده است. مثال‌های زبانی در این منابع، عمدتاً نمونه‌های دست دوم تلقی می‌شوند، چون در اکثر مواقع رده‌شناسان از داده‌ها و مثال‌های افراد دیگر استفاده کرده و قضاوت و یافته خود را بر آن بنا نهاده‌اند. در حدود دو دهه اخیر، با گسترش امکان ساختن پیکره‌های زبانی مختلف از زبان‌های گوناگون، مطالعات رده‌شناسی زبان به سمت استفاده از پیکره‌ها کشیده شده است. این گرایش به حدی است که اکنون از نوعی رده‌شناسی متفاوت با نام رده‌شناسی پیکره‌بنیاد^۴، رده‌شناسی موردبنیاد^۵ یا رده‌شناسی مدرج^۶ سخن به میان می‌آید. این دیدگاه در مقابل نگاه سنتی به رده‌شناسی مطرح می‌شود که از آن با عنوان رده‌شناسی مشخصه‌بنیاد^۷ یا رده‌شناسی دستوربنیاد^۸ یاد می‌شود (هیگ^۹ و همکاران، ۲۰۲۲؛ اشل و شیبور^{۱۰}، ۲۰۲۲؛ لوشینا^{۱۱}، ۲۰۱۹ و ۲۰۲۱؛ لوشینا و همکاران، ۲۰۲۳).

هدف این مقاله مرور مطالعات رده‌شناختی توالی کلمات با تکیه بر یافته‌های جدید در رده‌شناسی پیکره‌بنیاد است. نویسنده در نظر دارد نشان دهد با توجه به مطالعات پیکره‌ای انجام شده، وضعیت توصیف و تبیین رده‌شناسی توالی کلمات فارسی چگونه است، چه یافته‌هایی به دست آمده است و با یافته‌های رده‌شناسی دستوربنیاد چه تفاوت‌هایی دارد، و چه فرصت‌هایی برای پژوهش‌های بیشتر وجود دارد؟ در اصل این مقاله مرور یافته‌هایی است که در رده‌شناسی پیکره‌بنیاد به دست آمده و ارتباط آن با زبان فارسی. این نگاه باعث می‌شود تصویری درست و گاهی متفاوت از یافته‌های مطالعات رده‌شناسی دستوربنیاد به دست آید. ساختار مقاله به این شکل است: پس از مقدمه حاضر، در بخش (۲) به مرور مطالعات رده‌شناسی دستوربنیاد توالی کلمات می‌پردازیم و وضعیت زبان فارسی را از دید این رویکرد بررسی

-
1. Croft, W.
 2. word order
 3. Greenberg, J.
 4. corpus-based typology
 5. token-based typology
 6. gradient typology
 7. feature-based typology
 8. grammar-based typology
 9. Haig, G.
 10. Schnell, S. & Schiborr, N. N.
 11. Leshina, N.

می‌کنیم. در بخش (۳) رده‌شناسی پیکره‌بنیاد، مفاهیم اصلی آن، تفاوت آن با رده‌شناسی دستوربنیاد و وضعیت کلی پیکره‌ها در این نگرش معرفی شده است. بخش (۴) به مهمترین مطالعات پیکره‌بنیادی اشاره می‌کند که در آنها به برخی ویژگی‌های توالی کلمات در زبان فارسی نیز پرداخته شده است و از این حیث نشان داده می‌شود این نگرش با نگرش دستوربنیاد در رده‌شناسی چه تفاوت‌هایی دارد. در این بخش توالی‌های مختلف و بی‌نظمی‌های موجود در آنها بررسی شده است. در بخش (۶) سعی شده به این پرسش پاسخ داده شود که چرا زبان فارسی و بسیاری از زبان‌های ایرانی در قالب دو رده فعل‌آغاز^۱ یا فعل‌پایان^۲ نمی‌گنجند و به نوعی زبان آمیخته^۳ هستند؟ تبیین این بخش بر اساس مطالعات پیکره‌ای است که در منطقه غرب آسیا صورت گرفته است و بر تماس زبانی^۴ یا رده‌شناسی منطقه‌ای^۵ به عنوان عاملی مهم و مغفول در مطالعات رده‌شناسی توالی کلمات در فارسی تاکید می‌کند. بخش (۵) نتیجه‌گیری مقاله است.

۲. رده‌شناسی دستوربنیاد

رده‌شناسان به دنبال یافتن شباهت‌های بین زبان‌ها، یا همان همگانی‌ها^۶ هستند که در اکثر زبان‌ها مشترک‌اند. همچنین تعیین گوناگونی‌های بین زبان‌ها و محدودیت‌های حاکم بر آنها نیز از اهداف آنهاست. بحث توالی کلمات و اصولاً رده‌شناسی نوین با مقاله گرینبرگ (۱۹۶۳) با نام «برخی همگانی‌های دستور با اشاره مشخص به توالی عناصر معنی‌دار» شروع شد. وی بر اساس مطالعه ۳۰ زبان، ۴۵ همگانی پیشنهاد کرد که ۲۸ مورد از آنها به توالی عناصر جمله اشاره داشت و بقیه به ویژگی‌های صرفی مربوط می‌شد. اکثر این همگانی‌ها، تلویحی^۷ بودند یعنی به صورت «اگر الف، آن‌گاه ب» گفته شده‌اند. همگانی‌های تلویحی یکی از مهم‌ترین ویژگی‌های پژوهش‌های رده‌شناختی است زیرا پیش‌بینی می‌کند که حضور یک ویژگی در زبان‌ها احتمال حضور یک یا چند ویژگی دیگر را هم مطرح می‌کند. به عبارتی زبان‌ها را در رده‌های مختلف قرار می‌دهد. گرینبرگ سه معیار اساسی را در رده‌شناسی زبان‌ها به کار گرفت؛ یکی داشتن پس‌اضافه^۸ یا پیش‌اضافه، دوم توالی‌های محتمل فاعل و مفعول و فعل، و سوم توالی اسم و صفت. اکثر همگانی‌های تلویحی که در مقاله آمده‌اند، بر اساس سه معیار فوق

-
1. verb initial
 2. verb final
 3. mixed
 4. language contact
 5. areal typology
 6. universal
 7. implicational
 8. preposition

بیان شده‌اند. چند مورد از همگانی‌های گرینبرگ (۱۹۶۳) که به توالی کلمات می‌پردازد عبارت‌اند از:

همگانی اول:

توالی غالب در جمله‌های خبری که فاعل و مفعول اسمی دارند تقریباً همیشه به صورتی است که فاعل قبل از مفعول قرار می‌گیرد.

همگانی سوم:

زبان‌هایی که توالی غالب آن‌ها «فاعل+مفعول»^۱ است، همیشه پیش‌اضافه^۲ دارند.

همگانی چهارم:

به احتمال زیاد و فراتر از شانس، زبان‌هایی که دارای توالی «فاعل+مفعول+فاعل»^۳ هستند، پس‌اضافه دارند.

همگانی‌های تلویحی گرینبرگ چهار امکان به دست می‌داد، اما محدودیت امکان آنها را نیز مشخص می‌کرد. به عنوان نمونه، طبق همگانی چهارم، زبان‌ها در چند رده قرار می‌گیرند. آنهایی که توالی فاعل+مفعول+فاعل و پس‌اضافه دارند (طبق پیش‌بینی همگانی). موارد نادری که فاعل+مفعول+فاعل دارند ولی پیش‌اضافه^۴ دارند. زبان‌هایی که توالی فاعل+فاعل+مفعول^۵ و پیش‌اضافه دارند و زبان‌هایی که فاعل+فاعل+مفعول و پس‌اضافه دارند. رابطه بین توالی سه سازه اصلی (فاعل، فاعل و مفعول) با توالی‌های دیگر (مانند توالی اسم و صفت، اسم و بند موصولی یا اسم و حرف اضافه و نظایر آن)، رابطه‌ای منطقی نیست اما مطالعات رده‌شناختی نشان داد که بین این توالی‌ها، رابطه‌ها یا هم‌بستگی‌هایی^۶ وجود دارد که نمی‌تواند اتفاقی باشد. به این ترتیب که اگر یکی از ویژگی‌ها، مثلاً توالی فاعل و مفعول را در نظر بگیریم، می‌توانیم برخی توالی‌های دیگر را با احتمال بالا پیش‌بینی کنیم (درایر^۷، ۱۹۹۲).

درایر (۱۹۹۲) در مقاله مهمی با عنوان «هم‌بستگی‌های توالی کلمات گرینبرگی» به بررسی این هم‌بستگی‌ها در ۶۲۵ زبان پرداخت. وی با تقسیم‌بندی زبان‌ها به فعل‌آغاز و فعل‌پایان نشان می‌دهد اکثر زبان‌های دنیا در معیارهای معرفی شده به یکی از این دو رده تعلق دارند. او بحث هم‌بستگی‌ها را با هم‌بستگی حرف اضافه^۸ و گروه اسمی وابسته آن آغاز می‌کند. درایر

-
1. OSV
 2. postposition
 3. SOV
 4. preposition
 5. SVO
 6. correlation
 7. Dryer, M.
 8. adposition

می‌گویند که شواهد کافی وجود دارد که نشان می‌دهد توالی حرف اضافه و گروه اسمی متمم آن با توالی فعل و مفعول هم‌بستگی بالایی دارد (درایر، ۱۹۹۲: ۸۵). به این ترتیب، او عناصر را به دو طبقه‌ی فعل‌الگو^۱ و مفعول‌الگو^۲ تقسیم می‌کند؛ که به ترتیب به مواردی اشاره دارد که عنصر مورد نظر در جایگاه هم‌راستا با فعل یا هم‌راستا با مفعول در توالی مورد بحث قرار می‌گیرد. جدول (۱) از درایر (۱۹۹۲) نشان می‌دهد که حرف اضافه فعل‌الگو و گروه اسمی متمم آن مفعول‌الگو است:

جدول ۱. هم‌بستگی حرف اضافه با توالی فعل و مفعول (درایر، ۱۹۹۲: ۸۳)

مجموع	آمریکای جنوبی	آمریکای شمالی	استرالیا و گینه نو	جنوب شرقی آسیا و اقیانوسیه	اروپا-آسیا	آفریقا	
۱۰۷	۱۹	۲۵	۱۷	۵	۲۶	۱۵	مفعول فعل و پس‌اضافه
۷	۰	۰	۱	۰	۳	۳	مفعول فعل و پیش‌اضافه
۱۲	۴	۳	۰	۰	۱	۴	فعل مفعول و پس‌اضافه
۷۰	۵	۲۰	۶	۱۵	۸	۱۶	فعل مفعول و پیش‌اضافه

داده‌های درایر به‌وضوح نشان می‌دهد که در ردیف‌های اول و آخر هم‌بستگی مشخصی دیده می‌شود. یعنی در ۱۰۷ گروه زبانی (نه زبان) که توالی «مفعول+فعل^۳» داشته‌اند، پس‌اضافه وجود دارد و در ۷۰ گروه زبانی که توالی «فعل+مفعول^۴» داشته‌اند، پیش‌اضافه وجود دارد. هرگونه مورد خلاف آن‌ها، باعث به‌وجود آمدن الگوی کم‌بسامد یا نادر در زبان‌ها می‌شود، که در ردیف‌های دوم و سوم دیده می‌شود. ادامه مقاله درایر معرفی سایر هم‌بستگی‌ها به سیاق فوق است.

خلاصه حرف درایر (۱۹۹۲) این است که زبان‌های فعل‌آغاز (یا با اندکی مسامحه هسته‌آغاز^۵)، در اکثر توالی‌های دیگر، مانند حرف اضافه و اسم، اسم و صفت، اسم و وابسته اضافی، اسم و بند موصولی، هسته را قبل از وابسته قرار می‌دهند و زبان‌های فعل‌پایان، یا همان

1. verb patterner
2. object patterner
3. OV
4. VO
5. head-initial

هسته‌پایان^۱، هسته را پس از وابسته قرار می‌دهند. برخی از این هم‌بستگی‌ها قوی هستند، یعنی در اکثریت بالایی از زبان‌ها رعایت می‌شوند، و برخی ضعیف هستند یعنی در تعداد زبان‌هایی که آن را رعایت نمی‌کنند هم قابل توجه است. اما در کل، هم‌بستگی‌ها نشان دادند که زبان‌های جهان در یکی از رده‌های فعل‌آغاز یا فعل‌پایان قرار می‌گیرند. مطالعه رده‌شناختی توالی کلمات در زبان فارسی نیز بر اساس معیارهای درایر انجام شده است (دبیرمقدم، ۱۳۹۲، ۲۰۱۸؛ راسخ‌مهند، ۱۳۹۶).

پرسشی که در بحث توالی کلمات در زبان فارسی مطرح است این است که آیا زبان فارسی در رده زبان‌های فعل‌آغاز قرار می‌گیرد یا رده زبان‌های فعل‌پایان؟ زبان فارسی در توالی مفعول و فعل، فعل‌پایان است و فعل را بعد از مفعول قرار می‌دهد و توالی «فاعل+مفعول+فعل» دارد، اما مطالعه سایر توالی‌ها نشان می‌دهد در تعدادی از این توالی‌ها، فارسی الگوی متفاوتی دارد و هسته را پس از فعل قرار می‌دهد. برای نمونه، زبان فارسی بیش‌تر دارای پیش‌اضافه است. حال اگر این ویژگی را با توالی فعل و مفعول مقایسه کنیم (جدول ۱) به این نتیجه می‌رسیم که زبان فارسی مفعول+فعل و پیش‌اضافه‌ای است (ردیف دوم)، و از این حیث این زبان در گروه آسیا-اروپا (منطقه‌ی جغرافیایی خود)، در این ویژگی زبانی، فعل‌پایانی ضعیف است، زیرا فقط سه گروه زبانی این‌گونه هستند، و در مجموع نیز این هم‌بستگی را ندارد. دبیرمقدم با بررسی معیارهای درایر (۱۹۹۲) در جفت‌های هم‌بسته توالی کلمات به این نتیجه رسیده است که: «زبان فارسی به لحاظ مؤلفه‌های رده‌شناختی، زبانی آمیخته و غیریکدست است. این زبان از برخی از مؤلفه‌های زبان‌های فعل‌پایانی برخوردار است و از برخی (بیش‌تر) از مؤلفه‌های فعل-میانی^۲» (دبیرمقدم ۱۳۹۲: ۱۲۸). به عبارتی فارسی مطابق انتظار عمل نمی‌کند، مفعول را قبل از فعل قرار می‌دهد، اما همه وابسته‌ها را قبل از هسته قرار نمی‌دهد. مطالعه رده‌شناختی توالی کلمات در زبان فارسی نشان می‌دهد این زبان در چهار پارامتر مانند زبان‌های فعل‌پایان عمل می‌کند:

-توالی گروه حرف اضافه‌ای و فعل:

(۱) علی از تهران رفت.

- توالی فعل و قید حالت^۳:

(۲) دوستم تند غذا می‌خورد.

1. head-final

۲- او به جای «فعل‌آغازی» از «فعل‌میانی» استفاده کرده است.

3. manner adverb

- توالی فعل ربطی^۱ و مسند:
(۳) دخترم دانشجو است.

- جایگاه کلمهٔ پرسشی^۲:
(۴) کتاب را به کی دادی؟

اما این زبان در نه هم‌بستگی زیر مانند زبان‌های فعل‌آغاز رفتار می‌کند:

- توالی هستهٔ اسمی و بند موصولی:
(۵) مردی که دیروز دیدی همسایه‌مان بود.

- توالی مضاف و مضاف‌الیه:
(۶) کتاب برادرم

- توالی فعل «خواستن» و فعل بند پیرو^۳:
(۷) می‌خواهم که داستانی بنویسم.

- توالی ادات پرسشی^۴ و جمله:
(۸) آیا از این بیمارستان خوش می‌آید؟

- توالی حرف ربط قیدی^۵ و بند قیدی^۶:
(۹) وقتی علی را در بازار دیدم، شناختمش.

- توالی فعل اصلی و فعل کمکی^۷ «توانستن»:
(۱۰) می‌تونم به فقرا کمک کنم.

-
1. copula
 2. wh-word
 3. dependent clause
 4. question particle
 5. adverbial connector
 6. adverbial clause
 7. auxiliary verb

- توالی وند منفی و ستاک فعل:

(۱۱) علی نمی‌رود.

- توالی متمم‌نما^۱ و بند متممی^۲:

(۱۲) می‌دانم که دوستم مغازه‌اش را بسته است.

- توالی حرف اضافه و گروه اسمی:

(۱۳) به احمد

علاوه بر این مواردی نیز هست که رفتاری متفاوت از این دو الگو دارد. مثلاً در چهار توالی زیر، زبان فارسی هر دو هم‌بستگی متفاوت را دارد، یعنی زبان فارسی بیش از یک توالی دارد و از این رو با هر دو رده هم‌بستگی دارد:

- توالی صفت و مبنای مقایسه^۳. زبان فارسی در برخی ساخت‌ها صفت را قبل از مبنای

مقایسه قرار دهد (۱۴) و در برخی دیگر بعد از آن (۱۵). از این رو، هم شبیه زبان‌های فعل‌آغاز است و هم فعل‌پایان:

(۱۴) جوان‌تر از برادرم

(۱۵) از برادرم جوان‌تر

- توالی فعل اصلی و فعل کمکی نشان‌دهنده زمان^۴ و نمود^۵. برخی افعال کمکی قبل از

فعل اصلی و برخی بعد از آن قرار می‌گیرند:

(۱۶) فردا به شمال خواهند رفت.

(۱۷) پارسال به مسافرت رفته بودند.

- توالی وند نشان‌دهنده زمان و نمود و فعل. در این معیار هم زبان فارسی هر دو امکان را

دارد:

1. complementizer
2. complement clause
3. standard of comparison
4. tense
5. mood

(۱۸) پریدم (-ید پسوند است)

(۱۹) می‌رفتم (می - پیشوند است)

- توالی مفعول و فعل. زبان فارسی مفعول اسمی را قبل از فعل و مفعول بندی را پس از فعل قرار می‌دهد:

(۲۰) احمد را دیدم.

(۲۱) می‌دانم که احمد در خانه نبود.

زبان فارسی در هفت توالی دیگر هم به صورت خاص به دو رده فوق تعلق ندارد. تمام این توالی‌ها به ساخت‌هایی اشاره می‌کنند که فاقد انشعاب^۱ هستند (درایر ۱۹۹۲). به عبارتی توالی آنها قبل یا بعد از هسته اهمیت چندانی ندارد زیرا فقط یک کلمه (عمدتاً نقشی) هستند:

- توالی صفت و موصوف.

(۲۲) داماد جوان

- توالی حرف تعریف و اسم.

(۲۳) پسری

(۲۴) پسرِ

- توالی اسم و تکواژ ملکی.

(۲۵) برادرم

(۲۶) برادر او

- توالی صفت اشاره و اسم.

(۲۷) آن کتاب

- توالی قید درجه و صفت.

(۲۸) خیلی زیبا

- توالی فعل و فاعل.

(۲۹) مریم رفت.

- توالی عدد و اسم.

(۳۰) چهار مرد

با توجه به بررسی فوق، دبیرمقدم (۲۰۱۸) پارامترهای درایر در زبان فارسی را در جدول زیر نشان می‌دهد:

جدول ۲. پارامترهای درایر در زبان فارسی (دبیرمقدم، ۲۰۱۸: ۶۴)

توالی کلمات در زبان فارسی	
۴	پارامترهای فعل‌پایان
۹	پارامترهای فعل‌آغاز
۷	پارامترهای مشترک فعل‌پایان و فعل‌آغاز
۴	پارامترهای بینابینی

زبان فارسی فقط در چهار توالی منحصرأ با زبان‌های فعل‌پایان هم‌بستگی دارد. در نه مورد هم با زبان‌های فعل‌آغاز هم‌بستگی دارد و در یازده مورد دیگر رفتاری بینابینی دارد. به این ترتیب این زبان بیشتر گرایش به زبان‌های فعل‌آغاز دارد، هرچند نمی‌توان رده توالی کلمات این زبان را به قطعیت تعیین کرد. در واقع فارسی را نمی‌توان در رده فعل‌آغاز یا فعل‌پایان قرار داد. حال پرسش این است که چرا زبان فارسی رفتار یکدستی از خود نشان نمی‌دهد؟ یعنی چرا با قاطعیت نمی‌توان گفت این زبان متعلق به رده فعل‌آغاز یا فعل‌پایان است؟ دبیرمقدم برای پاسخ به این پرسش دو فرضیه را مطرح کرده است: نخست این که زبان فارسی در مرحله تغییر رده است، و دوم این که زبان فارسی اساساً زبانی با توالی کلمات آزاد^۱ است. دبیرمقدم (۱۳۹۲) برای تعیین صحت و سقم این فرضیه‌ها به بررسی توالی کلمات در فارسی باستان و فارسی میانه پرداخته و به این نتیجه می‌رسد که «فارسی باستان زبانی یکدست و یکپارچه به لحاظ مشخصه‌های رده‌شناختی نیست، اما به فعل‌میانی (فعل‌آغازی) بودن گرایش دارد» (دبیرمقدم ۱۳۹۲: ۱۳۴). در مورد زبان فارسی میانه نیز نتیجه مشابهی گرفته است (دبیرمقدم ۱۳۹۲: ۱۳۸). وی به عنوان نتیجه‌گیری می‌گوید که «زبان فارسی به لحاظ مؤلفه‌های رده‌شناختی نحوی زبانی است با گرایش غالب به سوی زبان‌های دارای ترتیب فاعل-فعل-مفعول (یعنی فعل‌آغازی). این تمایل در فارسی میانه و باستان نیز مشهود است لیکن در دوره نو تقویت شده است. در این دوره، هم‌چنین شاهد تثبیت و پایداری بیشتر توالی ارکان اصلی جمله (یا همان ترتیب واژه‌ها) هستیم» (دبیرمقدم ۱۳۹۲: ۱۴۲). بر این اساس، وی

1. free word order

فرضیه نخست، یعنی این که زبان فارسی در حال تغییر رده (از فعل‌پایانی به فعل‌آغازی) است، را تأیید می‌کند، و فرضیه دوم را رد می‌کند. البته تأکید می‌کند که سرعت این گرانش و رانش بسیار آهسته و کند بوده است.

به طور خلاصه می‌توان گفت مطالعه رده‌شناختی دستوربنیاد توالی کلمات نشان داده است که زبان‌ها دارای هم‌بستگی‌هایی هستند که بر اساس آن آنها را در دو رده بزرگ فعل‌آغاز و فعل‌پایان قرار می‌دهند. اعمال این معیارها بر زبان فارسی نشان داده است فارسی در هیچ یک از این دو رده قرار نمی‌گیرد و با این که از حیث تاریخی زبان‌های ایرانی باستان فعل‌پایان هستند، اما اکنون توالی‌هایی که عکس آن عمل می‌کنند در زبان فارسی (و بسیاری از زبان‌های ایرانی) غالب هستند. به همین دلیل، محققان آنها را در رده زبان‌های آمیخته طبقه‌بندی کرده‌اند. با این که سعی شده گاه تبیین‌هایی برای این عدم یکدستی عنوان شود، اما نمی‌توان با قطعیت گفت چرا زبان‌های ایرانی، از جمله فارسی، رفتاری خاص خود دارد. نکته مهم قابل ذکر دیگر این است که، با این که همه محققان می‌دانند که یک زبان ممکن است همزمان بیش از یک الگو داشته باشد، مثلاً مفعول را قبل از فعل (احمد رو دیدم) و گاهی پس از آن قرار دهد (دیدم احمد رو)، اما در مطالعات خود، یک الگو را غالب دانسته و الگوی دیگر را معمولاً نشاندار، خاص و وابسته به بافتی مشخص دانسته‌اند. به عبارتی، در بحث توالی کلمات، گوناگونی در توالی‌ها نادیده گرفته شده است. این ویژگی مطالعات رده‌شناختی دستوربنیاد است (درایر، ۱۹۹۷ و ۲۰۱۳؛ درایر و هسپلمث، ۲۰۱۳).

۳. رده‌شناسی پیکره‌بنیاد

در بخش قبل به رده‌شناسی دستوربنیاد و مطالعه توالی کلمات در زبان فارسی بر اساس این دیدگاه پرداختیم. از حدود دو دهه اخیر، با گسترش امکان ساختن پیکره‌های زبانی مختلف از زبان‌های گوناگون، مطالعات رده‌شناسی زبان به سمت استفاده از پیکره‌ها کشیده شده است. همان گونه که عنوان شد این گرایش به حدی است که اکنون از نوعی رده‌شناسی متفاوت با نام رده‌شناسی پیکره‌بنیاد، رده‌شناسی موردبنیاد، رده‌شناسی مدرّج، یا مطالعات تطبیقی در مقابل نگاه سنتی به رده‌شناسی با عنوان رده‌شناسی مشخصه‌بنیاد یا رده‌شناسی دستوربنیاد یاد می‌شود (هیگ و همکاران، ۲۰۲۲؛ اشنل و شیپور، ۲۰۲۲؛ لوشینا، ۲۰۱۹؛ لوشینا و همکاران، ۲۰۲۳؛ ابرت^۱ و همکاران، ۲۰۲۴؛ جینگ^۲ و همکاران، ۲۰۲۲؛ ایوانی و بیکل^۳، زیر چاپ). البته همچنان محدودیت پیکره‌ها، به‌ویژه در مورد زبان‌هایی که پیکره‌های بزرگ از آنها در دست

1. Ebert, C.

2. Jing, Y.

3. Ivani, J. & Bickel, B.

نیست، باعث شده است نتوان این حوزه را به طور کامل جایگزین حوزه رده‌شناسی دستوربنیاد کرد (لوشینا، ۲۰۲۱)، اما رده‌شناسی پیکره‌بنیاد با سرعت در حال رشد است و مطالعات فراوانی در این حوزه صورت گرفته است. در ادامه برخی ویژگی‌های رده‌شناسی پیکره‌بنیاد ذکر می‌شود و از این رهگذر می‌توان به تفاوت‌های آن با رده‌شناسی دستوربنیاد پی برد:

- در رده‌شناسی دستوربنیاد، محقق کم و بیش بر قضاوت‌ها و نتیجه‌گیری‌های شخص یا اشخاص دیگر از داده‌های زبانی تکیه دارد و به عبارتی داده‌های دست دوم در اختیار دارد. در مطالعات پیکره‌ای، داده‌ها دست اول است و شخص می‌تواند به طور مستقیم و با جزییات بیشتر به واکاوی داده‌های زبانی بپردازد.

- در برخی موارد، می‌توان از پیکره‌ها برای تقویت و تکمیل مطالعات دستوربنیاد استفاده کرد. در واقع در این موارد پیکره همان کاری را می‌کند که مطالعه دستوربنیاد هم انجام می‌دهد. -پیکره‌ها اطلاعات موجز درباره بافتی که یک ساخت مشخص به کار می‌رود در اختیار ما قرار می‌دهند. به عبارتی آنها بافت‌بنیاد هستند.

- نتایج حاصل از مطالعات پیکره‌ای می‌تواند کمی^۱ باشد، و به این ترتیب ماهیت مدرج^۲ و پیوستاری ویژگی‌های زبانی را نشان دهد. در واقع رده‌شناسی دستوربنیاد اغلب مقوله‌ای است. مثلاً تعیین می‌کند یک زبان در ساخت (الف) در کدام رده زبانی قرار می‌گیرد. به فرض آیا فعل آغاز است یا فعل پایان؟ آیا پیش‌اضافه‌ای است یا پس‌اضافه‌ای؟ کمی بودن کمک می‌کند درک درست‌تری از ماهیت ویژگی‌های زبانی به دست آید (لوشینا و همکاران، ۲۰۲۳).

- رده‌شناسی پیکره‌بنیاد به تعیین تعمیم‌های بین‌زبانی جدید منجر می‌شود و حتی هم‌بستگی‌های جدید و گرایش‌های جدید را در یک زبان و بین چند زبان مشخص می‌کند. بر این اساس حتی می‌توان همگانی‌های مشخص‌شده قبلی را نیز دوباره بررسی کرد.

- مطالعات پیکره‌ای می‌تواند در ارائه تبیین‌های شناختی و نقشی کمک کند، و علاوه بر این برای تبیین‌های در زمانی و منطقه‌ای در رده‌شناسی مفیدتر است. پیکره‌های تاریخی و پیکره‌ها از زبان‌های یک منطقه می‌توانند دورنمای بهتری از تغییرات زبانی و تماس زبانی در اختیار ما قرار دهند.

ما در بخش (۴)، به مطالعات پیکره‌بنیادی اشاره می‌کنیم که در آنها به فارسی (و گاهی سایر زبان‌های ایرانی) پرداخته شده و در هر مورد نشان می‌دهیم کدام یک از ویژگی‌های مطرح در این بخش در این مطالعات برجسته است. قبل از آن، بهتر است به طور اجمالی به معرفی پیکره‌هایی بپردازیم که مطالعات اخیر رده‌شناسی پیکره‌بنیاد بر اساس آنها صورت گرفته است.

1. quantitative

2. gradient

البته ایجاد پیکره‌های زبانی فرایندی سریع است و این معرفی نمی‌تواند همه موارد را پوشش دهد (ایوانی و بیکل، زیر چاپ).

-پیکره‌های موازی^۱: این پیکره‌ها متونی یکسان در چند زبان هستند، مانند پیکره اعلامیه جهانی حقوق بشر که به ۳۷۲ زبان ترجمه شده است^۲ (واتنان و همکاران، ۲۰۱۰)، پیکره اخبار و سخنرانی‌های پارلمان اتحادیه اروپا^۳ (تیدمان، ۲۰۱۲)، پیکره زیرنویس فیلم‌ها^۴ (لوشینا، ۲۰۱۵)، و پیکره ترجمه کتاب آلیس در سرزمین عجایب و کیمیاگر (ورکرک، ۲۰۱۴).

-پیکره‌های قابل مقایسه^۵: این پیکره‌ها ترجمه متنی واحد نیستند، اما متونی با موضوع یا نوع مشابه هستند. مثلاً پیکره چندزبانه متون گفتاری برجسب‌زده‌شده^۶ (هیگ و اشنل، ۲۰۱۴ و ۲۰۱۵)، و پیکره لایپزیگ^۷ (گلدهان و همکاران، ۲۰۱۲).

-پیکره‌هایی با برجسب‌زنی یکسان^۸: این پیکره‌ها از نظام برجسب‌زنی یکسانی استفاده می‌کنند، مانند پیکره وابستگی‌های همگانی^۹ (زمان و همکاران، ۲۰۲۴).

از آن‌جا که مطالعات پیکره‌ای حوزه‌ای نوپا و در عین حال با رشدی سریع است، مطالعات اخیر در رده‌شناسی و حوزه‌های دیگر زبان‌شناسی، بسیار وابسته به پیکره است. در بخش بعد، در هر مطالعه، پیکره مورد استفاده نیز معرفی شده است.

۴. مطالعات پیکره‌بنیاد توالی کلمات

در این بخش سعی شده مطالعاتی که در آن بحث توالی کلمات زبان فارسی نیز مطرح شده معرفی شود. همه این مطالعات پیکره‌بنیاد هستند و ماهیت بین‌زبانی دارند. مرور این مطالعات باعث می‌شود تصویر درست‌تری از توالی کلمات مختلف در زبان فارسی به دست آید. مطالعات این بخش به طور مشخص به فارسی یا زبان‌های ایرانی نپرداخته‌اند، اما یکی از زبان‌های مورد بررسی در آنها فارسی و یا برخی از زبان‌های ایرانی بوده‌اند. ما ضمن مرور این مطالعات، به طور مشخص در مورد وضعیت زبان فارسی در هر توالی بحث می‌کنیم.

1. parallel

2. <http://research.ics.aalto.fi/cog/data/udhr/>.

3. <http://opus.nlpl.eu/>.

4. <https://github.com/levshina/ParTy-1.0>.

5. comparabale

6. Multilingual Corpus of Annotated Spoken Texts (Multi-CAST)

7. <https://wortschatz.uni-leipzig.de/en/download>.

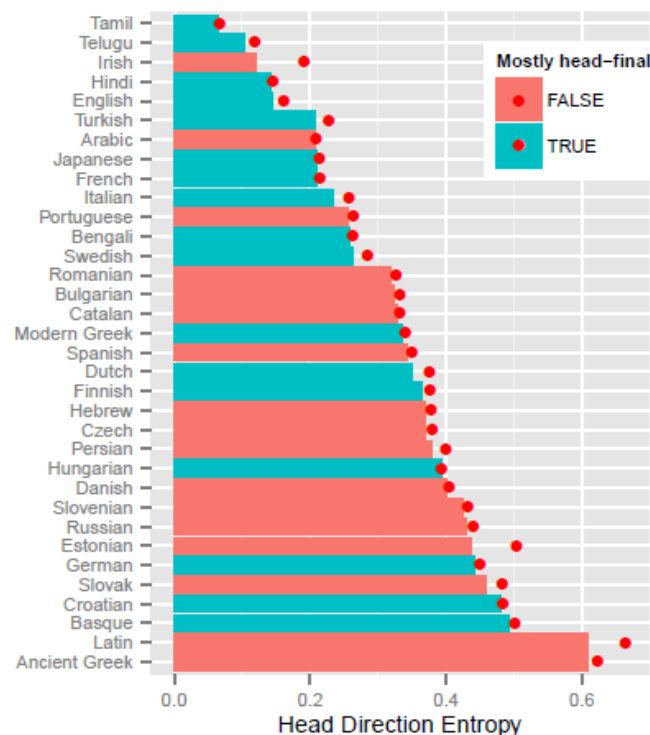
8. unified annotation

9. <https://universaldependencies.org/>.

۴-۱. توالی هسته و وابسته

فوترل و همکاران (۲۰۱۵) در یکی از نخستین مطالعاتی که به آزادی توالی کلمات اختصاص دارد، به میزان بی‌نظمی توالی هسته و وابسته در زبان‌ها پرداختند. اولین نوع داده‌های این مقاله در نمودار (۱) قابل مشاهده است. این نمودار میزان بی‌نظمی توالی هسته و وابسته را در زبان‌های مختلف نشان می‌دهد. بر این اساس، زبان‌هایی که در بالای نمودار هستند (مانند تامیلی، تلوگویی، ایرلندی، هندی و انگلیسی) دارای توالی ثابت‌تر^۱ هسته و وابسته هستند (یعنی یا هسته‌آغاز یا هسته‌پایان هستند) و هر چه به پایین نمودار می‌رویم (زبان‌هایی مانند یونانی باستان، لاتینی، باسکی) دارای توالی آزادتر هسته و وابسته هستند (یعنی برخی ساخت‌ها هسته‌آغاز و برخی هسته‌پایان هستند). نکته مهم این است که قرار دادن زبان‌ها در دو مقوله مجزا یا دو رده مجزا (مثلاً توالی هسته و وابسته ثابت یا آزاد)، تصویر نادرستی از زبان‌ها به دست می‌دهد.

نمودار ۱. بی‌نظمی جهت هسته در ۳۴ زبان (Futrell 2015: fig.1)



1. rigid

همان گونه که مشخص است زبان فارسی در نیمه پایین این نمودار قرار دارد، یعنی جزو زبان هایی است که توالی هسته و وابسته در ساخت های مختلف آن یکدست نیست و بی نظم است. همچنین نمی توان گفت ساخت های این زبان عمدتاً هسته پایان هستند (رنگ قرمز نمودار). به عبارتی با این که این زبان فعل را در انتهای جمله قرار می دهد، اما نمی توان آن را زبانی هسته پایان دانست. اگر این زبان را با ترکی و انگلیسی مقایسه کنیم متوجه می شویم که این دو زبان اکثراً دارای توالی هسته پایان هستند و بی نظمی بسیار کمتری در مقایسه با فارسی دارند. همچنین زبان عربی زبانی هسته آغاز است و بی نظمی آن مانند زبان ترکی است و بسیار کمتر از زبان فارسی است. در واقع دو زبان مهم همسایه فارسی، یعنی ترکی و عربی، یکی هسته پایان و دیگری هسته آغاز است و میزان بی نظمی هر دو حدود نصف فارسی است. البته بی نظمی زبان فارسی در این متغیر کمی بیش از حد متوسط است. همچنین این نمودار نشان می دهد زبان هایی که توالی هسته آنها قابل پیش بینی است (ثابت است)، بیشتر هسته پایان هستند (رنگ آبی).

۴-۲. توالی فاعل و مفعول

توالی فاعل و مفعول در زبان های مختلف یکی از متغیرهایی است که در بررسی پیکره بنیاد مورد توجه قرار گرفته است. فوترل و همکاران (۲۰۱۵) به بررسی توالی آزاد فاعل و مفعول و رابطه آن با حضور نشانه های حالت^۱ فاعلی-مفعولی در ۳۴ زبان پرداختند. آنها نشان دادند توالی آزاد کلمات بیشتر نیازمند این است که آن زبان نشانه حالت فاعلی-مفعولی داشته باشد، هر چند عکس آن صادق نیست، یعنی زبان هایی که نشانه حالت فاعلی-مفعولی دارند لزوماً توالی کلمات آزاد ندارند. با این که قبلاً چنین ارتباطی در مباحث رده شناختی عنوان شده بود (کیپارسکی^۲، ۱۹۹۷)، اما این مقاله این هم بستگی را به صورت کمی نشان می دهد. فوترل و همکاران (۲۰۱۵) در ادامه به بررسی بی نظمی در توالی فاعل و مفعول و ارتباط آن با داشتن نشانه حالت پرداختند (نمودار ۲).

بر اساس این نمودار زبان ها به سه گروه تقسیم شده اند: اگر در یک صیغگان کامل نشانه حالت همیشه هست، رنگ سبز نشانه آن است. رنگ قرمز نشانه زبان هایی است که نشانه حالت ندارند. رنگ آبی نیز زبان هایی را نشان می دهد که دارای مفعول نمایی افتراقی^۳ هستند، یعنی مفعول بسته به شرایطی مانند معرفگی^۴ و جاننداری^۵ می تواند نشانه حالت بگیرد. همان گونه که

1 case

2. Kiparsky, P.

3. Differential Object Marking (DOM)

4. definiteness

5. animacy

مشخص است زبان فارسی جزو زبان‌هایی است که دارای مفعول‌نمایی افتراقی است و نشانه «را» این نقش را بر عهده دارد (راسخ‌مهند و پریزاده، ۲۰۲۴). زبان‌هایی که در پایین جدول هستند دارای بی‌نظمی بالای توالی فاعل و مفعول هستند و هر چه بالاتر می‌رویم بی‌نظمی در توالی کمتر است. مشخص است که زبان‌های دارای توالی آزاد بالای فاعل و مفعول (پایین جدول) همگی دارای نشانه حالت کامل هستند (تامیلی، لاتینی، مجاری، روسی و مانند آن). اما داشتن نشانه حالت کامل به معنی داشتن توالی آزاد فاعل و مفعول نیست (مانند عربی). یعنی زبانی می‌تواند توالی کلمات ثابت داشته باشد اما دارای نشانه‌های حالت نیز باشد. زبان فارسی در کنار هندی و ترکی زبان‌های فعل‌پایانی هستند که سطح بی‌نظمی در توالی فاعل و مفعول در آنها پایین است، هرچند ترکی آزادی بیشتری دارد. یعنی فارسی معمولاً فاعل را قبل از مفعول قرار می‌دهد و توالی برعکس کم است. در این بین فارسی و ترکی دارای مفعول‌نمایی افتراقی هستند.

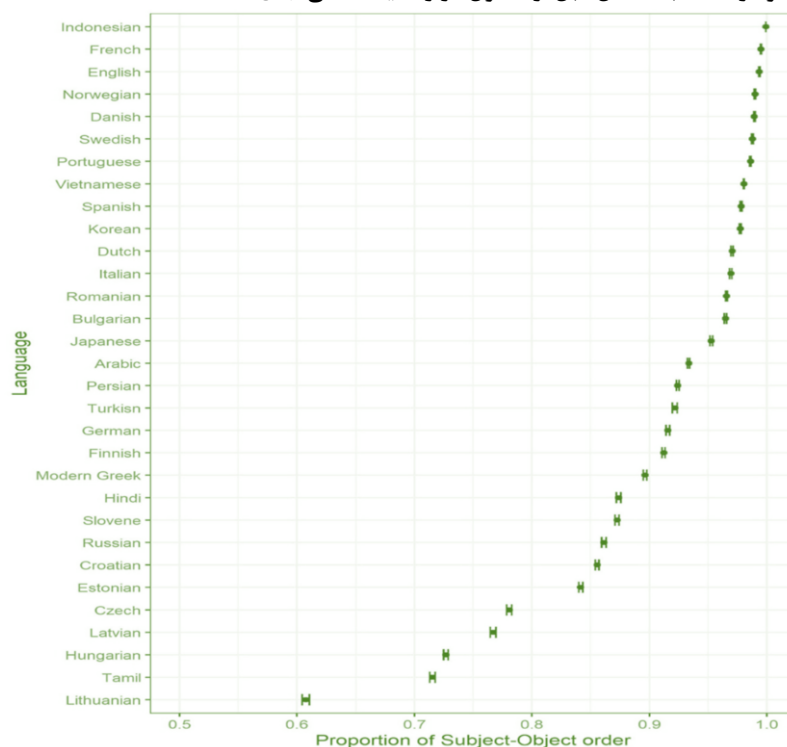
نمودار ۲. بی‌نظمی توالی فاعل و مفعول در ۳۴ زبان (Futrell et al. 2015: fig.3)



مقایسه نمودار (۱) و (۲) نشان می‌دهد اگرچه زبان فارسی در توالی هسته و وابسته جزو زبان‌های آزاد (بابی‌نظمی بالا) طبقه‌بندی می‌شود، اما این زبان در توالی فاعل و مفعول چندان آزاد نیست و در واقع توالی ثابتی دارد. شاید این تصویر کمی متفاوت از آنچه در مورد توالی فاعل و مفعول در ذهن داشتیم، باشد. و نکته دیگر اینکه، با وجود نشانه «را» همراه مفعول در فارسی، این زبان معمولاً فاعل (البته به صورت اسم) را قبل از مفعول می‌آورد و این توالی چندان تغییر نمی‌کند.

در مورد توالی فاعل و مفعول لوشینا و همکاران (۲۰۲۳: ۸۳۳) با بررسی ۳۱ زبان تا حدودی یافته‌های فوترل و همکاران (۲۰۱۵) را تایید می‌کنند. در نمودار (۳)، زبان‌هایی که در سمت راست هستند (مانند لیتوانیایی، تامیلی و مجاری) دارای توالی آزادتر فاعل و مفعول نسبت به زبان‌های سمت چپ (مانند اندونزیایی، فرانسوی، انگلیسی) هستند.

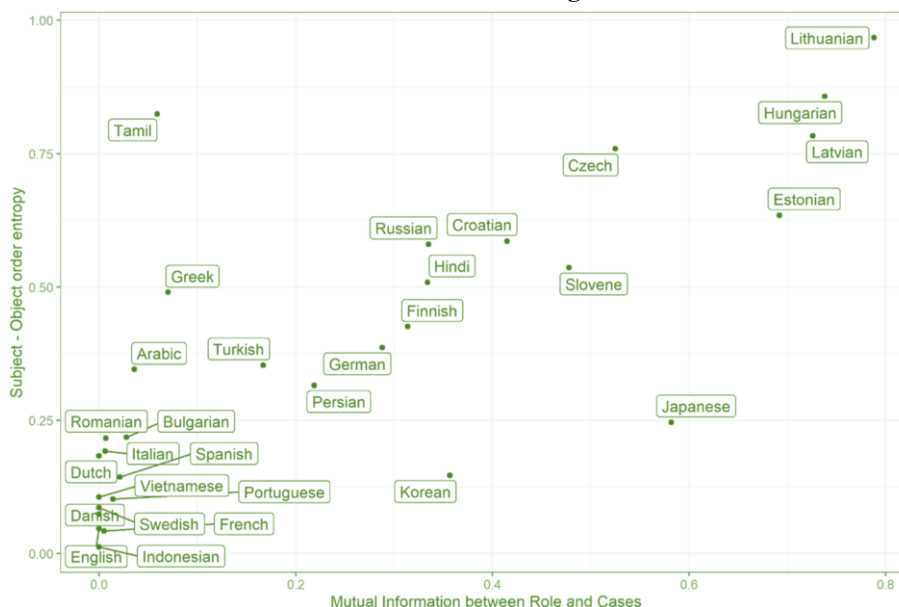
نمودار ۳. نسبت فاعل قبل از مفعول در وضعیت اسمی (Levshina et al. 2023, fig.1)



همان‌گونه که مشخص است توالی فاعل و مفعول در زبان فارسی در بیش از ۹۰ درصد موارد فاعل قبل از مفعول است. به عبارتی فارسی از این حیث زبانی دارای توالی ثابت است.

نمودار (۴) به هم‌بستگی توالی فاعل و مفعول (نمودار ۳) و میزان حالت‌نمایی (نمودار ۲) می‌پردازد. همان‌گونه که عنوان شد زبان‌هایی که توالی فاعل و مفعول آزادتری دارند (سمت راست بالای نمودار، مانند لیتوانیایی، مجاری و نظایر آن) دارای نشانه‌های حالت مشخص‌تری هستند و زبان‌های دارای توالی کلمات ثابت (سمت چپ پایین نمودار، مانند انگلیسی، دانمارکی و نظایر آن) نشانه‌ای کمتری دارند. همان‌گونه که مشخص است زبان فارسی از حیث داشتن نشانه‌ی حالت در نزدیک مرز یک چهارم پایین قرار دارد، چون مفعول‌نمایی افتراقی دارد. به همین صورت، توالی فاعل و مفعول در این زبان چندان آزاد نیست و در یک چهارم زبان‌های دارای توالی ثابت قرار می‌گیرد.

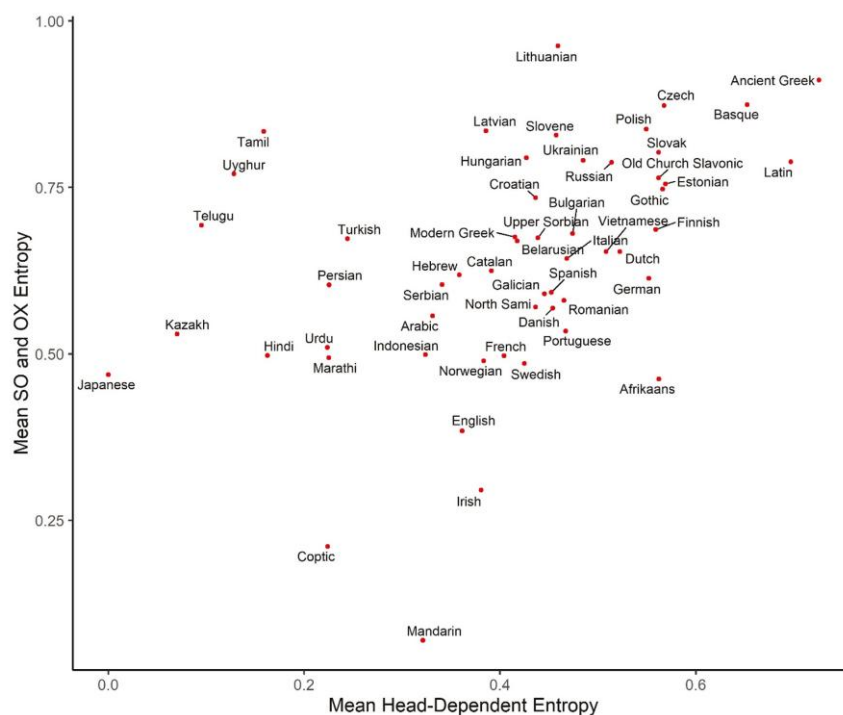
نمودار ۴. رابطه‌ی دوسویه بین بی‌نظمی فاعل و مفعول و داشتن نشانه‌ی حالت
(Levshina et al 2023: fig. 3)



لوشینا (۲۰۱۹) به بررسی هم‌بستگی بین بی‌نظمی در توالی هسته و وابسته و بی‌نظمی در توالی فاعل و مفعول و مفعول و متمم غیرفاعلی^۱ پرداخته است. نمودار (۵) این ارتباط را نشان می‌دهد.

1. oblique

نمودار ۵. بی‌نظمی در توالی هسته و وابسته و وابسته‌های فعلی در پیکره وابستگی‌های همگانی (Levshina, 2019: fig. 2) (UD)



محور افقی بی‌نظمی در توالی هسته و وابسته است و محور عمودی به بی‌نظمی در توالی فاعل و مفعول و متمم غیرفاعلی اشاره دارد. هر چه عدد بالاتر باشد، بی‌نظمی بیشتر است. همان‌گونه که مشخص است زبان‌های معدودی در هر دو محور دارای بی‌نظمی بالا هستند و در بالای سمت راست نمودار قرار دارند (مانند یونانی باستان و باسکی). در مقابل، زبان‌هایی که بی‌نظمی پایینی دارند، این بی‌نظمی یا حاصل محور افقی یا محور عمودی است، اما نه هر دو. مثلاً ماندارین به دلیل توالی ثابت هسته و وابسته و ژاپنی به دلیل توالی ثابت در محور عمودی در این بازه قرار می‌گیرند. زبان‌هایی که از حیث صرفی قوی هستند در بالای سمت راست، زبان‌های تحلیلی در وسط و زبان‌های فعل‌پایان در سمت چپ هستند. همان‌گونه که فوترل و همکاران (۲۰۱۵) عنوان کردند زبان‌هایی که توالی هسته و وابسته در آنها ثابت است معمولاً هسته‌پایان هستند. همچنین این زبان‌ها معمولاً دارای نظام حالت قوی هستند. این ویژگی‌ها، در کنار توالی کلمات ثابت، در تعیین ساختار موضوعی کمک می‌کنند. به عبارتی هرچقدر احتمال ابهام در شناسایی فاعل و مفعول بیشتر شود (نشانه‌های حالت کمتر)، توالی

آنها ثابت‌تر می‌شود. البته ابهام بیشتر به معنی بی‌نظمی کمتر است، اما عکس آن، یعنی ابهام کمتر به معنی بی‌نظمی بیشتر نیست. این فرایند می‌تواند دلیل در زمانی داشته باشد. از دست دادن نظام حالت منجر به ثابت‌تر شدن توالی کلمات می‌شود تا ابهام در شناسایی موضوع‌ها به وجود نیاید. اما ظهور نشانه‌های حالت لزوماً منجر به توالی کلمات منعطف نمی‌شود (لوشینا، ۲۰۱۹: ۵۴۷).

در نمودار فوق زبان فارسی تقریباً در وسط قرار گرفته است. یعنی زبانی است که توالی هسته و وابسته در آن چندان قابل‌پیش‌بینی نیست، و از طرفی دارای نشانه‌ی حالت برای مفعول هست. به این ترتیب توالی کلمات در زبان فارسی در میانه قرار دارد، یعنی نمی‌توان آن را آزاد دانست، اما کاملاً هم ثابت نیست. ممکن است این سوال پیش بیاید که با وجود «را» به عنوان نشانه‌ی مفعول در اکثر موارد، چرا توالی مفعول و فاعل در فارسی آزاد نیست و حتی نزدیک به ثابت است؟ استیلو (۲۰۰۹ب) نشان داده است که زبان فارسی در دوره‌ی باستان دارای نظام حالت قوی بوده است و از دوره‌ی میانه نشانه‌های حالت را از دست داده است. دبیرمقدم (۱۳۹۲: ۱۳۴) نیز عنوان می‌کند توالی کلمات در فارسی باستان آزاد بوده است، اما با از دست دادن نشانه‌های حالت در دوره‌های بعدی، به‌ویژه در دوره‌ی نو، این زبان دارای توالی کلمات ثابت شده است تا همان‌گونه که لوشینا (۲۰۱۹: ۵۴۷) گفته ابهام در شناسایی فاعل و مفعول رفع شود. اما ظهور نشانه‌های جدید حالت، مانند «را» در فارسی، باعث نشده که توالی کلمات این زبان آزادتر شود. به همین دلیل فارسی با وجود داشتن نشانه برای تمیز فاعل و مفعول، در میان زبان‌هایی است که توالی کلمات آنها تا حدودی ثابت است. زبان‌هایی که مفعول‌نمایی افتراقی دارند تقریباً در وسط نمودار هستند (مانند فارسی و ترکی)، درحالی‌که زبان‌هایی که تمایز فاعل و مفعول در آنها مشخص است، توالی آزادتری دارند (مانند یونانی باستان و لیتوانیایی). این نکته نشان می‌دهد نظام حالت و توالی کلمات ویژگی‌های ذاتی زبان نیستند، بلکه حاصل تمایل زبان‌ها برای کارایی بیشتر در ارتباط هستند. به عبارتی این تحلیل تبیین‌نقشی-سازگاری برای ویژگی‌های زبان را تایید می‌کند و مدرج بودن این ویژگی‌ها هم دلیلی برای پذیرش تبیین‌های نقشی است.

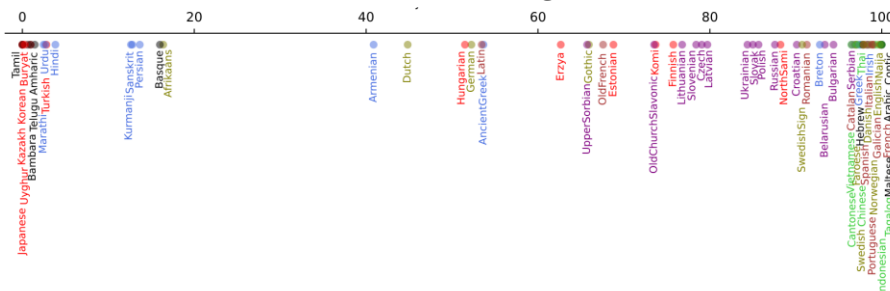
۳-۴. توالی فاعل و مفعول با فعل

گردس^۱ و همکاران (۲۰۲۱) به مطالعه‌ی توالی فاعل و مفعول در حالت اسمی، ضمیری و بندی با فعل پرداختند. نمودار (۶) میزان مفعول‌های اسمی پس از فعل را نشان می‌دهد:

1. Gerdes, K.

نمودار ۶. میزان مفعول اسمی پس از فعل. در سمت چپ زبان‌های فعل پایان و سمت راست فعل آغاز

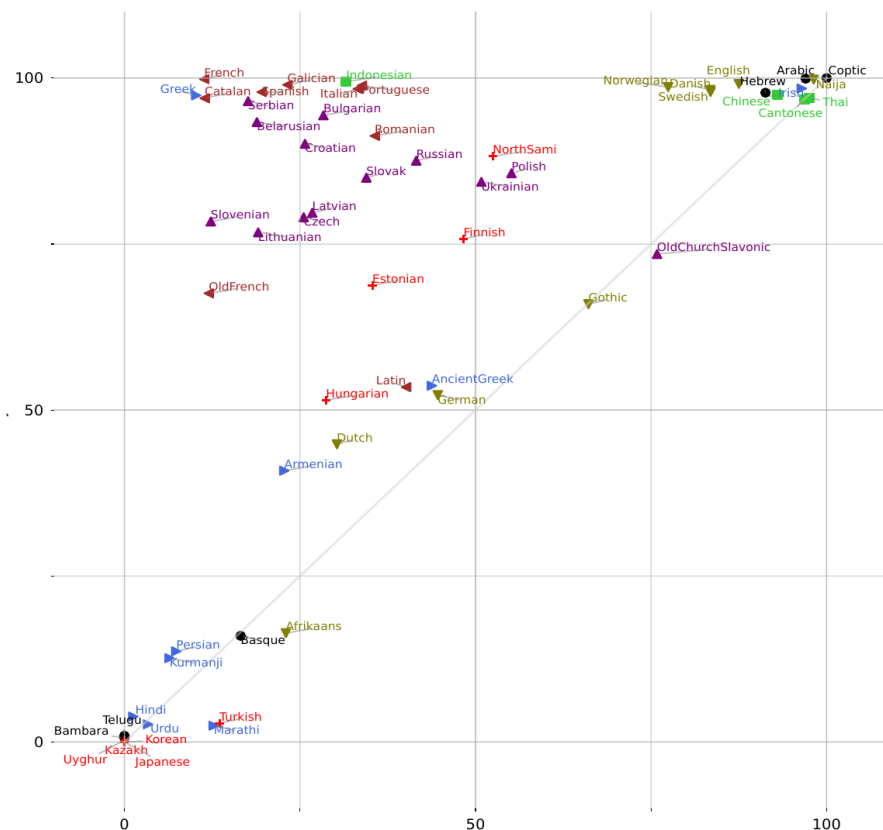
دیده می‌شوند (Gerdes 2021:1)



زبان‌های سمت چپ فعل پایان و زبان‌های سمت راست فعل آغاز هستند. همان‌گونه که مشخص است کمتر از ۲۰ درصد مفعول‌ها در زبان فارسی پس از فعل قرار می‌گیرد، در حالی که این میزان برای زبان ترکی بسیار کمتر است و در زبان عربی، تمام مفعول‌های اسمی پس از فعل هستند. از طرفی زبان آلمانی تقریباً در وسط نمودار است و ۵۳ درصد مفعول‌های آن پس از فعل قرار می‌گیرند. به این ترتیب صحبت از دو رده مشخص در این متغیر نیز اشتباه است. عربی در رده زبان‌های فعل آغاز و ترکی در رده زبان‌های فعل پایان است. فارسی به رده دوم نزدیک‌تر است، در حالی که آلمانی یک زبان دارای توالی کلمات آزاد است.

نمودار (۷) به توالی مفعول اسمی و ضمیری در کنار فعل می‌پردازد. این نمودار تصویری واضح‌تر از نمودار قبلی به دست می‌دهد. محور افقی توالی مفعول ضمیری قبل و بعد از فعل را نشان می‌دهد و محور عمودی توالی مفعول اسمی قبل از فعل در پایین و بعد از فعل در بالا را نشان می‌دهد. مثلاً در زبان فرانسوی تمام مفعول‌های اسمی پس از فعل قرار می‌گیرند، اما فقط ۷ درصد از مفعول‌های ضمیری پس از فعل قرار می‌گیرند. در مورد زبان فارسی حدود ۱۸ درصد از مفعول‌های اسمی پس از فعل و حدود ۸ درصد از مفعول‌های ضمیری پس از آن قرار می‌گیرند. در واقع متغیر اسم و ضمیر در جایگاه مفعول نسبت به فعل در زبان فارسی تاثیری ندارد و در بیش از ۹۰ درصد اوقات آنها قبل از فعل واقع می‌شوند. از این حیث زبان فارسی با ترکی قرابت نزدیک دارد و کاملاً در مقابل عربی و انگلیسی است که مفعول ضمیری و اسمی را پس از فعل قرار می‌دهند. به این ترتیب توالی غالب فارسی فاعل+مفعول+فعل است و کمتر از ۲۰ درصد مفعول‌ها پس از فعل قرار می‌گیرند. ترکی نیز در همین رده قرار می‌گیرد. اما عربی در رده‌ای کاملاً متفاوت، یعنی فعل+فاعل+مفعول قرار می‌گیرد.

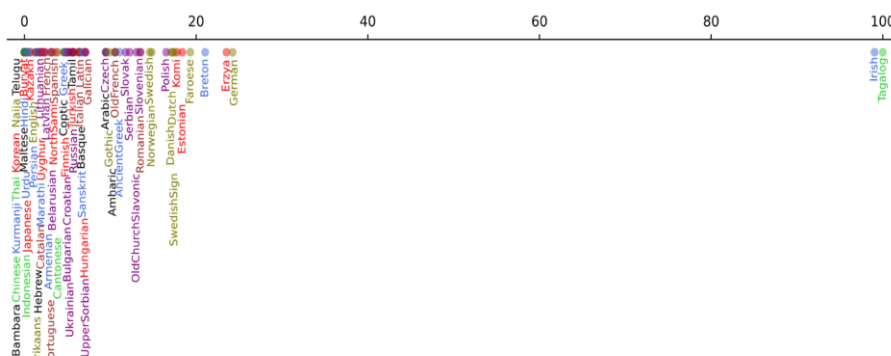
نمودار ۷. پراکندگی میزان توالی فعل + مفعول ضمیری (محور افقی) و فعل + مفعول اسمی (محور عمودی) (Gerdes et al. 2021: fig. 2)



گردس و همکاران (۲۰۲۱) عنوان می‌کنند وقتی زبانی مانند فرانسوی صد در صد مفعول اسمی را پس از فعل و ۷ درصد مفعول ضمیری را پس از فعل قرار می‌دهد (یعنی ۹۳ درصد را قبل از فعل قرار می‌دهد)، نمی‌توان گفت این زبان توالی کلمات آزاد دارد، زیرا متغیر موجود قابل پیش‌بینی است. در اصل این زبان زبانی آمیخته است. در مقابل، با توجه به نمودار (۶)، زبان‌هایی که در وسط قرار دارند (مانند آلمانی، مجاری، لاتینی و یونانی باستان) در طبقه توالی کلمات آزاد قرار می‌گیرند، زیرا رفتار جایگاه مفعول در آنها قابل پیش‌بینی نیست. بر این اساس، زبان فارسی هم که مفعول اسمی و ضمیری را در حدود ۹۰ درصد اوقات قبل از فعل و مفعول بندی را در اکثر اوقات پس از فعل قرار می‌دهد، دارای توالی کلمات آزاد نیست، زیرا رفتار آن قابل پیش‌بینی است.

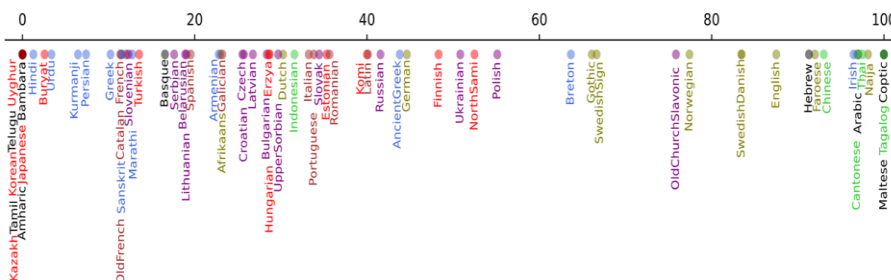
نمودار (۸) رفتار متفاوت فاعل ضمیری و توالی آن با فعل را نشان می‌دهد. همان‌گونه که مشخص است اکثر زبان‌ها (به جز ایرلندی و تاگالوگ) ضمیر فاعلی را قبل از فعل قرار می‌دهند و در مورد فارسی هم این عدد نزدیک به صد در صد است. در واقع این یک همگانی کیفی است.

نمودار ۸. پراکندگی زبان‌ها بر اساس ضمایر فاعلی (Gerdes 2021: fig.14)



اگر این نمودار را با نمودار (۹) مقایسه کنیم، مشخص می‌شود که توزیع مفعول‌های ضمیری در جایگاه قبل و بعد از فعل متوازن است.

نمودار ۹. پراکندگی زبان‌ها در توالی مفعول ضمیری و فعل (Gerdes s et al. 2021: fig. 15)



در سمت چپ جدول زبان‌هایی هستند که مفعول ضمیری را به ندرت پس از فعل قرار می‌دهند و هر چه به سمت راست می‌رویم میزان آنها پس از فعل بیشتر می‌شود. طبق نمودار (۹)، توزیع مفعول‌های ضمیری پراکنده است و در زبان فارسی حدود ۱۰ درصد از آنها پس از فعل قرار می‌گیرند. در واقع این یک همگانی کمی است. در عین حال، نمودار (۱۰) به توزیع مفعول اسمی قبل و بعد از فعل اشاره دارد.

در این حالت فارسی نیز جزو زبان‌هایی است که تمام بندهای مفعولی را پس از فعل قرار می‌دهد. مقایسه نمودارهای (۹) تا (۱۱) نشان می‌دهد چگونه مفعول با توجه به متغیر ضمیر، اسم یا بند بودن در جایگاه پیش یا پس از اسم قرار می‌گیرد. در واقع این نیز یک همگانی کمی است و با مطالعه پیکره‌ای تصویری درست‌تر از وضعیت زبان‌ها به دست می‌آید. بر این اساس نیز نمی‌توان گفت زبان فارسی دارای توالی کلمات آزاد است، زیرا میزان حضور مفعول پس از فعل کاملاً قابل‌پیش‌بینی است.

۴-۴. رابطه مطابقه و توالی کلمات

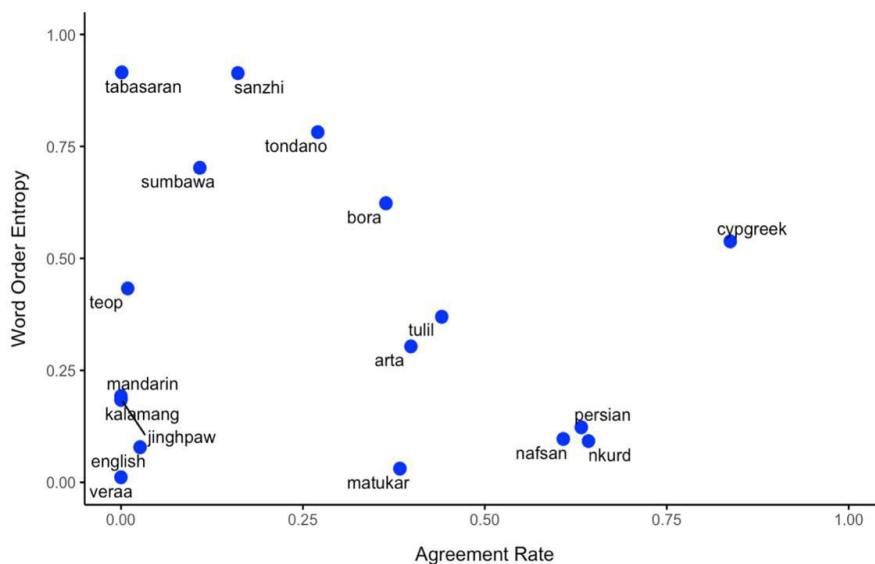
اسپیرگات^۱ (۲۰۲۵) به بررسی تعامل بین میزان مطابقه^۲ و میزان آزادی توالی کلمات بر اساس داده‌های پیکره‌ای پرداخته است. در بخش (۲-۴) عنوان شد که حالت‌نمایی و توالی کلمات آزاد دارای هم‌بستگی منفی هستند، یعنی اگر حالت‌نمایی بیشتر باشد توالی کلمات آزادتر است (البته عکس آن صادق نیست). در مورد مطابقه، داده‌ها نشان می‌دهد زبان‌هایی که نشانه مطابقه بیشتری دارند، توالی کلمات ثابت‌تری دارند. یعنی مطابقه عکس حالت‌نمایی عمل می‌کند. نمودار (۱۲) رابطه مطابقه و توالی کلمات را در ۱۸ زبان نشان می‌دهد.

همان‌گونه که مشخص است محور افقی میزان مطابقه و محور عمودی میزان بی‌نظمی در توالی کلمات را نشان می‌دهد. در این نمودار، فارسی دارای مطابقه قوی است و در یک‌چهارم سوم قرار دارد، اما میزان بی‌نظمی در توالی کلمات (توالی فاعل و مفعول) حدود ۲۰ درصد است. از طرفی زبان‌هایی که بی‌نظمی بالاتری دارند (مانند تبرسانی و سانژی)، دارای مطابقه ضعیف‌تری هستند. نمودار (۱۳) با بررسی خانواده زبانی زبان‌های نمودار (۱۲)، نشان می‌دهد که این دو متغیر هم‌بستگی منفی دارند، یعنی یا میزان مطابقه بالاست یا بی‌نظمی بالاست. در واقع این دو، دارای نقش مشابه در زبان هستند و در تعیین جریان اطلاعات یا همان مدیریت اطلاعی^۳ نقش دارند.

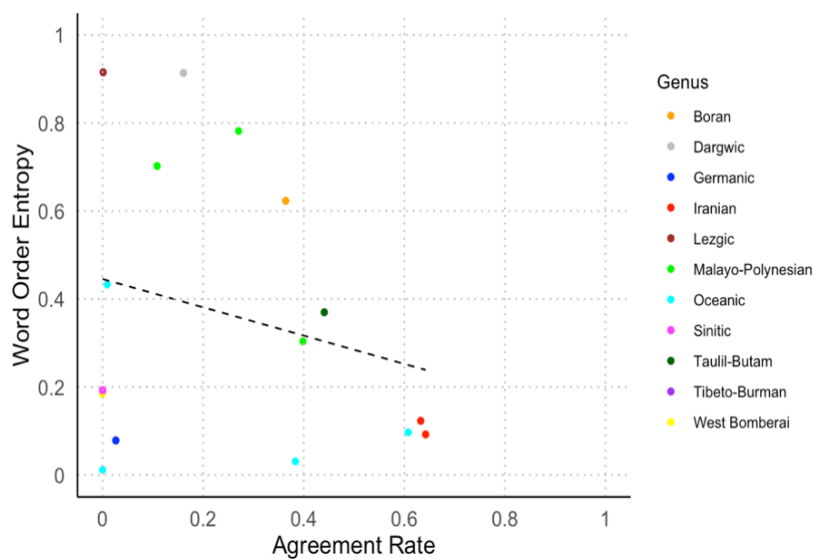
اسپیرگات (۲۰۲۵: ۱۵) عنوان می‌کند هرگاه میزان مطابقه افزایش پیدا می‌کند، میزان بی‌نظمی کمتر می‌شود. این ویژگی در مقوله حالت متفاوت بود، زیرا افزایش نشانه‌های حالت به معنی بی‌نظمی بالاتر در توالی کلمات است. در واقع، غنی بودن مطابقه به معنای آزادتر بودن توالی کلمات در زبان فارسی نیست.

1. Spirgath, A.
2. agreement
3. information management

نمودار ۱۲. رابطه بی‌نظمی در توالی کلمات و مطابقت. محور افقی میزان مطابقت و محور عمودی میزان بی‌نظمی (Spiragath, 2025:fig.1)



نمودار ۱۳. مدل خطی تعامل دو عامل مطابقت و بی‌نظمی در توالی کلمات (Spiragath, 2025:fig.1)



۴-۵. چند توالی یک‌بعدی

گرینبرگ (۱۹۶۳) در همگانی شماره ۱۹ خود فقط یک متغیر را در نظر گرفته است:

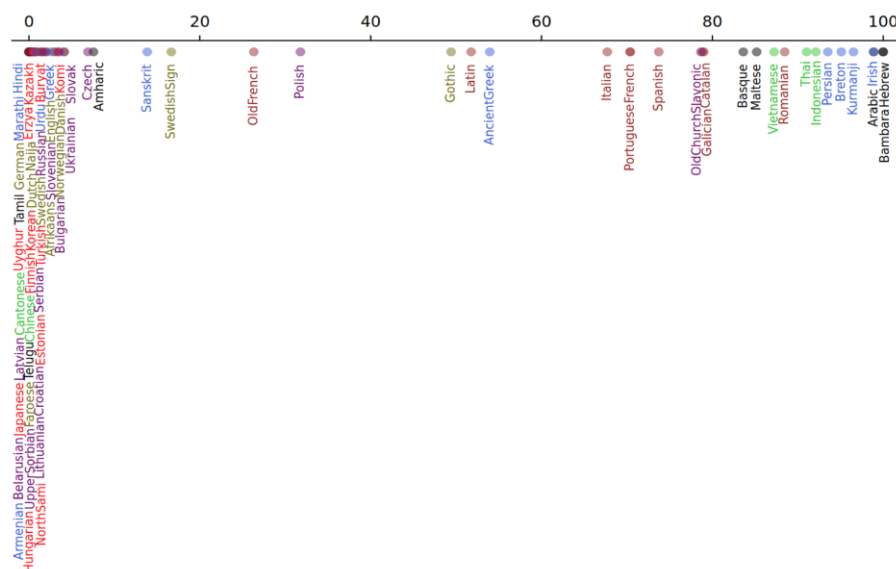
همگانی شماره ۱۹. اگر در زبانی توالی غالب این باشد که صفت پس از اسم بیاید، ممکن است تعداد کمی از صفات قبل از اسم هم قرار بگیرند، اما اگر توالی غالب صفت قبل از اسم باشد، استثنایی وجود ندارد.

بر این اساس، زبان‌های هسته‌پایان (مانند ترکی)، نباید تعداد زیادی ساخت وصفی هسته‌آغاز داشته باشند. نمودار (۱۴) نشان می‌دهد همگانی گرینبرگ تا حد زیادی تایید می‌شود. در سمت چپ زبان‌های هسته‌پایان دیده می‌شوند و همان‌گونه که دیده می‌شود تقریباً به صورت یکدست قرار دارند. اما توزیع زبان‌های هسته‌آغاز این گونه نیست و تقریباً در تمام نمودار پخش است.

همان‌گونه که دیده می‌شود زبان فارسی در این نمودار در سمت راست دیده می‌شود و حدود ۹۵ درصد از صفات را پس از اسم قرار می‌دهد. عربی از این حیث یکدست‌تر است و تقریباً تمام صفات را پس از اسم قرار می‌دهد، و در مقابل ترکی در طرف مقابل قرار دارد. به این ترتیب می‌توان گفت این تحلیل کمی، همگانی ۱۹ گرینبرگ را تایید می‌کند و شواهد آماری مشخص برای آن فراهم می‌کند.

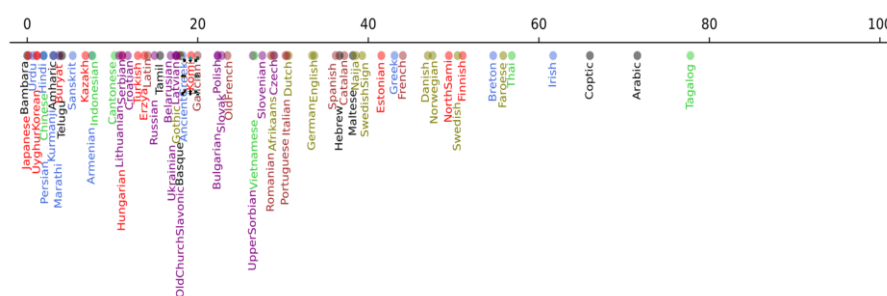
نمودار ۱۴. پراکندگی زبان‌ها بر اساس توالی صفت و اسم. سمت چپ زبان‌های صفت+اسم و سمت

راست زبان‌های اسم+صفت (Gerdes et al. 2021:fig.12)

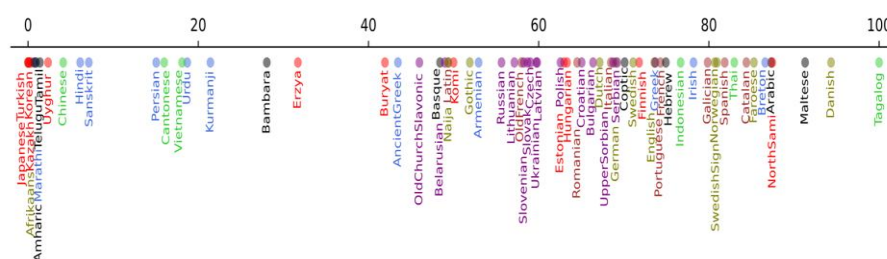


بررسی جایگاه قید واژگانی و بند قیدی نیز نکته‌ی مشابهی را روشن می‌سازد. قیدهای واژگانی بیشتر در سمت چپ بند قرار می‌گیرند (نمودار ۱۵)، اما بندهای قیدی بیشتر تمایل به سمت راست دارند (نمودار ۱۶).

نمودار ۱۵. پراکندگی زبان‌ها بر اساس جهت حضور معرف‌های قیدی (Gerdes et al. 2021:fig.18)



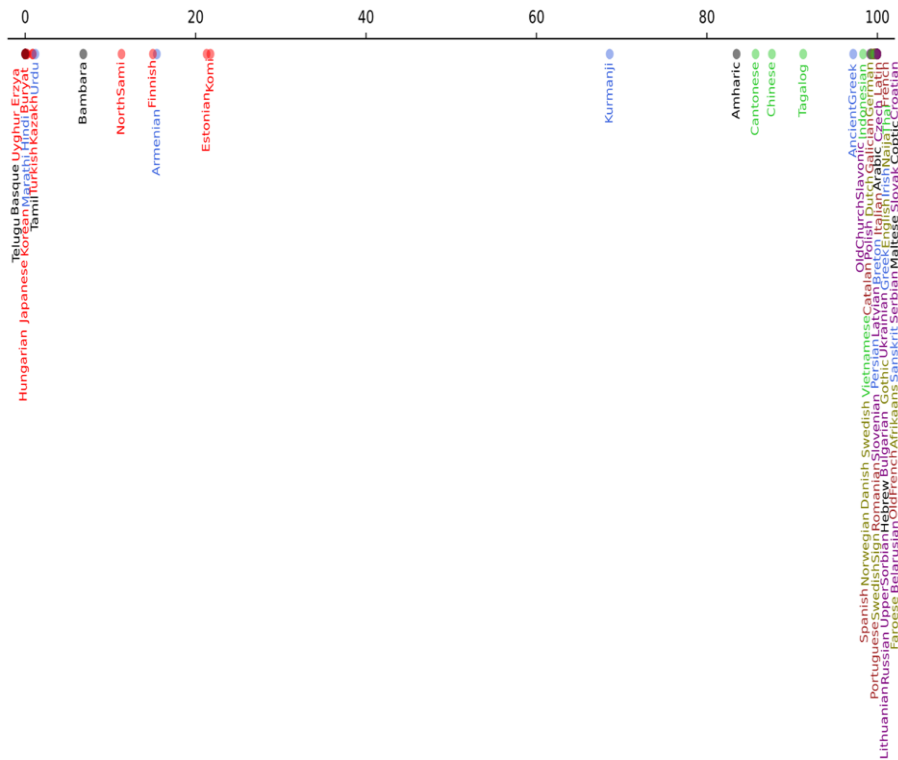
نمودار ۱۶. پراکندگی زبان‌ها بر اساس جهت حضور بندهای قیدی (Gerdes et al. 2021:fig.19)



فارسی کمتر از ۴ درصد از قیدهای واژگانی را پس از فعل قرار می‌دهد، اما این عدد در بندهای قیدی حدود ۱۷ درصد است. عربی از این حیث در نقطه‌ی مقابل فارسی است و اکثر قیدها را در جایگاه سمت راست قرار می‌دهد ولی ترکی رفتاری شبیه با فارسی دارد. توالی یک‌بعدی سوم وجود پیش‌افزافه یا پس‌افزافه در زبان‌هاست. نمودار (۱۷) نشان می‌دهد زبان‌ها یا انتخاب می‌کنند پیش‌افزافه داشته باشند (سمت راست) یا پس‌افزافه (سمت چپ).

زبان‌های بسیار کمی هستند که این قاعده را نقض می‌کنند، به همین دلیل وسط نمودار خالی است. فارسی از این حیث در سمت راست در کنار عربی است و در مقابل ترکی قرار دارد. داده‌های این بخش نشان می‌دهد تحلیل پیکره‌ای در برخی موارد همگانی‌های دستوربنیاد را تایید می‌کند، اما بی‌نظمی موجود در بین زبان‌ها و درون زبانی را بهتر نشان می‌دهد.

نمودار ۱۷. پراکندگی زبان‌ها بر اساس جهت حضور پیش‌افزافه (در چپ) و پس‌افزافه (در راست)
(Gerdes et al. 2021:fig.13)



۴-۶. سازه‌های بعد از فعل

یکی از مطالعات رده‌شناختی پیکره‌بنیاد که به طور مشخص به بررسی زبان‌های ایرانی و از جمله فارسی پرداخته مطالعه هیگ و همکاران (۲۰۲۴ الف) است. این مطالعه به بررسی حضور برخی از سازه‌ها در جایگاه پس از فعل اختصاص دارد که فرامر^۱ (۱۹۸۱) آن را پدیدهٔ پس‌افعل^۲ و هیگ و راسخ‌مهند (۲۰۱۹) و راسخ‌مهند و همکاران (۲۰۲۴) عناصر پس‌گزاره^۳ می‌نامند. راسخ‌مهند و همکاران (۲۰۲۴) با استفاده از پیکرهٔ گفتاری هم‌بام (هیگ و راسخ‌مهند، ۲۰۲۲) به مطالعهٔ سازه‌های پس از فعل در زبان فارسی پرداخته‌اند. این پیکره مجموعه‌ای از گفتار غیررسمی است. آنها جدول زیر را بر اساس بسامد نقش سازهٔ پس از فعل ارائه کردند:

1. Frommer, P.
2. postverbal
3. post-predicate elements

جدول ۳. بسامد سازه‌های پس از فعل (راسخ‌مهند و همکاران، ۲۰۲۴: ۲۰۴)

نوع سازه	کل	فعل — سازه غیرفعل	درصد فعل — سازه غیرفعل
هدف سببی	۶۰	۵۵	۹۱/۷ %
هدف	۲۰۶	۱۶۷	۸۱/۱ %
مفعول مستقیم	۴۳۷	۱۸	۴/۱ %
مکانی	۱۴۶	۲۹	۱۹/۹ %
ازی	۵۰	۵	۱۰ %
سایر (غیرمشخص)	۳۱۵	۹۸	۳۱/۱ %
همراهی	۴۶	۱۲	۲۶/۱ %
ایزار	۲۸	۴	۱۴/۳ %
متمم (شدن)	۲۴	۶	۲۵ %
مخاطب	۶۹	۸	۱۱/۶ %
بهره‌ور	۱۱	۳	۲۷/۳ %
گیرنده + بهره‌ور	۱۳	۲	۱۵/۴ %
متمم فعل ربطی	۲۰۵	۵	۲/۴ %
محرك	۴	۱	۲۵ %
گیرنده	۱۱	۰	۰ %

همان‌گونه که از جدول مشخص است، سازه‌های دارای دو نقش هدف^۱ و هدف سببی^۲ بیشترین بسامد را در میان سازه‌های پس از فعل داشته‌اند و تفاوت آنها با سایر نقش‌ها به حدی است که باید آنها را در طبقه مجزایی قرار داد. این یافته با دیدگاه فرامر (۱۹۸۱) همسو است و نشان می‌دهد در فارسی گفتاری، بیش از ۸۰٪ سازه‌های دارای نقش هدف و هدف سببی در جایگاه پس از فعل قرار می‌گیرند. نمونه‌هایی از این سازه‌ها در مثال‌های زیر آمده است:

(۱) رد کرده بودم توی چوب.

(۲) این را بگذار روی سرت.

(۳) رفتم دکتر.

(۴) پریدند روی میز.

جدول (۴) پربسامدترین افعال حرکتی سببی را نشان می‌دهد و مشخص می‌کند به چه میزان سازه هدف در این افعال پس از فعل آمده‌اند:

1. goal

2. caused goal

جدول ۴. پربسامدترین افعال حرکتی سببی (راسخ‌مهند و همکاران، ۲۰۲۴: ۲۰۶)

افعال حرکتی سببی	کل	فعل - سازه غیرفعل	درصد کل فعل - سازه غیرفعل
آوردن	۱۱	۱۱	۱۰۰ %
گذاشتن	۱۲	۱۱	۹۲ %
بردن	۱۱	۱۰	۹۱ %

همان‌گونه که از جدول مشخص است، برای فعل «آوردن» این میزان صد درصد بوده است، یعنی تمام سازه‌های هدف با این فعل پس از فعل آمده‌اند. در مورد دو فعل «گذاشتن» و «بردن» نیز این میزان بالاست. مثال‌های زیر، هدف را در افعال حرکتی نشان می‌دهد:

(۵) رفتم دکتر.

(۶) پریدند روی میز.

بسامد هدف‌های پس از فعل در افعال حرکتی با گروه هدف‌ها در افعال حرکتی سببی یکسان است و در مورد دو فعل «رفتن» و «آمدن» به این صورت است:

جدول ۵. پربسامدترین افعال حرکتی (راسخ‌مهند و همکاران، ۲۰۲۴: ۲۰۷)

افعال حرکتی	کل	فعل - سازه غیرفعل	درصد کل فعل - سازه غیرفعل
رفتن	۸۴	۶۹	۸۲ %
آمدن	۳۲	۲۶	۸۱ %

به این ترتیب می‌توان گفت در فارسی گفتاری جایگاه بی‌نشان برای سازه‌های نقش هدف و هدف سببی، جایگاه پس از فعل است. جدا از سازه هدف، سایر سازه‌هایی که به مکان اشاره می‌کنند، مانند مکانی^۱ و «از»ی^۲ هم به نسبت بسامد بالایی دارند:

(۷) توقف داشته است تو اندیمشک.

(۸) شهید شد تو جبهه.

(۹) دستش را گرفت از من.

موارد زیر، موضوع‌های غیراصلی هستند که نقش مکان ندارند و پس از فعل قرار گرفته‌اند.

1. locative

2. ablative

(۱۰) که یکی بشیم با همدیگه.

(۱۱) باز زندگی می‌کنم باهاشون.

(۱۲) بعد اشاره کرد با دست.

(۱۳) حسودی نکنند به همدیگه.

در مجموع می‌توان گفت در سازه‌های غیر از مفعول مستقیم، محتمل‌ترین سازه‌هایی که به جایگاه پس از فعل می‌روند هدف و هدف سببی است، و البته سازه‌های دیگری که به مکان اشاره دارند نیز در حدود ۲۰ درصد در این جایگاه قرار می‌گیرند. به عبارتی دیگر در این جایگاه سازه‌هایی که به مکان اشاره دارند بیش از سازه‌های دیگر قرار دارند.

راسخ‌مهند و همکاران (۲۰۲۴) تحلیلی از تاثیر عوامل مختلف بر حضور سازه پس از فعل انجام داده‌اند. عوامل مورد بررسی شامل نقش، سبک (رسمی و غیررسمی)، صورت (اسمی و ضمیری)، نشان‌گذاری (وجود یا نبود حرف اضافه)، انسان‌بودن (انسان در برابر غیرانسان) و وزن سازه است. نتایج نشان می‌دهد مهم‌ترین و موثرترین عامل، نقش معنایی سازه است و در این بین هدف و هدف سببی رفتاری کاملاً متفاوت با سایر نقش‌ها دارند. عوامل دیگر همچون وزن سازه، جاننداری و سبک تاثیر چندانی بر این فرایند ندارند.

هیگ و همکاران (۲۰۲۴ الف) توالی سازه پس از فعل را در چندین زبان آسیای غربی بررسی کرده‌اند. این پژوهش‌ها با بهره‌گیری از دو پیکره^۱ Wowa و HamBam^۲ انجام شده است. Wowa یک پیکره چندزبانه حاوی حدود ۴۰ مجموعه داده از زبان‌های مختلف در منطقه انتقالی غرب آسیا^۳ است که برای بررسی پدیده‌های خاص این منطقه طراحی شده است. پیکره HamBam (پیکره همدان-بامبرگ، فارسی گفتاری معاصر) پیکره‌ای است از فارسی گفتاری غیررسمی، که بر اساس الگوی Multi-CAST بنا شده است. این پیکره امکان بررسی‌های دقیق‌تر در مورد توالی کلمات، آهنگ گفتار و سبک‌های مختلف زبان فارسی را فراهم می‌کند.

در کنار مطالعه مجزای هر یک از زبان‌های پیکره، هیگ و همکاران (۲۰۲۴ الف) به چند پرسش کلی درباره سازه‌های پس از فعل در این زبان‌ها پرداخته‌اند. یکی از پرسش‌ها این است که آیا الگوی مشترکی برای سازه‌های پس از فعل در این زبان‌ها بر اساس نقش دستوری سازه قابل مشاهده است؟ پرسش دوم هم این است که در زبان‌های غیرایرانی این منطقه وضعیت

1. <https://multicast.aspra.uni-bamberg.de/resources/wowa/>.

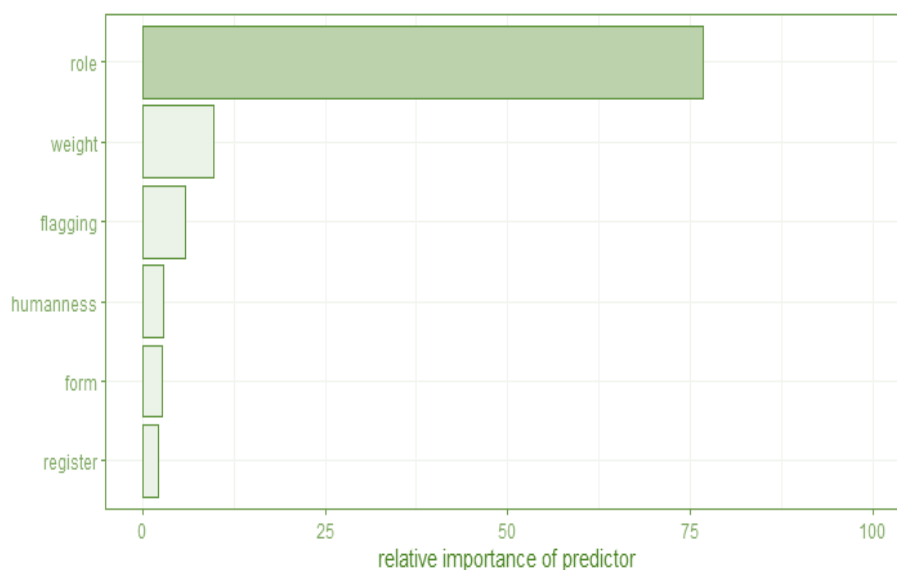
2. <https://multicast.aspra.uni-bamberg.de/resources/hambam/>.

3. WATZ: Western Asia Transition Zone

راسخ‌مهند | ۴۳

سازه‌های پس از فعل چگونه است و آیا می‌توان الگوی مشخصی مشاهده کرد؟ جدول (۶) به نوعی الگوهایی را از سازه‌های پس از فعل در زبان‌های مورد بررسی نشان می‌دهد.

نمودار ۱۸. اهمیت نسبی عوامل مختلف بر سازه‌های پس از فعل (راسخ‌مهند و همکاران، ۲۰۲۴: ۲۲۰)



اولین گرایش کلی که از جدول ۳ قابل استخراج است این است که سازه‌های دارای نقش هدف (ستون آبی) بیشترین میزان قرار گرفتن در جایگاه پس از فعل را دارند، و در جایگاه‌های بعدی به ترتیب گیرنده و مخاطب (ستون زرد)، سایر متمم‌های غیرفاعلی (ستون سبز) و مفعول مستقیم (ستون قرمز) قرار می‌گیرند. به عبارتی در بیش از دو سوم زبان‌ها سازه‌های هدف در اکثر مواقع پس از فعل قرار می‌گیرد و در واقع هدف‌ها بیشترین احتمال وقوع در این جایگاه را دارند. این عدد برای زبان فارسی معاصر ۸۴ درصد است. در مورد سوال دوم، یعنی دلیل وجود سازه‌های پس از فعل و اهمیت تماس زبانی، در بخش (۵) صحبت می‌کنیم.

جدول ۳. سلسله‌مراتب سازه پس از فعل در ۳۳ گونه زبانی در منطقه انتقالی غرب آسیا (Haig et al. 2024a)

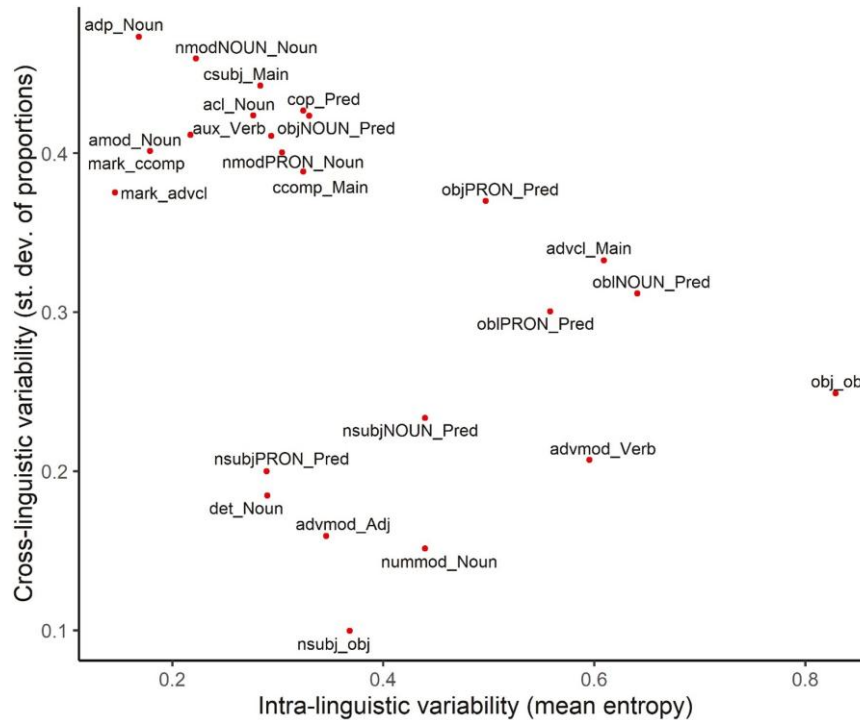
DOCULECT	GOALS	% post-verbal		
		RECIPIENTS + ADDRESSEES	OTHER OBLIQUES	DIRECT OBJECTS
◆ Laz (Arhavi)	4	6	1	3
◆ Persian (New, Early Classical)	5	0	9	2
● Oghuz (Ankara)	7	19	7	2
● Oghuz (Erzurum)	38	8	9	7
◆ Balochi (Turkmen)	48	14	6	2
◆ Kurdish (Northern, Ankara)	59	13	6	0
✱ Kholosi (Kholos)	62	39	1	1
◆ Mazandarani (Kordxeyl)	63	12	8	3
◆ Balochi (Coastal)	63	28	18	7
◆ Bashkardi (Northern)	63	56	31	27
● Oghuz (Bayat)	64	43	13	4
◆ Zazaki (Cewlig)	91	15	0	5
● Oghuz (Tabriz)	75	21	13	1
◆ Persian (New)	84	25	19	5
◆ Kurdish (Southern, Bijar)	97	27	11	2
◆ Tat (Daykuseu)	78	34	30	15
◆ Kurdish (Northern, Mus)	89	36	5	3
◆ Bashkardi (Southern)	80	38	36	11
◆ Vafsi (Gurchani)	88	38	11	2
◆ Kurdish (Northern, Lachin)	81	40	4	2
◆ Balochi (Koroshi)	90	42	27	2
◆ Talysi (Lerik)	72	45	26	2
● Oghuz (Gagauz)	73	47	38	51
◆ Zazaki (Siwereg)	100	52	8	5
◆ Tati (Hazarrudi)	92	53	16	3
● Oghuz (Qashgai)	71	88	16	8
◆ Kurdish (Central, Sanandaj)	94	95	22	1
◆ Gorani (Gawraju)	96	82	36	5
▲ NENA (Jewish, Sanandaj)	92	81	40	5
◆ Kumzari (Musandam)	100	97	87	7
▲ NENA (Christian, Barwar)	96	100	74	83
▲ NENA (Jewish, Dohok)	99	98	93	90
▲ Arabic (Jewish, Baghdad)	100	100	85	97

Table 4: Post-possibility hierarchy of 33 doculects, based on a classification threshold of 66% post-verbal positioning. Goals and direct objects consider only nominal expressions; recipients, addressees, and other oblique roles (locations, sources, instrumentals, benefactives, and comitatives) also include pronominal expressions. Two doculects with less than 8 observations in each role category are excluded. See Appendix B for raw data. NENA = Northeastern Neo-Aramaic.

۴-۷. بی‌نظمی بین‌زبانی و درون‌زبانی

لوشینا (۲۰۱۹) به تفکیک توالی‌هایی را نشان داده است که بی‌نظمی بین‌زبانی یا بی‌نظمی درون زبانی بیشتر یا کمتری دارند. نمودار (۲۰) این توالی‌ها را نشان می‌دهد.

نمودار ۲۰. گوناگونی درون‌زبانی (محور افقی) و بین‌زبانی (محور عمودی) از برخی توالی‌های منتخب (Levshina, 2019: fig.4)



محور افقی توالی‌هایی را نشان می‌دهد که درون هر زبان می‌توانند بی‌نظم باشند. مثلاً توالی مفعول و متمم بیشترین میزان بی‌نظمی را درون یک زبان می‌تواند داشته باشد، و توالی حرف اضافه و اسم کمترین میزان را دارد. به عبارتی یک زبان اغلب یا پیش‌اضافه دارد یا پس‌اضافه، به همین دلیل بی‌نظمی درون زبان کم است. محور عمودی بی‌نظمی بین‌زبانی را مشخص می‌کند. به همین ترتیب، توالی حرف اضافه و اسم بیشترین بی‌نظمی بین‌زبانی را دارد زیرا تقریباً نصف زبان‌ها پیش‌اضافه و نصف زبان‌ها پس‌اضافه دارند. به این ترتیب توالی‌های بالای سمت چپ بیشترین بی‌نظمی بین‌زبانی را دارند، فعل کمکی و فعل اصلی، مفعول و فعل، و نظایر آن. مواردی که در قسمت پایین چپ قرار دارند، کمترین بی‌نظمی بین‌زبانی و درون‌زبانی را دارند، مانند توالی فاعل و مفعول اسمی، حرف تعریف و اسم، قید و صفت و نظایر آن. در قسمت راست جدول، متمایل به وسط، توالی‌هایی هستند که بیشترین بی‌نظمی درون‌زبانی را دارند، مانند توالی مفعول و متمم غیرفاعلی، قید و فعل، یا متمم غیرفاعلی و فعل. در مطالعات رده‌شناختی دستوربنیاد بیشتر توالی‌هایی مورد نظر بوده که بی‌نظمی بین‌زبانی دارند و چندان

به بی‌نظمی‌های درون‌زبانی توجه نشده است. اما رده‌شناسی پیکره‌بنیاد به بی‌نظمی‌های درون‌زبانی نیز به همان اندازه اهمیت می‌دهد.

۴-۸. خلاصه بخش

در بخش (۴) به مطالعاتی اشاره شد که به‌صورت پیکره‌بنیاد به بی‌نظمی در توالی کلمات در زبان‌های مختلف پرداخته‌اند. در این مطالعات زبان فارسی نیز یکی از زبان‌های مورد بحث بوده است. داده‌های این بخش به طور کلی نشان می‌دهد اکثر همگانی‌های گرینبرگی که زبان‌ها را در رده‌های مختلف قرار می‌دهد، به نوعی نگاه ساده شده به زبان‌هاست. در واقع باید وجود بی‌نظمی‌های بین‌زبانی و درون‌زبانی را پذیرفت و در پی همگانی‌های مقوله‌ای (صفر و صدی) در رده‌شناسی زبان‌ها نبود. در این میان، نکاتی نیز درباره توالی کلمات در زبان فارسی آشکار شد که به صورت خلاصه این‌گونه است:

- زبان فارسی جزو زبان‌هایی است که توالی هسته و وابسته در ساخت‌های مختلف آن یکدست نیست و بی‌نظم است. همچنین نمی‌توان گفت ساخت‌های این زبان عمدتاً هسته‌پایان هستند. به عبارتی با این‌که این زبان فعل را در انتهای جمله قرار می‌دهد، اما نمی‌توان آن را زبانی هسته‌پایان دانست.

- زبان فارسی جزو زبان‌های فعل‌پایانی است که سطح بی‌نظمی در توالی فاعل و مفعول در آنها پایین است. یعنی فارسی معمولاً فاعل را قبل از مفعول قرار می‌دهد و توالی برعکس کم است. اگرچه زبان فارسی در توالی هسته و وابسته جزو زبان‌های دارای بی‌نظمی بالا طبقه‌بندی می‌شود، اما این زبان در توالی فاعل و مفعول چندان آزاد نیست و در واقع توالی ثابتی دارد.

- مفعول‌های اسمی گرایش بیشتری به جایگاه پس از فعل دارند و در مورد زبان فارسی هم این عدد حدود ۱۸ درصد است. یعنی احتمال وقوع مفعول اسمی نسبت به مفعول ضمیری در جایگاه پس از فعل در زبان فارسی حدود دو برابر بیشتر است. در مورد مفعول بندی، مشخص شد که احتمال وقوع این بندها در جایگاه پس از فعل بسیار بالاتر است.

- با این‌که فارسی دارای مطابقت قوی است، اما میزان بی‌نظمی در توالی کلمات (توالی فاعل و مفعول) حدود ۲۰ درصد و میزان آن کم است.

- فارسی در حدود ۹۵ درصد از صفات را پس از اسم قرار می‌دهد.

- فارسی کمتر از ۴ درصد از قیدهای واژگانی را پس از فعل قرار می‌دهد، اما این عدد در بندهای قیدی حدود ۱۷ درصد است.

- در فارسی ۸۴ درصد از سازه‌های هدف پس از فعل قرار می‌گیرند و در واقع هدف‌ها بیشترین احتمال وقوع در این جایگاه را دارند. به عبارتی بی‌نظمی در جایگاه پس از فعل در این زبان بسیار بالاست.

۵. تبیین منطقه‌ای

در مورد چرایی نظم و بی‌نظمی موجود در توالی کلمات، هم در رده‌شناسی دستوربنیاد و هم در رده‌شناسی پیکره‌ای تبیین‌های نقشی مختلفی عنوان شده است. این تبیین‌ها عمدتاً بر اساس ساختار اطلاعاتی و مدیریت اطلاعات در بند، دستوری‌شدگی و دلایل تاریخی، و عوامل شناختی موثر بر پردازش طبقه‌بندی شده‌اند (لوشینا، ۲۰۱۹ و ۲۰۲۱؛ لوشینا و همکاران، ۲۰۲۳). من در این بخش به تبیین‌های مذکور نمی‌پردازم، بلکه سعی می‌کنم با توجه به داده‌های این مقاله، به نقش جغرافیا، یا همان رده‌شناسی منطقه‌ای و یا به عبارتی نقش تماس زبانی در برخی از این بی‌نظمی‌ها بپردازم.

همهٔ زبان‌ها معمولاً در طول حیات خود با زبان‌های مختلفی همسایه هستند و به عبارتی بین آنان تماس زبانی برقرار است. زبان‌های ایرانی به دلیل تاریخ طولانی خود و وسعت جغرافیایی، در طول تاریخ با زبان‌های مختلفی از خانواده‌های زبانی مختلف همسایه بوده‌اند، و گاهی مورد هجوم آنها واقع شده‌اند، و در سالیان طولانی با آنان هم‌زیستی داشته‌اند. این تماس زبانی در حوزه‌های مختلف زبان اثر خود را می‌گذارد و باعث دادوستد بین زبان‌ها می‌شود. آشکارترین این دادوستد را می‌توان در وام‌گیری واژگانی مشاهده کرد و با مطالعهٔ واژگان مختلف هر زبان، واژه‌هایی یافت که از زبان‌های دیگر قرض گرفته است. تاثیر تماس زبانی محدود به حیطهٔ وام‌گیری واژگانی نیست و بر ساخت آوایی و واجی (نغزگوی کهن ۱۳۹۲؛ استیلو، ۱۹۹۴)، صرفی (راسخ‌مهند، ۱۴۰۴) و نحوی زبان‌ها نیز تاثیر می‌نهد. یکی از معدود مجموعه مقالات در این زمینه را سساتو^۱ و همکاران گردآوری کرده‌اند. البته مطالعهٔ این تاثیرات مشکل‌تر از حوزهٔ قرض‌گیری واژگانی است. از آن‌جاکه زبان‌های ایرانی، و از جمله زبان فارسی، در حیات بلندمدت خود با زبان‌های زیادی در تماس بوده‌اند، گمان می‌رود این همسایگی بر توالی کلمات این زبان و الگوهایی که در این مقاله به آنها اشاره کردیم، تاثیر گذاشته باشد. به عبارتی می‌توان فرض کرد برخی از این بی‌نظمی‌ها در زبان آمیخته‌ای مثل فارسی، که به راحتی در ردهٔ فعل‌آغاز یا فعل‌پایان قرار نمی‌گیرد (رک بخش ۲)، حاصل تماس زبانی است. این زبان در طول سالیان با خانوادهٔ زبان‌های سامی، از جمله عربیو زبان‌های آرامی^۲، و با زبان ترکی همسایه بوده است. به

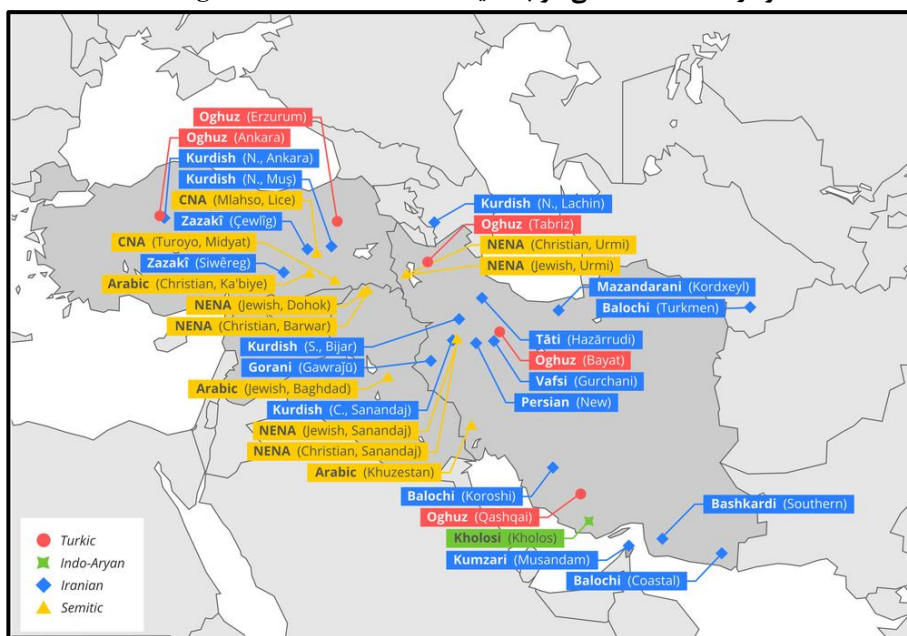
1. Csató, E. A.

2 Aramaic

همین دلیل برخی از بی‌نظمی‌های معرفی شده در این مقاله را با این دو خانواده زبانی مقایسه می‌کنیم، تا ببینیم شواهدی از تاثیر این زبان‌ها بر توالی کلمات در فارسی می‌توان یافت یا خیر؟

هیگ و همکاران (۲۰۲۴ الف و ب) از محدود مطالعاتی هستند که با مطالعه زبان‌ها در منطقه‌ای خاص سعی در معرفی تبیین‌های منطقه‌ای برای برخی ویژگی‌های این زبان‌ها در توالی کلمات کردند. آنها این منطقه را «منطقه انتقالی غرب آسیا (WATZ)» (هیگ، ۲۰۱۷؛ هیگ و خان، ۲۰۱۸) می‌نامند. نمودار (۲۱) این منطقه و برخی زبان‌های موجود در آن را نشان می‌دهد.

نمودار ۲۱. منطقه انتقالی غرب آسیا (WATZ) (Haig et al. 2024b)



منطقه غرب آسیا که در سال‌های اخیر به عنوان یک منطقه انتقالی^۱ (هیگ و خان، ۲۰۱۸) یا حائل^۲ (استیلو، ۲۰۰۵) از آن نام برده می‌شود، تعریف کاملاً دقیق جغرافیایی ندارد. همان‌گونه که در نمودار (۲۱) مشخص است این منطقه شامل بخش‌هایی از ایران، شرق ترکیه و شمال عراق است. در واقع این منطقه از بالای بین‌النهرین تا رشته کوه‌های زاگرس و البرز را دربرمی‌گیرد. منطقه‌ای که در طول تاریخ خاستگاه امپراتوری‌های مختلفی بوده است، از

1. transition zone
2. buffer zone

سومریان (۳۰۰۰ سال قبل از میلاد) تا امپراطوری عثمانی در اوائل قرن بیستم. به این ترتیب این منطقه همیشه محل دادوستد زبانی و فرهنگی بوده است، چنان‌که هیگ و خان (۲۰۱۸: ۱) از آن با عنوان فضای فرهنگی^۱ یاد می‌کنند. در این فضای فرهنگی زبان‌هایی از خانواده‌های زبانی مختلف با یکدیگر در تماس بوده‌اند و بر هم تاثیر گذاشته‌اند. زبان‌هایی از چهار خانواده ترکی، سامی، هندواروپایی و کارتولی^۲ در این منطقه وجود دارند.

مناطق انتقالی از حیث زبان‌شناختی، فقط مناطقی از زبان‌های مختلف نیستند، بلکه می‌توانند تصویر رده‌شناختی مخصوص به خود در برخی ویژگی‌ها داشته باشند و از دل این تماس زبان‌هایی به وجود آیند که ویژگی‌های آمیخته دارند. منطقه انتقالی غرب آسیا بین دو بخش مجزا قرار گرفته است. بخش اول گروهی از زبان‌ها که توالی فعل‌پایان و پس‌افزافه دارند و شامل ترکی و مغولی در شمال و شمال‌غرب ایران و زبان‌های ایرانی شرقی، دراویدی^۳ و هندوآریایی در شبه‌قاره هند هستند. در بخش دوم، زبان‌های فعل‌آغاز و پیش‌افزافه‌ای هستند که در شمال آفریقا، اروپای غربی و اطراف مدیترانه یافت می‌شوند، مانند خانواده زبانی سامی، رومیایی و یونانی. همسایگی آنها با یکدیگر و با زبان‌های خانواده ایرانی، باعث شده تغییراتی در این زبان‌ها به وجود آید. در ادامه به چند مورد از تغییرات حاصل تماس زبانی اشاره می‌شود. استیلو (۲۰۰۵ و ۲۰۰۹) از نخستین افرادی است که به مطالعه تماس زبانی در این منطقه پرداخته است. او با بررسی دو رده زبانی مقابل در این منطقه، رده زبانی فعل‌پایان (ترکی، هندی، قفقازی، دراویدی و جز آن) و فعل‌آغاز (عربی، آلبانیایی و جز آن)، عنوان می‌کند که زبان فارسی و بسیاری از زبان‌های ایرانی از این حیث رده آمیخته هستند، یعنی ویژگی‌هایی از هر دو رده را در خود دارند. در واقع این زبان‌ها در منطقه انتقالی بین دو رده زبانی مخالف به وجود می‌آیند. زبان‌های ایرانی در شرق و شمال با زبان‌های فعل‌پایان همسایه بوده‌اند و در غرب و جنوب با زبان‌های فعل‌آغاز. به عنوان یکی از متغیرهایی که متاثر از تماس زبانی در این منطقه است، استیلو (۲۰۰۵: ۴۹) به توالی صفت و اسم می‌پردازد. وی عنوان می‌کند در زبان‌هایی مانند گیلکی، مازندرانی، و تاتی که در شمال با ترکی و زبان‌های قفقازی همسایه‌اند، و همچنین زبان‌هایی مانند بلوچی، پشتو و پامیر در جنوب و شرق که همسایه زبان‌های هندی هستند، صفت قبل از اسم قرار می‌گیرد (الگوی فعل‌پایان)، اما در زبان‌هایی مانند فارسی، لری، کرمانجی، و نظایر آن که همسایه زبان‌های فعل‌آغاز (مانند عربی و آرامی نو) هستند، صفت

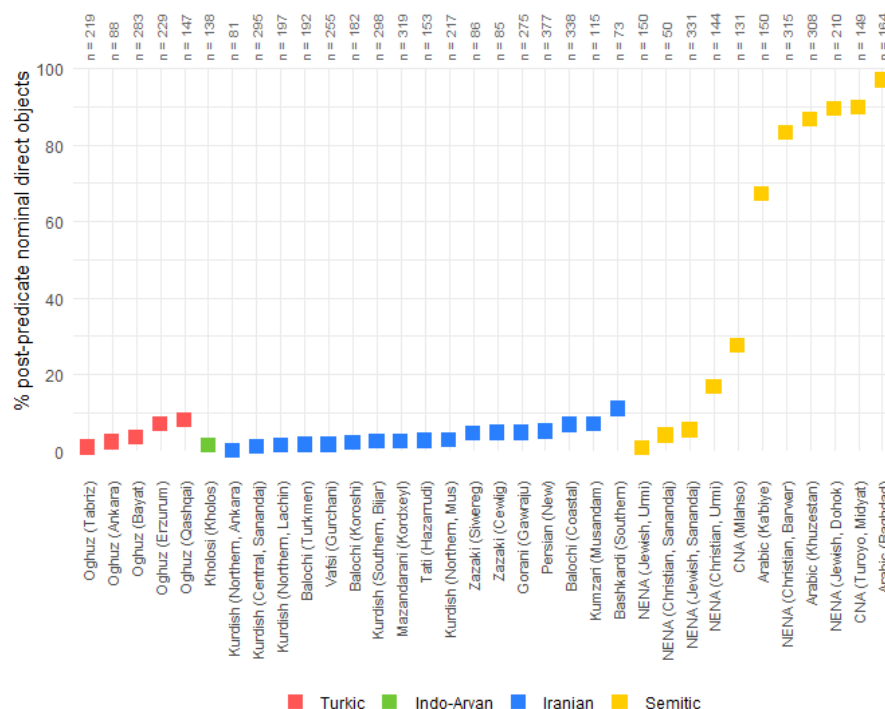
1. Kulturraum

2. Kartvelian

3. Dravidian

پس از فعل قرار می‌گیرد (الگوی فعل‌آغاز). وی همین مشاهده را در مور توالی اسم و وابسته اضافی دریافته است، و همین زبان‌ها کم‌وبیش چنین الگویی را نشان می‌دهند. پس از استیلو، هیگ و همکاران (۲۰۲۴) الف و ب و ۲۰۲۵) در مجموعه مقالات مختلفی به بررسی تاثیر جغرافیا بر توالی کلمات در این منطقه پرداختند. عمده این مطالعات بر اساس پیکره WOA انجام شده است. نمودار (۲۲) (هیگ و همکاران، ۲۰۲۵) به جایگاه مفعول مستقیم و فعل در زبان‌های این منطقه اشاره دارد.

نمودار ۲۲. بسامد مفعول‌های اسمی پس از فعل در ۳۵ گونه زبانی (Haig et al., 2025: fig.2)

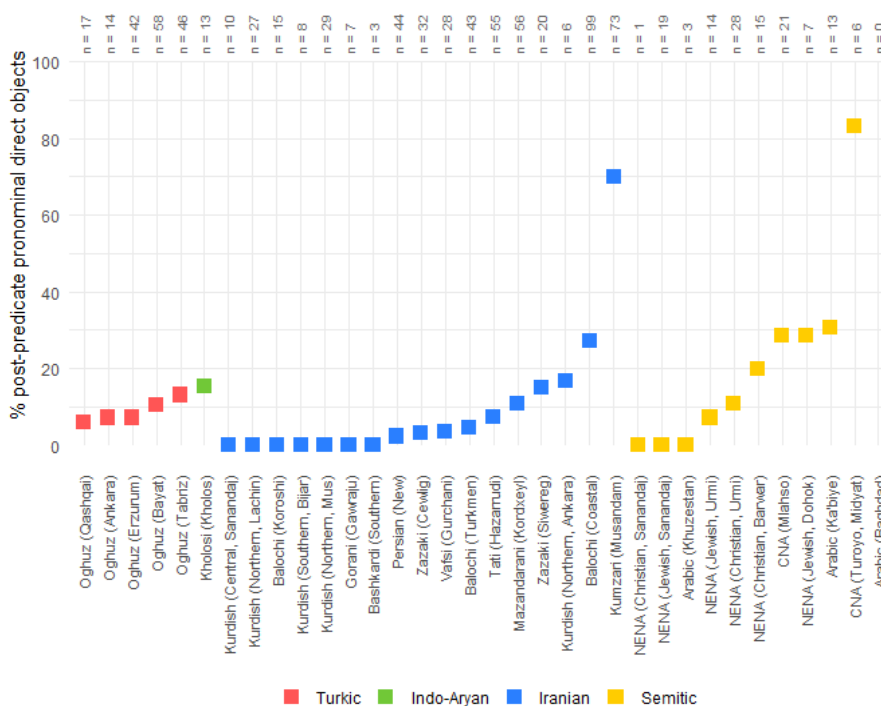


همان‌گونه که مشخص است زبان‌های ترکی و ایرانی (مربع‌های قرمز و آبی) از حیث تاریخی فعل‌پایان هستند و داده‌های پیکره‌ای نشان می‌دهد این ویژگی تقریباً ثابت است. به عبارتی مفعول قبل از فعل به‌ندرت به جایگاه پس از فعل می‌رود. این یافته برخی دیدگاه‌های پیشین را به چالش می‌کشد که عنوان می‌کنند توالی مفعول+فعل بی‌ثبات است و راحت‌تر به توالی فعل+مفعول تبدیل می‌شود. تقریباً در زبان‌هایی که از حیث تاریخی مفعول+فعل بوده‌اند، حتی در داده‌های گفتاری این ویژگی حفظ شده است (به جز بشاکردی که در حدود سی درصد

موارد مفعول به بعد از فعل رفته است). اما در مورد زبان‌های سامی این منطقه، تصویر متفاوت است. با اینکه زبان‌های سامی از حیث تاریخی فعل+مفعول هستند، در برخی گونه‌ها، جایگاه مفعول به قبل از فعل تغییر کرده است. این گونه‌ها در مناطقی صحبت می‌شوند که توسط زبان‌های ایرانی محصور شده‌اند (مانند زبان‌های آرامی نو شمال شرقی^۱ در سنج و ارومیه. به عبارتی، تماس زبانی باعث شده این ویژگی در زبان‌های سامی شبیه به زبان‌های ایرانی و ترکی منطقه شود.

وقتی نمودار (۲۲) را با نمودار (۲۳) مقایسه می‌کنیم، متوجه می‌شویم که در اکثر زبان‌های ایرانی و ترکی که مفعول اسمی جایگاه ثابت پیش از فعل داشت، جایگاه مفعول ضمیری نیز ثابت است.

نمودار ۲۳. بسامد مفعول‌های ضمیری پس از فعل در ۳۵ گونهٔ زبانی (Haig et al., 2025: fig.3)



اما این نمودار دو تفاوت با نمودار قبلی دارد. یکی اینکه اکثر زبان‌های سامی، حتی مواردی که مفعول اسمی را پس از فعل قرار می‌دادند، مفعول ضمیری را قبل از فعل قرار می‌دهند.

¹ North-Eastern Neo-Aramaic (NENA)

یعنی رفتار این زبان‌ها در مورد مفعول‌های ضمیری به زبان‌های فعل‌پایان شبیه شده است. به عبارتی احتمال حضور مفعول ضمیری قبل از فعل در این زبان‌ها بیش از مفعول اسمی است. نکته جالب دیگر در این نمودار زبان کمزاری است. کمزاری زبانی ایرانی است که در عمان صحبت می‌شود. محصور شدن این زبان با زبان‌های سامی باعث شده ویژگی‌های جالبی از دو رده فعل‌آغاز و فعل‌پایان داشته باشد. با اینکه مفعول اسمی در این زبان در همان جایگاه تاریخی قبل از فعل قرار می‌گیرد، مفعول ضمیری بیشتر به جایگاه پس از فعل رفته و متاثر از تماس با زبان‌های سامی اطراف خود است (آنونبی^۱، ۲۰۱۵).

ویژگی دیگری که در منطقه انتقالی غرب آسیا دیده می‌شود و به قول هیگ و خان (۲۰۱۸: ۵) راهکاری برای آشتی زبان‌های فعل‌آغاز و فعل‌پایان در این منطقه است، وجود سازه پس از فعل است. در بخش (۴-۶) درباره این ویژگی در فارسی صحبت کردیم، و در این بخش تبیین جغرافیایی این ویژگی را در فارسی و سایر زبان‌های این منطقه مرور می‌کنیم. در واقع نکته مهم در این متغیر، بررسی جایگاه هدف نسبت به فعل یا گزاره است که می‌تواند هدف+فعل^۲ یا فعل+هدف^۳ باشد. در یک جمله هرچقدر به سمت غرب این منطقه پیش می‌رویم، میزان قرار گرفتن هدف پس از فعل بیشتر می‌شود (هیگ و همکاران، ۲۰۲۴ ب) (نمودار ۲۴). مثلاً زبان‌های ایرانی که در تماس بلندمدت با زبان‌های سامی، در بخش جنوبی و غربی، قرار دارند، تقریباً در حدود صد در صد هدف‌ها پس از فعل قرار می‌گیرند. اما زبان‌هایی که در شرق و شمال ایران قرار دارند و به زبان‌های فعل‌پایان نزدیک‌ترند، میزان حضور سازه هدف پس از فعل در آنها کمتر است. نمودار (۲۴) این دو گروه را به طور مشخص نشان داده است.

همان‌گونه که در نمودار (۲۴) مشخص است، حتی در زبانی مانند ترکی قشقایی که تحت تاثیر زبان‌های ایرانی غربی است، میزان زیادی (۶۰ درصد) از سازه‌های هدف پس از فعل قرار می‌گیرند. به عبارتی، با این که زبان‌های ترکی و ایرانی فاعل+مفعول+فعل هستند، اما در این منطقه، درصد بالایی از سازه هدف را پس از فعل قرار می‌دهند. اما در مورد زبان‌های فعل‌پایان که در قسمت شمال و شرق منطقه صحبت می‌شوند، این میزان کم می‌شود (مانند بلوچی ترکمنستان یا مازندرانی). وقتی نمودارهای (۲۲) و (۲۳) را با نمودار (۲۵) مقایسه می‌کنیم، متوجه می‌شویم که اکثر زبان‌های ایرانی و ترکی سازه هدف را به جایگاه پس از فعل برده‌اند و از این حیث شبیه زبان‌های سامی منطقه شده‌اند.

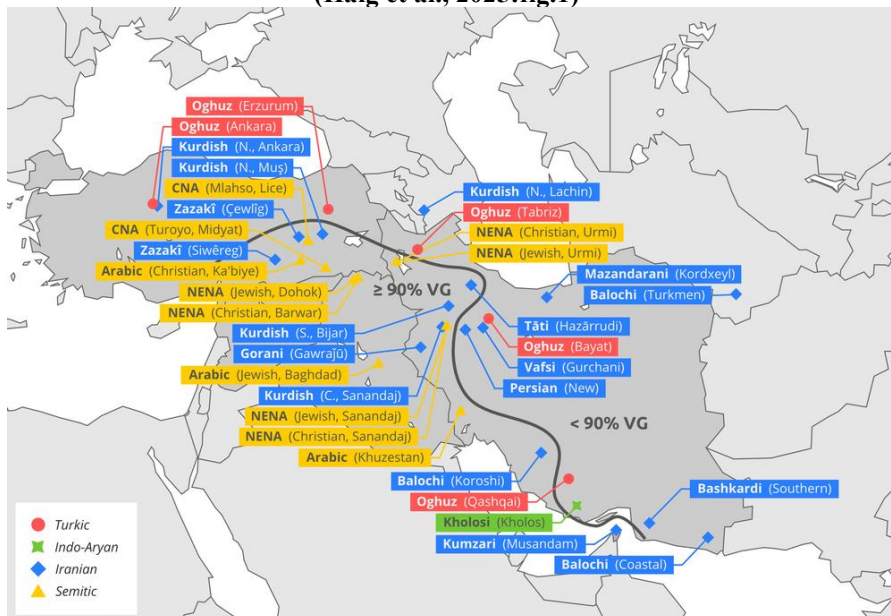
1. Anonby, C.

2. GV

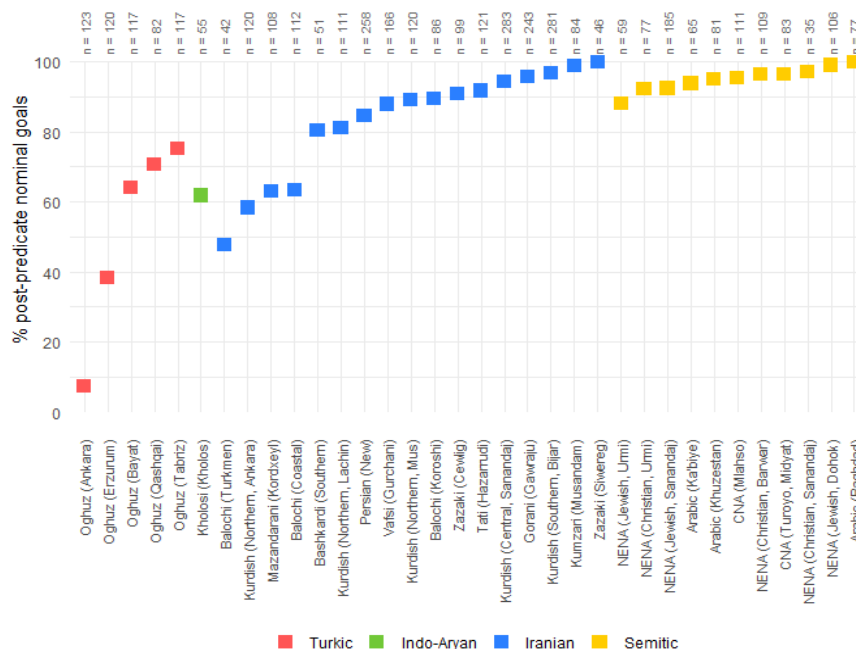
3. VG

نمودار ۲۴. نمای زبان‌هایی که سازه هدف را پس از فعل قرار می‌دهند (بیش و کمتر از ۹۰ درصد)

(Haig et al., 2025:fig.1)



نمودار ۲۵. بسامد سازه هدف پس از فعل در ۳۵ گونه زبانی (Haig et al., 2025: fig.4)



همان‌گونه که مشخص است حتی در زبان‌های ترکی هم سازه هدف به جایگاه پس از فعل رفته است، اما این اتفاق در گونه‌هایی روی داده است که نزدیک زبان‌های ایرانی هستند مانند اویغوز بیات، اویغوز تبریز و قشقایی. این ویژگی در ترکی آنکارا دیده نمی‌شود و در ترکی ارزروم هم کمتر است. نکته دیگر اینکه حتی زبان‌های سامی که در مورد جایگاه مفعول به زبان‌های فعل‌پایان نزدیک شده بودند (نمودار ۲۲)، در مورد جایگاه هدف مانند سایر زبان‌های منطقه عمل می‌کنند و آن را پس از فعل قرار می‌دهند. پس می‌توان گفت زبان‌هایی که از حیث تاریخی فعل+هدف بوده‌اند، ثبات بیشتری داشته‌اند اما زبان‌هایی که از حیث تاریخی هدف+فعل بوده‌اند، ثبات کمتری داشته و تغییر کرده‌اند.

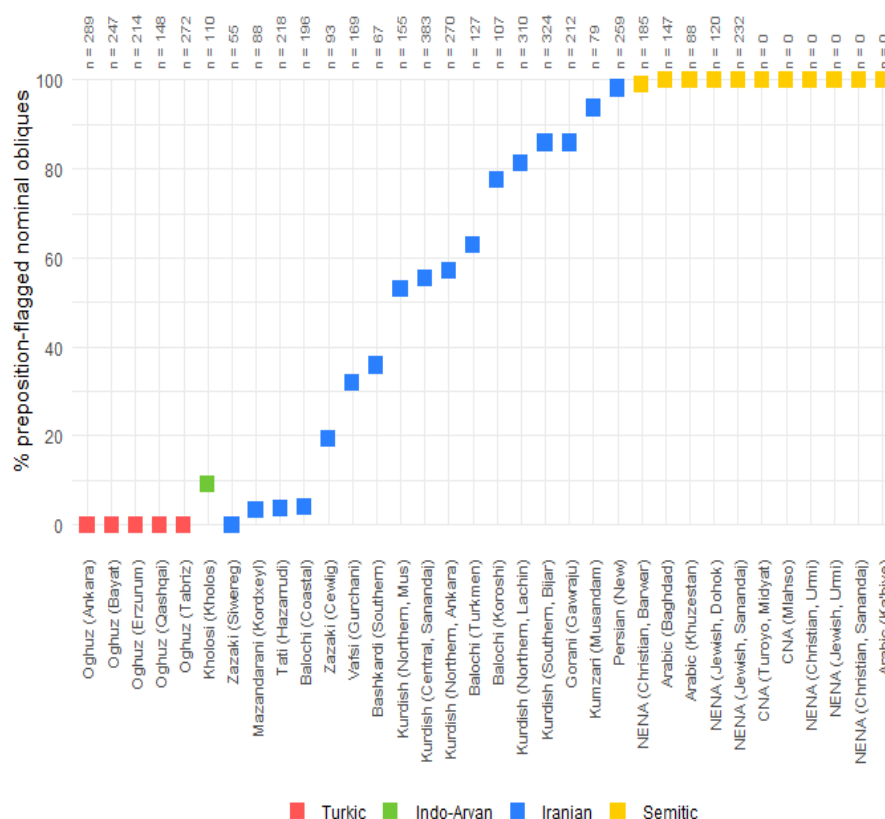
نکته جالب دیگر در این منطقه به توالی حرف اضافه^۱ و اسم برمی‌گردد. استیلو (۲۰۰۵): ۵۵ و ۲۰۱۲ نشان داده است که زبان‌های ایرانی بین زبان‌هایی که پیش‌اضافه‌ای هستند (یعنی عربی و سایر زبان‌های سامی در جنوب و جنوب غربی ایران غربی) و زبان‌های که پس‌اضافه‌ای هستند (یعنی ترکی و قفقازی در شمال و شمال غربی و زبان هندی در شرق) قرار دارند. به همین دلیل زبان‌های ایرانی ویژگی‌های آمیخته‌ای از این دو گروه دارند. زبان‌هایی در این منطقه یافت می‌شوند که هم‌زمان پس‌اضافه، پیش‌اضافه و پیرااضافه دارند (مانند وفسی)، در برخی زبان‌ها پس‌اضافه غالب است (مانند مازنی و گیلکی) و در برخی پیش‌اضافه غالب است (مانند فارسی و لری). هیگ و همکاران (۲۰۲۵) این ویژگی را به صورت پیکره‌ای بررسی کردند که نتایج آن در نمودار (۲۶) منعکس شده است. نمودار (۲۶) نشان می‌دهد زبان‌های ترکی پس‌اضافه‌ای و زبان‌های سامی پیش‌اضافه‌ای هستند، اما زبان‌های ایرانی در این بین آمیخته هستند. آن زبان‌هایی که با زبان‌های سامی از حیث تاریخی تماس داشته‌اند، مانند کمزاری و فارسی بیشتر پیش‌اضافه دارند، اما زبان‌های که به گونه‌های ترکی نزدیک بوده‌اند، پس‌اضافه دارند مانند زازاکی و مازندرانی و تاتی. و البته زبان‌های بسیاری هم هستند که هم پیش‌اضافه و هم پس‌اضافه دارند، مانند وفسی و برخی گونه‌های کردی و بلوچی.

زبان‌شناسان مولفه‌های دیگری را نیز از حیث تماس زبانی در این منطقه بررسی کرده‌اند. مثلاً هیگ و خان (۲۰۱۸) نشان داده‌اند که وجود اجباری فعل ربطی در انتهای جملات می‌تواند حاصل تماس زبانی باشد که در سراسر منطقه گسترش پیدا کرده است. در واقع تمام زبان‌های این منطقه، فارغ از خانواده زبانی، این ویژگی را دارند. همچنین محمدی راد (۲۰۲۰) مالکیت اسنادی در زبان‌های ایرانی غربی بررسی کرده است و نشان داده زبان‌های نزدیک به ترکی و عربی از فعل «بودن» و زبان‌های مناطق مرکزی از فعل «داشتن» برای بیان این رابطه استفاده می‌کنند. یافته‌های ارائه شده در این بخش نشان می‌دهد در کنار سایر تبیین‌های نقشی، تماس

1 adposition

زبانی و تبیین‌های منطقه‌ای می‌تواند تصویر کامل‌تری از آنچه در توالی کلمات زبان فارسی و سایر زبان‌های ایرانی دیده می‌شود، نشان دهد.

نمودار ۲۶. بسامد پیش‌اضافه و پس‌اضافه (Haig et al., 2025: fig. 6)



۶. نتیجه‌گیری

با افزایش روزافزون پیکره‌های زبانی مختلف، این امکان برای زبان‌شناسان به وجود آمده است که به داده‌های طبیعی دسترسی پیدا کنند و حجم بالای داده‌ها باعث شده آنها بتوانند تحلیل‌های کمی از ویژگی‌های مختلف زبان‌ها به دست دهند. رده‌شناسان نیز از این قافله عقب نمانده‌اند و پیکره‌های مختلف از زبان‌های گوناگون به آنها کمک کرده شباهت‌ها و تفاوت‌های زبان‌ها را با دید بهتری بررسی کنند.

در این مقاله به مرور مطالعاتی پرداختیم که با استفاده از پیکره‌های زبانی به بررسی توالی کلمات در زبان‌های مختلف اختصاص دارند. این مطالعات تصویر درست‌تری از نظم و بی‌نظمی‌های توالی کلمات در زبان فارسی، در پرتو رده‌شناسی پیکره‌بنیاد به دست می‌دهند. به طور کلی می‌توان عنوان کرد وجود بی‌نظمی در توالی کلمات بین و درون‌زبانی امری عادی است و تلاش برای نگاه مقوله‌ای به رده‌های مختلف زبانی، امری تصنعی است. داده‌های پیکره‌ای به ما نشان داد زبان فارسی در چه متغیرهایی بی‌نظمی بالایی دارد (مانند توالی هسته و وابسته یا سازه‌های پس از فعل) و در چه متغیرهایی بی‌نظمی کمتری دارد (مانند توالی فاعل و مفعول، صفت پس از اسم و داشتن پیش‌افزافه). همچنین تحلیل منطقه‌ای این زبان و سایر زبان‌های ایرانی نشان داد چرا این زبان‌ها، با وجود قرار دادن فعل در انتهای جمله، تمام ویژگی‌های رده زبانی فعل‌پایان را ندارند. با این که قبلاً به این ویژگی‌ها کم‌وبیش در رده‌شناسی دستور بنیاد اشاره شده بود، اما تبیین منطقه‌ای نشان می‌دهد دلیل آمیخته بودن این زبان‌ها، لزوماً تاریخی نیست و به معنی داشتن توالی آزاد هم نیست. این زبان‌ها، به دلیل حضور درازمدت در منطقه انتقالی غرب آسیا، متأثر از خانواده‌های زبانی مختلف با ویژگی‌های مختلف هستند، و همین امر باعث شده از دو رده مختلف فعل‌آغاز و فعل‌پایان، ویژگی‌های مختلفی در خود داشته باشند. با ایجاد پیکره‌های بیشتر از زبان فارسی و سایر زبان‌های ایرانی، طبق معیارهای جهانی، می‌توان از ویژگی‌های مختلف این زبان‌ها تصویر درستی به دست آورد.

منابع

- دبیرمقدم، محمد (۱۳۹۲). رده‌شناسی زبان‌های ایرانی، دو جلد. تهران: سمت.
- راسخ‌مهند، محمد (۱۳۹۶). نحو زبان فارسی: نگاهی نقشی-رده‌شناختی. تهران: آگه.
- راسخ‌مهند، محمد (۱۴۰۴). چرا زبان فارسی فعل ساده نمی‌سازد؟ پژوهش‌های زبان‌شناسی تطبیقی، ش ۲۹، ص ۳۸-۶.
- نغزگوی کهن، مهرداد (۱۳۹۲). بررسی تاثیر زبان عربی بر نظام واجی زبان فارسی و گونه‌های آن. پژوهش‌های زبان‌شناسی تطبیقی، ش ۵، ص ۶۵-۹۴.
- Anonby, C. V. D. W. (2015). *A Grammar of Kumzari: A mixed Perso-Arabian language of Oman*. Leiden: Leiden University PhD Dissertation.
- Croft, W. (2003). *Typology and universals*, 2nd ed. Cambridge: Cambridge University Press.
- Csató, É. Á, Isaksson, B. & Jahani, C. (eds.) (2005). *Linguistic Convergence and Areal Diffusion: Case studies from Iranian, Semitic and Turkic*. London, New York: Routledge.
- Dabir Moghaddam, M. (2018). Typological approaches and dialects. In Anousha Sedighi & Pounesh Shabani-Jadidi (eds.), *The Oxford handbook of Persian linguistics*, 52-88. Oxford: Oxford University Press.

- Dryer, M. S. (1992). The Greenbergian Word Order Correlations. *Language* 68: 81-138.
- Dryer, M. S. (1997). On the six-way word order typology. *Studies in Language*, 21(2), pp. 69–103.
- Dryer, M. S. (2013). Order of Subject, Object and Verb. In Dryer, M. S. & Haspelmath, M. (Eds.), *WALS Online* (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591>. Retrieved from <http://wals.info/chapter/81>.
- Dryer, M. S., & Haspelmath, M. (Eds.) (2013). *WALS Online* (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591>. Retrieved from <https://wals.info>.
- Ebert, C., Bickel, B., & Widmer, P. (2024). "Areal and phylogenetic dimensions of word order variation in Indo-European languages" *Linguistics*, vol. 62, no. 5, pp. 1085-1116. <https://doi.org/10.1515/ling-2022-0146>
- Frommer, P. (1981). Post-verbal phenomena in colloquial Persian syntax. Los Angeles: University of Southern California. (Doctoral dissertation).
- Futrell, R., Mahowald, K., Edward Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *PNAS*, 112(33), pp. 10336–10341. 22
- Gerdes, K., Sylvain K., & Xinying, C. (2021). Typometrics: From Implicational to Quantitative Universals in Word Order Typology. *Glossa: a journal of general linguistics* 6(1): 17. 1–31. DOI: <https://doi.org/10.5334/gjgl.764>
- Goldhahn, D., Eckart, Th., & Quasthoff, U. (2012). Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In: Calzolari, N., Choukri, Kh., Declerck, Th. et al. (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pp. 759–765. Istanbul: ELRA.
- Greenberg, J. (1963/1966). Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. *Universals of Language*, London: MIT Press, pp. 110-113.
- Haig, G. (2017). Western Asia: East Anatolia as a transition zone. In Hickey, Raymond (ed.), *The Cambridge Handbook of Areal Linguistics*, 396-423. Cambridge: CUP.
- Haig, G., & Khan, G. (2018). Introduction. In: Haig, Geoffrey & Geoffrey Khan (eds.) *The languages and linguistics of Western Asia. An areal perspective*, 1-29. Berlin: Mouton de Gruyter.
- Haig, G., & Rasekh-Mahand, M. (2019). Post-predicate constituents in Iranian and neighboring languages. Position paper for the project Post-predicate constituents in Iranian and neighboring languages. https://www.uni-bamberg.de/fileadmin/aspra/05_Events/2019_Post
- Haig, G., & Rasekh-Mahand, M. (eds.). (2022). *HamBam: The Hamedan-Bamberg Corpus of Contemporary Spoken Persian*. Bamberg: University of Bamberg. (multicast.aspra.uni-bamberg.de/resources/hambam/).

- Haig, G., & Schnell, S. (2014). Annotations using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and guidelines for annotators. Version 7.0. <https://multicast.aspra.uni-bamberg.de/>
- Haig, G., & Schnell, S. (2015). Multi-CAST: Multilingual Corpus of Annotated Spoken Texts. Bamberg: University of Bamberg. <https://multicast.aspra.uni-bamberg.de/>
- Haig, G., & Schnell, S. (eds.). (2023). Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/
- Haig, G., Schnell, S., & Seifart, F. (eds.). (2022). *Doing Corpus-Based Typology With Spoken Language Corpora: State of the Art*. Language Documentation & Conservation Special Publication 25. Honolulu: University of Hawai'i Press. <https://doi.org/10.5167/uzh-217096>
- Haig, G., Schnell, S. & Schiborr, N. N. (2021). Universals of reference in discourse and grammar: Evidence from the Multi-CAST collection of spoken corpora. In G. Haig, S. Schnell & F. Seifart (eds.), *Doing Corpus-Based Typology with Spoken Language Corpora*. Language Documentation & Conservation Special Publication 25. Honolulu: University of Hawai'i Press.
- Haig, G., Vollmer, M., & Thiele, H. (2019). Multi-CAST Northern Kurdish. In G. Haig & S. Schnell (eds.), Multi-CAST: Multilingual Corpus of Annotated Spoken Texts (Version 2311). Bamberg: University of Bamberg. multicast.aspra.uni-bamberg.de/#nkurd
- Haig, G., Stilo, D., Schiborr, N. N., & Dogan, M. (eds.). (2022). WOVA: Word Order in Western Asia: A spoken-language-based corpus for investigating areal effects in word order variation. Bamberg: University of Bamberg. (multicast.aspra.uni-bamberg.de/resources/wowa/).
- Haig, G., Rasekh-Mahand, M., Stilo, D., Schreiber, L. & Schiborr, N. N. (eds.). (2024a). Post-predicate elements in the Western Asian Transition Zone: A corpus-based approach to areal typology (Contact and Multilingualism 8). Berlin: Language Science Press. <https://langsci-press.org/catalog/book/395>
- Haig, G., Rasekh-Mahand, M., Stilo, D., Schreiber, L. & Schiborr, N. N. (eds.). (2024b). Post-predicate elements in the Western Asian Transition Zone: Data, theory, and methods. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr (eds.). 2024. Post-predicate elements in the Western Asian Transition Zone: a corpus-based approach to areal typology. Berlin: Language Science Press. 1-48. <https://langsci-press.org/catalog/book/395>
- Haig, G., Noorlander, P., & Schiborr, N. N. (2025). "10. Which word order features are stable in a contact setting? Corpus-based evidence from the Western Asian Transition Zone". *Language contact, an international handbook, Volume 2*, edited by Jeroen Darquennes, Joseph C. Salmons and Wim Vandenbussche, Berlin, Boston: De Gruyter Mouton. 159-184. <https://doi.org/10.1515/9783110443011-010>

- Ivani, J., & Bickel, B. (In press). Databases for comparative syntactic research. To appear in *The Cambridge Handbook of Comparative Syntax*, ed. Barbiers, S., N. Corver & M. Polinsky
- Jing, Y., Blasi, D. E., & Bickel, B. (2022). Dependency-length minimization and its limits: A possible role for a probabilistic version of the final-over-final condition. *Language*, 98(3), pp. 397–418. DOI <https://doi.org/10.1353/lan.0.0267>.
- Kiparsky, P. (1997). The rise of positional licensing. In Ans von Kemenade and Nigel Vincent, editors, *Parameters of morphosyntactic change*, pages 460–494. Cambridge University Press.
- Levshina, N. (2015). European analytic causatives as a comparative concept. Evidence from a parallel corpus of film subtitles. *Folia Linguistica* 49(2). 487–520.
- Levshina, N. (2019). Token-based typology and word order entropy. *Linguistic Typology*, 23(3), pp. 533–572. DOI <https://doi.org/10.1515/lingty-2019-0025>.
- Levshina N (2021). Cross-Linguistic Trade-Offs and Causal Relationships Between Cues to Grammatical Subject and Object, and the Problem of Efficiency-Related Explanations. *Front. Psychol.* 12:648200. doi: 10.3389/fpsyg.2021.648200
- Levshina, N. (2022). Corpus-based typology: Applications, challenges and some solutions. *Linguistic Typology* 26(1). 129–160.
- Levshina, N. (2022). *Communicative Efficiency: Language Structure and Use*. Cambridge: Cambridge University Press.
- Levshina, N., Nambodiripad, S., Allasonnière-Tang, M., Kramer, M., Talamo, L., Verkerk, A., Wilmoth, S., Garrido Rodriguez, G., Gupton, T.M., Kidd, E., Liu, Z., Naccarato, Ch., Nordlinger, R., Panova, A., & Stoyanova, N. (2023). Why we need a gradient approach to word order. *Linguistics*, 61(4), pp. 825–883. <https://doi.org/10.1515/ling-2021-0098>.
- Mohammadirad, M. (2020). Predicative possession across Western Iranian languages, *Folia Linguistica*, 54(3): 497-526. <https://doi.org/10.1515/flin-2020-2038>
- Parizadeh, M., & Rasekh-Mahand, M. (2024). Post-predicate elements in Early New Persian. In Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr (eds.). *Post-predicate elements in the Western Asian Transition Zone: a corpus-based approach to areal typology*. Berlin: Language Science Press. 184-196. <https://langsci-press.org/catalog/book/395>
- Rasekh-Mahand, M., & Parizadeh, M. (2024). Different functions of ‘rā’ in New Persian A semantic map analysis. *Journal of Historical Linguistics*, 14(1): 31 – 57. <https://doi.org/10.1075/jhl.21056.ras>
- Rasekh-Mahand, M., Izadi, E., Parizadeh, M., Haig, G., & Schiborr, N. N. (2024). Post-predicate elements in colloquial Persian; a multi-variate analysis; in Geoffrey Haig, Mohammad Rasekh-Mahand, Donald Stilo, Laurentia Schreiber & Nils Schiborr (eds.). 2024. *Post-predicate elements in the Western Asian Transition Zone: a corpus-based approach to areal typology*. Berlin: Language Science Press. 151-182. <https://langsci-press.org/catalog/book/395>

- Schnell, S., & Schiborr, N. N. (2022). Crosslinguistic Corpus Studies in Linguistic Typology. *Annual Review Linguistics*. 8:171-191. <https://doi.org/10.1146/annurev-linguistics-031120-104629>
- Spirgath, A. (2025). The interaction of word order entropy and verb agreement: A token-based approach. *Linguistics in Amsterdam* 16 (1): 30-50.
- Stilo, D. (1994). Phonological Systems in Contact in Iran and Transcaucasia, [in] Mehdi, Marashi, Ed., *Persian Studies in North America: Studies in Honor of Mohammad Ali Jazayery*, Bethesda, Maryland: Iranbooks, 1994.
- Stilo, D. (2005). Iranian as buffer zone between the universal typologies of Turkic and Semitic. In Csató, Éva Ágnes, Isaksson, Bo & Jahani, Carina (eds.), *Linguistic Convergence and Areal Diffusion: Case studies from Iranian, Semitic and Turkic*, 35-63. London, New York: Routledge.
- Stilo, D. (2009a). Circumpositions as an areal response: The case study of the Iranian zone. *Turkic Languages* 13. 3-33.
- Stilo, D. (2009b). 'Case In Iranian: From Reduction and Loss to Innovation and Renewal', in Andrej L. Malchukov, and Andrew Spencer (eds), *The Oxford Handbook Case*. <https://doi.org/10.1093/oxfordhb/9780199206476.013.0049>, accessed 22 Dec. 2025.
- Stilo, D. (2012). Intersection zones, overlapping isoglosses, and 'Fade-out/Fade-in' phenomena in Central Iran. In: Behrad Aghaei & M. R. Ghanoonparvar (eds.), *Iranian Languages and Culture: Essays in Honor of Gernot Ludwig Windfuhr*, 3-33. Costa Mesa: Mazda Publishers.
- Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, 2214–2218. Istanbul: European Language Resources Association (ELRA).
- Vatani, T., Väyrynen, J. J., & Virpioja, S. (2010). Language identification of short text segments with n-gram models. In *Proceedings of the seventh international conference on language resources and evaluation (LREC'10)*, 3423–3430. Malta: European Language Resources Association (ELRA).
- Verkerk, A. (2014). The evolutionary dynamics of motion event encoding. PhD dissertation. Radboud University Nijmegen.
- Zeman, D., et al. (2024). Universal Dependencies 2.15, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. Retrieved from <http://hdl.handle.net/11234/1-5787>.