

Introduction

Using the computational tool developed by Stump and Finkel (2013) for measuring the complexity of inflection-class systems, we examine the complexity of Persian verb inflection. We regard the complexity of an inflectional system as the extent to which its paradigms are unpredictable according to a range of measurements. To facilitate these measurements, we represent the Persian conjugational system as a “plat”. We also make separate plats for simple and complex predicates, since we believe that these two types of verbs do not show the same level of complexity due to the differences in their inflection classes (ICs). In the following sections we will discuss the different level of complexity in Persian simple and complex predicates, and also we will compare the complexity of Persian verb system with those of Latin and French.

Persian Verb System

Persian Inflectional Morphology

- Poor nominal inflection (no gender, no case, etc.)
- Rich verbal morphology (modal/aspectual prefixes, different verbal stems, different personal endings, etc.)

The main tool for word formation in Persian is composition. So, in order to expand the verbal lexicon, the language uses compounding.

Types of Verbs in Persian	
✓ Simple	eg. aamadan: to come
✓ Complex	+ Compound Predicates
	- Compounding
	eg. aab shodan (water become): to melt
	- Incorporation
	eg. tasmim gereftan (decision take): to decide
	+ Impersonal Constructions
	eg. xosh aamadan (delight come): to be pleased/delighted

Although the conjugation of simple verbs is largely predictable and regular in Persian, the conjugation of complex predicates (compound verbs) is comparatively complex. Ordinarily, a complex predicate’s light verb is conjugated, while its nonverbal part remains uninflected. But in impersonal constructions, it is the nonverbal part that is inflected for person and number (by means of pronominal enclitics), while the verbal part invariably exhibits third-person singular morphology, inflecting only with respect to tense and aspect. The following table is a partial fragment of the plat of Persian verbs. You can see these different inflections in different and random paradigms.

In this study we focus on Persian formal or written verb forms from a speaker-oriented point. We investigate the same number (i.e. 70) of simple and complex verbs. Moreover, the group of complex verbs includes an equal number of compound verbs and impersonal constructions.

Fragment of a plat of Persian verbs						
Lexeme	Inf	IndActPast SimpSg1	IndActPast ImpfPl2	IndActPres ProgPl3	IndActFut SimpPl1	SubjActPres SimpSg2
say	goftan	goft am	migoft id	dar and migu ya nd	xaah im goft	begu yi
be	budan	bud am	mibu di d	!	xaah im bud	baash i
like	xosh aamada n	xosh am aamad	xosh eta an miaamad	daar ad xosh es ha an miaa ya d	xosh ema an xaah ad aamad	xosh at biaa ya d
disappea r	qeyb zadan	qeyb am zad	qeyb eta an miza d	daar ad qeyb es ha an miza n ad	qeyb ema an xaah ad zad	qeyb at beza n ad
dream	xaab didan	xaab did am	xaab midid id	daar and xaab mibi n and	xaab xaah im did	xaab bebini
play	baazi kardan	baazi kard am	baazi mi kard id	daar and baazi miko n and	baazi xaah im kard	baazi (bo)koni

Assumptions About Complexity

What is Complexity?

The notion of linguistic complexity, according to Stump and Finkel (2013), specifically relates to inflection-class (IC) systems. By definition, an inflection class is a class of lexemes such that its members are distinguished by a common pattern of inflection. So, in this paradigm-based view of complexity “we compare the individual inflection classes in a language’s inflectional system according to the manner and extent of their deviation from canonical ideals of maximal transparency and maximal opacity; the criteria for these comparisons in turn inform a more general comparison of whole inflection-class systems. We define the complexity of an inflection-class system as the extent to which it inhibits motivated (“certain”) inferences about a lexeme’s full paradigm from subsets of the forms in its paradigm.” The way to represent the paradigms of an IC system is to use a plat of that system. By definition, **plat** is a table in which

- morphosyntactic property sets are represented as columns,
- conjugation classes are represented as rows, and
- the exponence (morphological expression) of a property set A in a particular conjugation B is represented in column A of row B.

Furthermore, we define a set of **principal-parts** for a given lexeme as any set of cells from which one can deduce the realization of the remaining cells of a paradigm. Those principal-part sets which are uniformly predictive of the whole paradigms are called **static** principal parts; while sets that only adhere uniquely to certain paradigms are called **dynamic** principal-part sets.

Applying the Measurements

The following tables show the results of applying several measures to Persian simple and complex verbs’ plats. The statistics employed here have been generated by means of the Principal-Parts Analyzer, a computer program that performs various calculations based on an input plat, and is freely accessible at:

<http://www.cs.uky.edu/~raphael/linguistics/analyze.html>

Our hypothesis was that in view of the diversity of light verbs and the different manners of conjugating complex predicates, the formal complexity of complex predicates will exceed that of simple verbs. As the left table below shows, complex predicates are partly more complex than simple verbs, but this is not an absolute higher rate of complexity. On the contrary, we can clearly see an absolute higher complexity in Persian compound verbs in comparison with impersonal constructions, and it is worth mentioning that in Persian compounding is the common mechanism of adding to the inventory of verbs (the highlighted cells show the higher complexity).

The relative complexity of Persian verbs			The relative complexity of the types of complex verbs		
	Simple verbs	Complex verbs		Compounds	Impersonal Constructions
Number of distillation	18	28	Number of distillation	16	14
Static principal-part number	6	8	Static principal-part number	7	3
Density of the system’s optimal static principal-part sets	0.000	0.000	Density of the system’s optimal static principal-part sets	0.000	0.033
Ave. dynamic principal-part number	1.91	1.80	Ave. dynamic principal-part number	1.88	1.50
Ave. ratio of actual to possible dynamic principal part sets	8.73%	6.92%	Ave. ratio of actual to possible dynamic principal part sets	5.67%	18.96%
Cell predictor number	1.03	1.06	Cell predictor number	1.04	0.96
Ave. predictiveness	0.251	0.441	Ave. predictiveness	0.269	0.408
IC predictability	0.273	0.250	IC predictability	0.251	0.538
Ave. predictability	0.620	0.711	Ave. predictability	0.614	0.807
4-MPS entropy	37	26	4-MPS entropy	35	19

Interpreting the Measurements

Different inflectional systems vary in the degree and the causes of complexity, so we need to employ a variety of different measurements to have a complete picture of the systems we are dealing with. Here are the 10 ways of measuring different aspects of an IC system’s morphological complexity as proposed by Stump and Finkel (2013):

- The more distillations an IC system has, the more complex it is. (A distillation is a set of morphosyntactic property sets whose exponence is interpredictable across all ICs)
- The larger the size of an IC system’s optimal static principal parts, the more complex it is.
- The lower the density of an IC system’s optimal static principal-part sets, the more complex it is. (it is the ratio of actual to possible optimal static principal-part sets, given the system’s number of distillations and the size of its optimal static principal-part sets)
- The larger the size of an IC system’s optimal dynamic principal-part sets, the more complex it is.
- The smaller the average ratio of actual to possible optimal dynamic principal-part analyses for an IC system, the more complex it is.
- The higher an IC system’s cell predictor number (average across IC’s), the more complex it is. (it is the average of dynamic principal-parts needed for each cell across the whole IC)
- The lower an IC system’s average cell predictiveness, the more complex it is. (it is the fraction of other cells in the paradigm that are fully determined by a given cell)
- The lower an IC system’s average IC predictability, the more complex it is. (it is the fraction of adequate dynamic principal-part sets in a cell’s realized paradigm)
- The lower an IC system’s average cell predictability, the more complex it is.
- The higher an IC system’s average n-MPS entropy, the more complex it is. (where n-MPS is the required number of morphosyntactic property sets for calculating the entropy)

Conclusion

But how complex is the verb system of Persian? In order to answer this question in this proposed framework, we are going to compare the complexity of Persian verbs to those of French and Latin. We were amazed by the high number of distillations Persian paradigm showed (35), which exceeded the distillations of previously analyzed languages. As highlighted in the table below, Persian shows a higher complexity by any of the ten given scales of measurement.

The relative complexity of Persian, French and Latin verb systems			
	Persian	French	Latin
Number of distillation	35	17	13
Static principal-part number	10	5	4
Density of the system’s optimal static principal-part sets	0.000	0.032	0.034
Ave. dynamic principal-part number	1.87	1.25	1.74
Ave. ratio of actual to possible dynamic principal part sets	5.94%	42.20%	19.55%
Cell predictor number	1.06	1.05	1.04
Ave. predictiveness	0.404	0.597	0.528
IC predictability	0.209	0.722	0.423
Ave. predictability	0.708	0.909	0.842
4-MPS entropy	27	11	21

We tried to present a fair analysis of Persian written verb forms, sensitive to actual use of the language. The aim of the next step of our study would be to investigate other members of the Iranian language family (and specifically West Iranian languages). As we started to look through some of the Western Iranian languages, our first impression was that unlike Persian, dialects/languages like Varzane’i, Laki and Khonji use the impersonal construction as the common strategy of verb formation, i.e. they use clitics rather than verbal affixes. If our eventual plats document the same pattern, then we may expect a rather different level of complexity across these languages.

Acknowledgements

I take this opportunity to express my profound gratitude and deep regards to Professor Greg Stump for his help and constant encouragement. I also want to express a deep sense of gratitude to Professor Raphael Finkel, without whose help this study would not have been done.

References

Stump, Gregory, & Raphael A. Finkel. 2013. Morphological typology: From word to paradigm. Cambridge: Cambridge University Press.
Stump, Gregory, & Raphael Finkel. To appear. Contrasting modes of representation for inflectional systems: Some implications for computing morphological complexity. In Matthew Baerman, Dunstan Brown & Greville G. Corbett (eds.), Understanding and measuring morphological complexity. Oxford University Press.