Heaviness in a Verb-final Language: Evidence from Persian
Pegah Faghiri, Pollet Samvelian, Barbara Hemforth

To cite this version:

HAL Id: halshs-01131210
https://halshs.archives-ouvertes.fr/halshs-01131210
Submitted on 13 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Heaviness in a Verb-final Language: Evidence from Persian

Pegah Faghirī1,2, Pollet Samvelian1 & Barbara Hemforth2

1 Université Paris 3 - Mondes iraniens et indien (CNRS), 2 Université Paris 7 - Laboratoire de Linguistique Formelle (CNRS)

pegah.faghir@univ-paris3.fr

Introduction

• Is the short-before-long principle universal?
  → Heavy constituents require more processing resources (e.g. Arnold et al, 2000; Stallings et al, 1998)
  → Costly constituents tend to be postponed
  → Meanwhile, Hawkins’s Early Immediate Constituent (EIC) principles predicts opposite tendencies
    → For head-initial and head-final languages (Hawkins, 1990 a.o.)
    → Long-before-short in head-final languages
    → Confirmed for Japanese by corpus and experimental data (Hawkins, 1994; Yamashita & Chang, 2001)

Persian: Essential Properties

• Mixed head direction:
  Head-final in verbal domain (SOV)
  Head-initial elsewhere
  e.g. Det N Mod, Prep NP, Comp P

  • EIC does not (always) work for Persian:
    NP PP V
    ➢ PP > NP (by 2 words)

  • Canonical word order is SOV
    But variation is possible (SVO, VSO, etc.)
    Depending on register, information structure, prosody, etc.
    e.g. Goal arguments are post-verbal in oral/informal register
    • Clausal complements are strictly post-verbal

• Differential Object Marking (DOM)
  A definite and/or specific DO is always marked with nro (cf. ex 4)
  An indefinite non-specific DO is not marked (cf. ex 1 – 3)

Object of this study:
Preferential order between the DO and the IO in the preverbal domain

Corpus Study

Corpus:
Bijankhan corpus : 2,6m tokens, from newspapers, annotated for POS, freely available
Verbs were lemmatized and potentially ditransitive verbs were extracted (42k tokens, 122 types)
Dataset (908 tokens, 82lemmas):
Selection of sentences out of
2. All instances of verbs ‘to send’ and ‘to pour’ (low frequency)
3. Random samples of verbs ‘to give’ and ‘to take’ (very high frequency)

4 DO types (based on preliminary observations)
Bare
(1) Maryam be Nima ketlab diid
Maryam to Nima book gave

Bare-Modified
(2) Maryam be Nima ketlab-e tarih diid
Maryam to Nima book-of history gave

(3) Maryam be Nima ketlab-e qadim diid
Maryam some old books to Nima gave

(4) Maryam this book to Nima gave
Maryam gave this book to Nima.

DO type determines the relative order in 87% of cases

Indefinite

Marked ( = nro)

Beyond the strong effect of DO type, relative length shows a significant effect (p-value < 0.001) corresponding to the long-before-short tendency

The Relative Length

Discussion and Conclusions

• The short-before-long principle is not universal:
  Not only Japanese (strictly head-final) but also Persian (mixed head-direction) presents the long-before-short tendency
  → The verbal position has to be taken into account in the effect of relative length on preferential order between verbal complements
  → Theories solely based on general principles ignoring linguistic parameters would eventually fail cross-linguistic validity
  → Theories proposing accounts in terms of dependency seems to be more appropriate

• However Hawkins’s EIC principles fails to account for Persian data

• In Persian the relative length plays only a secondary role while the DO type, which depends on the information status of the NP, plays the essential role
  We are currently running a number of experiments to explore the effect of information structure and the relative length independently

The average preference of 59% for NP-PP-V order

AFLap 2013
Université Aix-Marseille, September 2nd – 4th, 2013

This study is part of a project on word order effects across languages in the Lexus Empirical Foundations of Linguistics (ANR/CGI)