



The study of relative clause schema in Persian language with data mining methods

Nasrin Azhideh¹, Mehrdad Naghzguy-Kohan²

Received: 06/06/2022

Accepted: 05/07/2022

1. INTRODUCTION

Relative clause as a complex sentence has always been focused in cognitive linguistics. A relative clause construction (RCC) consists of four main parts: nominal head, main clause (MC), relative clause (RC) and a clause connector "*keh*". In exemplar theory all tokens of a RCC can be classified in a variety of examples. Exemplar is an individual trace from previous experience in memory. An exemplar-based representation indicates a list of words happening in a certain slot. (Bybee, 2013, p. 58). That is to say, an actual utterance of a RCC always simultaneously instantiates a number of lower-level constitutive constructions and their properties. (Wiechmann, 2015, p.3). Following the study of English RCCs in Wiechmann (2015), the present paper aims to introduce exemplar clusters (closely related to schemas in cognitive grammar) of Persian RCCs. However, this study has focused more on explaining schema occurrence. Furthermore, variables have been chosen based on RCC features in Persian.

2. MATERIALS AND METHOD

In this study, 1400 relative clauses have been gathered from 30 hours of conversations on TV and radio programs. Data analysis was conducted with computational algorithms of ARM (Association rule mining) to find probabilistic rules and CFA (configurational frequency analysis in Stefanowitsch and Gries2003, Wiechmann, 2015) to find RC Schemas. In other words, the effect of bi-valued features of RCCs have been studied with the above mentioned techniques. Finally we displayed the neighboring configurations in clusters of philological trees.

3. Results and Discussion

Some probabilistic rules for semantic and syntactic features have been displayed in table 1. The first rule indicates" If the Head is lexical then it will be a contentful head by leverage of 0.140584827"

¹ Ph.D Candidate of General Linguistics, Department of Linguistics, Literature and Humanities Faculty, BuAli Sina University, Hamedan, Iran (correspondent author); N.azhideh@basu.ac.ir

² Full Professor of Linguistics, Department of General Linguistics, Literature and Humanities Faculty Bu-Ali Sina University, Hamdan, Iran; mehrdad.kohan@basu.ac.ir

Table 1. Rules obtained with ARM technique for RC. Syntactic and semantic features

Antecedents	consequents	Leverage
Lexical	Contentful	0.140584827
Contentful	Lexical	0.140584827
Contentful, restricted	Lexical	0.123506784
lexical	Contentful, restricted	0.123506784
...

Table 2 shows configurations of both syntactic and semantic features of RCCs; the first configuration refers to non-subject RCs with inanimate heads; here MC has a less number of arguments; yet RC is not extraposed. In the second configuration, MC and RC are of the same length, still there is no extraposition again; this indicates that one use of RC which could not be accomplished with other post-head modifiers¹ is to describe heads in detail.

Table 2. Most important configurations for syntactic and semantic features

index	RC .head	MC. head	words	Extraposition	Animacy	concreteness	Pre. Head. Mod.	Content	RC type	Freq	Dec	Q
1	Nonsubject	subject	less	extraposed	Inanimate	Abstract	C	Cont.	Restr.	24	***	0.02
2	Subject	Subject	less	Central	Animate	Concrete	C	Cont.	Restr.	23	***	0.02
3	nonsubject	nonsubject	more	extraposed	Inanimate	Abstract	A	Cont.	Restr.	23	***	0.01
4	Nonsubject	subject	more	extraposed	Inanimate	Abstract	C	Cont.	Restr.	23	***	0.01
...

The highly pronounced configurations are called *Types*, conceptually close to schemas in cognitive linguistics. There are also some *Antitypes* which occur interestingly less than expected ; they contradict usual RC configurations in Persian: in antitype {Subject , more , 1 argument , 2 arguments , central }for example, the number of words in RC doesn't match the number of arguments.

1. *in pelOn.ha.øe dOramØdi ke sOvar bOr in mohtOva*
 This plan. pl. KASREH income. RC. MARKER sit on this
Sodeh budØnd besqar EesarØt vared mikØrdØn.
 content become. PAST. pp. PAST. be. 3rd PL much Damage enter PRES.
 Do. PAST. 3RD PL.

"These income plans which were upon this content had caused a lot of damage."

¹ (In this table, Pre - head modifiers were labeled according to their influence on RC prediction. If pre-head Modifiers help RC presence, they were labeled as A, modifiers having negative collconstruction value with RC were labeled as B ; and finally, in the case of no modifier , the label was C)

The clusters of combined syntactic and semantic features are demonstrated in the following philological tree:

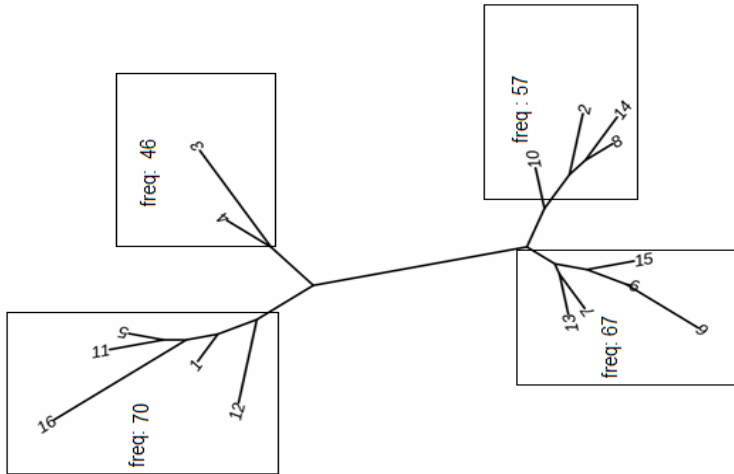


Figure 1. Dendrogram for fully specified types of semantic and syntactic features

Some of these clusters involve certain schemas: the most frequent cluster involves some representational constructions in which MC starts with **"/in.e ke "**:

2. */in.a.ee ke mi.g.Øm bØraφe /in.e*
 These.PL.INDF RC MARKER PRES. say. 1st sg for this. CL: be .PRES.3rd sg
ke *midunØm δZenab/ali Ø/hle motale/e.id.*
 ADVERBIAL MARKER PRES. Know.1st sg you group study. CL: be.
 2nd PL

"The things that I am saying is because I know you are used to studying."

The second cluster is characterized with the use of **"kØs.i / kØs.an.i"**:

3. *kØmtØr kØs.i hØst ke Σokr be δZa bφare.*
 Fewer one. INDF be. 3rs sg RC MARKER gratitude to place
 SUBJ.bring. 3rd sg

"There are few people to be grateful."

Some other schemas could be found which refer to existential sentences, general rules, etc.

4. Conclusions

Results of the present paper indicate that there are certain relative clause schemas in spoken Persian such as *ye + main clause (a copular verb) + relative clause* and they can ease similar RCC processing. The results also show that the configurations involving processing accelerator features are not essentially the easiest configurations to process. To find out the influence of each feature in an ambient space, they need to be studied extremely carefully; in this way exemplar theory helps a lot to explain RCC configural processing. It is recommended that these kinds of research be conducted over a larger corpus of Persian language.

Key words: Exemplar, Exemplar Cluster, Relative Clause Construction, Schema