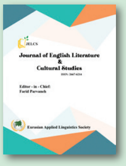
Journal of English Literature and Cultural Studies
jelcsjournal 2024;5(1):1–6
Original Research
doi: 10.14744/........

Dependency Parsing in Persian: A Rule-based Hybrid Model

Amouee, Ehsan[1]
Falahati Qadimi Fumani, Mohammad Reza[2]
Pahlevanzadeh, Bahareh Nafarzadehnafari[3]

[1]Department of Computational Linguistics, Islamic World Science and Technology Monitoring and Citation Institute, Shiraz, *Iran*

[2]Faculty Member of the Department of Computational Linguistics, Islamic World Science and Technology Monitoring and Citation Institute (ISC), Shiraz, *Iran*

[3]Faculty Member of the Department of System Operations, Islamic World Science and Technology Monitoring and Citation Institute, Shiraz, Iran

Corresponding Author: Falahati Qadimi Fumani, Mohammad Reza

Phone: 989173162506
e-mail: mrfalahat@yahoo.com

Article citation: Amouee, H., Falahati Qadimi Fumani, M. R., Pahlevanzadeh, B. (2024). Dependency parsing in Persian: A rule-based hybrid model, Journal of English Literature and Cultural.

Received Date: July 8, 2024
Accepted Date: September 19, 2024
Online Date: September 25, 2024

Publisher: Kare Publishing

E-ISSN: 2407-1155

ABSTRACT

Dependency Parsing in Persian is mostly data-driven. In this paper, we have presented a hybrid model of grammar-based parsing method consisting of Hyperedge Replacement and Constraint Dependency Grammars. The model parses Persian POS tagged sentences based on their main verb valency structure in a rule-based manner. The sample data of the study, selected through purposive sampling, included 81 Persian sentences covering all types of Persian basic sentence structures. The results of the comparison of each sentence in Gold Analysis and the corresponding output of the model with both UAS and LAS have been presented in Table 10. These scores pertain to the entire research corpus and have been calculated based on the average score of each sentence result. HRG is fully capable of representing all base sentence structures as hypergraphs and in a sentence, could generate all possibilities for any desired sequence of POS tags. In our hybrid model, HRG collaborated with CDG on dependency parsing, and with the assistance of the heuristic function, the model could achieve an accuracy of 100% among the research corpus. Further, the represented model could undertake dependency parsing with no train dataset. Also, we presented a machine-readable verb valency dictionary, based on our proposed HRG model and corpus data. Although the model's time complexity is quasi-polynomial time, the accuracy score in our sample data might be helpful.

Keywords: Hyperedge Replacement Grammar; Constraint Dependency Grammar; Dependency Parsing; Base Sentence Structure; Valency Structure.

1. Introduction

Dependency parsing, in Persian, has recently received a lot of interest among researchers. This interest is, of course, due to the availability of free word order property in this language (Rasooli, Moloodi, Kouhestani, & Minaei Bidgoli, 2011). According to Aduriz et al. (2003), dependency grammar is capable of analyzing PRO-drop languages of which Persian is a member. An overview of the literature on the topic revealed two major works (each utilizing a dataset) on dependency grammar in Persian. The first work by Rasooli, Kouhestani, and Moloodi (2013) represented Persian Dependency Treebank (PerDT) and used a fully tagged corpus as a train set. The second work was carried out by Seraji (2015) who designed the Uppsala Persian Dependency Treebank (UPDT). In her work, a fully tagged train dataset was used to achieve dependency parsing goals. Kübler, McDonald, and Nivre (2009) asserted that both of these research works used data-driven approaches. Recently, there have been some data-driven approaches for Persian language such as: Using neural pipelines for text analysis, including tokenization, multi-word token expansion, lemmatization, part-of-speech and morphological feature tagging, dependency parsing, and named entity recognition (Zhang et al. 2020), injecting semantic and syntactic features which is derived from phrase-structure parser to the MSTParser by the stacking method (Lazemi and Ebrahimpour-Komleh, 2019), a Bidirectional Long Short-Term Memory (Bi-LSTM) method to build the required features of the Maximum Spanning Tree Parser, MSTParser, (Lazemi and Ebrahimpour-Komleh, 2022) all of which require a fully tagged dataset to implement.

But, in the present article, the authors have used a grammar-based approach. Accordingly, due to the application of a rule-based method, we do not need any train dataset, and our model is capable of parsing each sentence based on the valency structure of its main verb (root). In this way, our model can recognize dependency relations between/among the proper tokens as the main valencies of the root under inspection. This was achieved by the application of Constraint Dependency Grammar, first introduced by Maruyama (1990). The represented model can also determine adjuncts, which do not participate in any valencies of the root. To do so, all sentences must first be tokenized appropriately and have POS tags assigned to them.

1. Method and Materials

In this paper, to address all the required complements, abbreviations were extracted from Tabibzadeh (2012) for a number of main Persian complement types for a verb. These complements have been listed in Table 1. Note that in our approach, the process to detect Persian 'Mosnad[1] ' is like that introduced by Nivre et al. (2016) regarding Universal Dependencies. In fact, they considered Mosnad as a root, unlike what was observed in PerDT.

Table 1: The Abbreviations Used

Abbr.	Description
advcp	Adverbial complement of a verb
clcomp	Clausal complement
cop	Copular verb (Related to Persian Mosnad)
gobj	The verb's genitive (Ezafe) complement
obj	Direct object
pobj	Prepositional complement
subj	The proposed subject
xcomp	Persian syntactic constituent: Tamiz

1 Predicate

Table 2: Desired Persian Base Structures

Row	Base Structure	Row	Base Structure
1	□*subj*□	14	□*subj,pobj,pobj*□
2	□*subj,clcomp*□	15	□*subj,pobj,gobj*□
3	□*subj,obj*□	16	□*subj,pobj,clcomp*□
4	□*subj,pobj*□	17	□*subj,pobj,xcomp*□
5	□*subj,gobj*□	18	□*subj,gobj,clcomp*□
6	□*subj,cop*□	19	□*subj,advcp,pobj*□
7	□*subj,advcp*□	20	□*subj,obj,pobj,pobj*□
8	□*subj,cop,clcomp*□	21	□*subj,pobj,pobj,advcp*□
9	□*subj,obj,pobj*□	22	□*subj,pobj,pobj,clcomp*□
10	□*subj,obj,gobj*□	23	□*subj,pobj,gobj,clcomp*□
11	□*subj,obj,clcomp*□	24	□*subj,obj,pobj,xcomp*□
12	□*subj,obj,xcomp*□	25	□*subj,obj,pobj,advcp*□
13	□*subj,obj,advcp*□	26	□*subj,pobj,pobj,gobj*□

In all, 26 base sentence structures were considered in this paper (Table 2). The subjects of all these base sentences were proposed to be obligatory even when the subject had been eliminated in the surface structure. These structures were extracted from Tabibzadeh (2012), Rasooli et al. (2011) and Tabibzadeh (2007). Two of these base structures (the structures No. 19 & 26) were proposed by us. There were a total of 81 sentences as the research corpus, which embodied at least 1 sample of each presented structure. All these instances were simple in structure and were intentionally selected from PerDT, UPDT and also from examples presented by Tabibzadeh (2012). The operation process has been described in Figure 1.

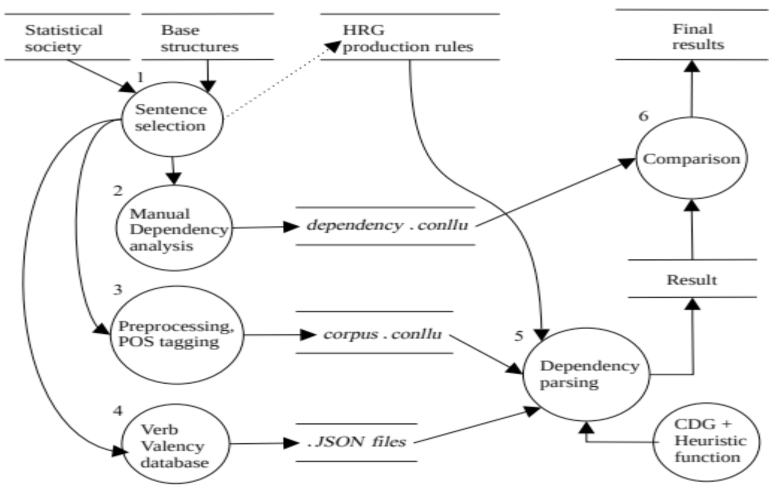

Figure 1. The Operation Process

2.1 Preprocessing

All punctuation marks were automatically eliminated and each compound syntactic constituent was manually tokenized as a single element. For instance, 'ân ketâb-e sabz-e ruy-e miz' ('that green book on the table') was considered as a single token and was labelled with a noun POS tag. All sentences were exported to CoNLL format semi-manually without any dependency relations annotated. All POS tags available in this research have been listed in Table 3.

Table 3: POS Tags

Tag	Description
ADP	Adposition
POSTP	Postposition (FPOS of ADP)
PREP	Preposition (FPOS of ADP)
NOUN	Noun phrase
VERB	Verb
ADV	Adverbial phrase
ADJ	Adjective phrase
CONJ	Conjunction
DEMP	Demonstration Pronoun
PRO	Pronoun

Table 4: Dependency Relations

Relation Tag	Description
root	Mostly the main verb
subj	Subject
clcomp	Clausal Complement
obj[+|–]	Direct Object with/without POSTP
gobj[+|–]	Genitive Object with/without POSTP
pobj	Prepositional Object
advcp	Adverbial Complement
cop	Copular Verb
xcomp	Persian Tamiz
adjunct	All possible collocations as Adjuncts
nve	None Verbal Element in Compound Verbs
adcl	Adjunct Clause

2.2 Gold Analysis

All sentences were manually analyzed for their dependencies and saved in CoNLL format with no POS tags. Each syntactic valency was described as a univalent token. Figure 2 showed a gold analysis (a dependency graph) for the sentence: "Kâmrân har ruz be daftaraš mi-ravad" (Kâmrân goes to his office every day). All dependency labels available here have been listed in Table 4. Further, all conjoining adjuncts together form a token and are considered as a dependent of the root.

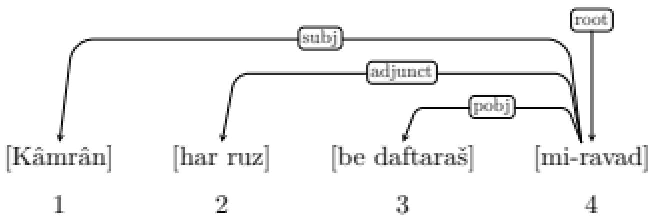

Figure 2. A Gold Analysis

2.3 Valency Database

The whole verbs present in the research corpus along with their infinitives, and the correspondence between them and their related base structure(s), collectively formed the verb valency database. In other words, the model determined the valency structure (every possible set of the base structure(s) of an infinitive) of each verb based on the information in the research corpus. All data was exported automatically in 'JSON' format according to Hyperedge Replacement Grammar (HRG), and there exists 1 file per individual infinitive presented including at least one of the basic forms introduced in Table 5.

Table 5: Base Structures Based on HRG (this research)

SUBJ.V	*SUBJ.VG*	*SUBJ.V.CLCOMP*
SUBJ.VG.CLCOMP	*SUBJ.OBJ.V*	*SUBJ.OBJ+.V*
SUBJ.OBJ+.VG	*OBJ+.SUBJ.V*	*SUBJ.POBJ.V*
SUBJ.POBJ.VG	*SUBJ.V.POBJ*	*SUBJ.VGP*
SUBJ.VRG	*SUBJ.ADVCP.VG*	*SUBJ.ADVCP.V*
SUBJ.POBJ.V.CLCOMP	*SUBJ.POBJ.VG.CLCOMP*	*SUBJ.VRC*
SUBJ.VRG.CLCOMP	*SUBJ.OBJ+.POBJ.VG*	*SUBJ.OBJ+.POBJ.V*
SUBJ.OBJ.POBJ.V	*SUBJ.POBJ.OBJ.VG*	*SUBJ.OBJ+.VGP*
SUBJ.OBJ.V.CLCOMP	*SUBJ.OBJ+.VG.CLCOMP*	*SUBJ.OBJ+.V.CLCOMP*
SUBJ.OBJ+.XCOMP.VG	*SUBJ.OBJ+.XCOMP.V*	*SUBJ.OBJ+.ADVCP.VG*
SUBJ.OBJ+.ADVCP.V	*SUBJ.POBJ.POBJ.VG*	*SUBJ.POBJ.V.POBJ*
SUBJ.POBJ.VGP	*SUBJ.POBJ.XCOMP.VG*	*SUBJ.VGP.CLCOMP*
SUBJ.POBJ.ADVCP.VG	*SUBJ.ADVCP.POBJ.V*	*SUBJ.ADVCP.POBJ.VG*
SUBJ.OBJ+.POBJ.POBJ.VG	*SUBJ.POBJ.POBJ.VG.CLCOMP*	*SUBJ.POBJ.POBJ.V.CLCOMP*
SUBJ.POBJ.VGP.CLCOMP	*SUBJ.POBJ.POBJ.ADVCP.VG*	*SUBJ.OBJ+.POBJ.XCOMP.V*
SUBJ.OBJ+.POBJ.XCOMP.VG	*SUBJ.POBJ.OBJ+.XCOMP.V*	*SUBJ.ADVCP.POBJ.OBJ.V*
SUBJ.OBJ+.POBJ.ADVCP.V	*SUBJ.POBJ.POBJ.VGP*	

2.4 Dependency Parsing

The hybrid model of this research consisted of 2 related but distinct parts: 1) HRG section which role played modeling of each existing base sentence structure, and 2) Constraint Dependency Grammar (CDG) unit with its heuristic function, which guaranteed finding the best possible results in the research corpus.

2.4.1 HRG

As posited by Habel (1992) and Rozenberg (1997), a Hyperedge Replacement Grammar requires a tuple: (N, T, P, S), in which N stands for non-terminal labels and T stands for terminal ones. Similarly, P and S stand for production rules and start symbol respectively. There were 34 labels including non-terminals (N) and terminals (T) in this paper, the list of which has been presented in Table 6.

Table 6: HRG Labels

Label	Description	Type	Label	Description	Type
NP	Noun Phrase	N	PP	Prepositional Phrase	N
ADJP	Adjective Phrase	N	ADVP	Adverbial Phrase	N
ADCL	Adjunct Clause	N	CLCOMP	Clausal Complement	N
ADJUNCT	Adjunct	N	ADVCP	Adverbial Phrase	N
OBJ	Object [–POSTP]	N	OBJ+	Object [+POSTP]	N
GOBJ	Genitive Object [–POSTP]	N	GOBJ+	Genitive Object [+POSTP]	N
POBJ	Prepositional Object	N	SUBJ	Subject	N
V	Governor Group: Simple Verb	N	VG	Governor Group: Compound Verb	N
VGP	Governor Group: Compound Verb + gobj	N	VRG	Governor Group: Copular Verb	N
VRC	Governor Group: Copular Verb + adcl	N	XCOMP	Persian Tamiz	N
NVE	None Verbal Element	N	CPE	Persian Mosnad	N
VERB	Verb	T	CONJ	Conjunction	T
HASVERB	Clause	T	ADJ	Adjective	T
ADV	Advderb	T	ADJCT	Non-empty Adjunct	T
EMPTY	Null-expression	T	POSTP	Persian 'râ'	T
PREP	Preposition	T	NOUN	Noun	T
PRO	Pronoun	T	DEMP	Demonstration Pronoun	T

Governor Group: This research assumes, based on its HRG primary aspect, that every syntactic constituent must have obvious and distinct boundaries, that is, no constituent could be inside another one.

Since there are some instances in the real world that violate this rule, we presented in our research an abstract syntactic constituent called Governor Group, which involves all those real syntactic constituents together as a single element. The Governor Group has the root of the sentence within itself and is not represented in the final dependency graph.

It is possible to have more governor groups in Persian Language than we have considered here. In fact, we considered five governor groups based on the research corpus of the study. For instance, in the sentence: "lotfan darxâst-e nesye nakonid!" (Please do not ask for credit!) its governor group, VGP, could be written as: gov(nve(darxast-e) + gobj(nesye) + root(nakonid)). Its compound main verb: "darxast nakonid" has a genitive object amongst.

2.4.2 CDG

The second part of our model concerned the Constraint Dependency Grammar (CDG) with its heuristic functions. CDG in this model checked every possible sequence of tokens against some desired syntactic constituents which were determined by the HRG section with a specific order of the POS tags. CDG checked several possibilities for dependency graphs per sentence, and the heuristic function served to determine the best possibility by enumerating the adjunct score.

2.5 Evaluation

The last process was to evaluate the results of the model and manual dependency analysis for each available sentence against Unlabeled and Labeled Attachment Scores.

3. Calculation

HRG, during the derivation process, starts from the axiom S (as shown in Figure 3) and pursuant to the production rules, replaces each non-terminal label and continues this process until all labels are terminal.

Figure 3. Axiom S

Definition 1 is required to represent our model's production rules. The set V is the vertex of a hypergraph, and L, union set of N and T. Production rules are shown in Table 7 (also c.f. Table 6).

Table 7: Production Rules

No.	Name		definition
0	S	::=	One of base structures according to HRG
1	NP	::=	hg ((0, NOUN, 1), (0, PRO,) 1), (0, DEMP, 1))
2	PP	::=	hg ((0, PREP, 1), (1, NP, 2))
3	ADJP	::=	hg ((0, ADJ, 1))
4	ADVP	::=	hg ((0, ADV, 1))
5	ADCL	::=	hg ((0, CONJ, 1), (1, HASEVERB, 2))
6	CLCOMP	::=	hg ((0, CONJ, 1), (0, EMPTY, 1), (1, HASVERB, 2))
7	ADJUNCT	::=	hg ((0, ADJCT, 1), (0, EMPTY, 1))
8	ADVCP	::=	hg ((0, ADJUNCT, 1), (1, ADV, 2))
9	OBJ	::=	hg ((0, ADJUNCT, 1), (1, NP, 2))
10	OBJ+	::=	hg ((0, ADJUNCT,1), (1, NP, 2), (2, POSTP, 3))
11	GOBJ	::=	hg ((0, NP, 1))
12	GOBJ+	::=	hg ((0, NP, 1), (1, POSTP, 2))
13	POBJ	::=	hg ((0, ADJUNCT, 1), (1, PP, 2))
14	SUBJ	::=	hg ((0, ADJUNCT, 1), (1, NP, 2), (1,) EMPTY, 2))
15	V	::=	hg ((0, ADJUNCT, 1), (1, VERB, 2))
16	VG	::=	hg ((0, ADJUNCT, 1), (1, NVE, 2), (2, VERB, 3))
17	VGP	::=	hg ((0, ADJUNCT, 1), (1, NVE, 2), (2, GOBJ, 3), (2, GOBJ+, 3), (3,VERB, 4))
18	VRG	::=	hg ((0, ADJUNCT, 1), (1, CPE, 2), (2, VERB, 3))
19	VRC	::=	hg ((0, ADJUNCT, 1), (1, CPE, 2), (2,) VERB, 3), (3, ADCL, 4))
20	XCOMP	::=	hg ((0, NP, 1), (0, PP, 1), (0, ADJP, 1))
21	NVE	::=	hg ((0, NP, 1), (0, ADJP, 1), (0, PP, 1))
22	CPE	::=	hg (0, NP, 1), (0, ADJP, 1), (0, ADVP, 1))

The function "hg" presents each possible hyperedge in each desired hypergraph of this research. All hyperedges here are (1,1)-edge and all hypergraphs are noted to be (1,1)-hypergraph in this paper. Therefore, the axiom S could be represented as: hg((0,S,1)) in which $S \in N$ and V={0,1}.

Definition 1.

$$hg(\{(i,l,j) \vee i,j \in V, l \in L\})$$

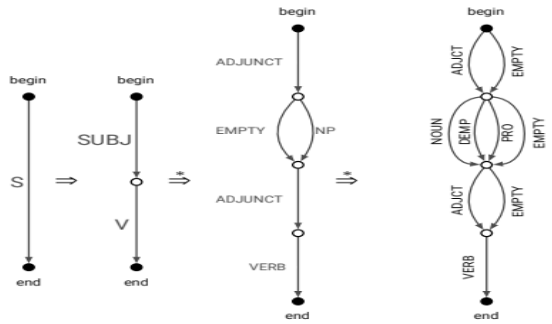

Figure 4. Derivation Process

In this paper, 50 base sentence structures were presented according to HRG (Tables 5 & 6). In HRG, there is an important point which any label related to the Governor Groups (mostly the corresponding main verb) of the structures, must be mentioned. This is unlike Table 2, where the verb part is not presented at all. For instance, the base structure: □subj□ according to HRG should be modeled in 2 distinguished ways: 'SUBJ.V' and 'SUBJ.VG'. The derivation process for the structure: 'SUBJ.V' has been illustrated in Figure 4.

Rows No. 15 to 19 in Table 7 refer to Governor Groups. Governor Groups are placed in each base sentence structure. If a sentence has k tokens, with the HRG base structure of n syntactic category, CDG at the very beginning calculates k-combinations with repetitions from category set of size n. Each multisubset might lead to a result candidate and must be well checked. In case all elements in a specific multisubset are confirmed by CDG against the desired constraints, it would be one of the possible results. To do so, by assuming that all phrases have an adjunct accompanied – even though they are null – for each element in a calculated multisubset, CDG considers 2-multicombinations from a set of 2 members: adjunct and phrase. If this phrase satisfies the constraint rule, the element is temporarily flagged as passed and its corresponding adjunct is considered separately.

The heuristic function calculates the adjunct score in 2 phases. First, it calculates the first-order adjunct score, when CDG detects passed multisubsets. Then, in the case of necessity, it calculates the second-order adjunct score. The first-order adjunct score is the sum of all tokens which stand in the adjunct parts within a given multisubset. As we proposed, the analysis with the least adjunct score would be the winner. Considering the first-order adjunct scores, each multisubset has only 1 victorious candidate. Sometimes, there is no unique smallest amount in all candidates' first-order adjunct scores. In this case, the second-order adjunct score could identify the best unique choice. An existing subject decreases, and a null-subject increases the adjunct score as we proposed in this paper. Furthermore, for Governor Groups, identification of their adjunct parts increases the score and vice versa. Table 8 has listed the competing candidates for the sentence: "?u tond david." (He/She ran fast.) leading to the first-order adjunct scores for the structure: 'SUBJ.V', considering multisubsets of size 3 (no. of tokens) from a set of two members: {SUBJ, V }. The sign 'gov' indicates the Governor Group. Further, the model detects empty phrases as λ, as they never appear in saved results. Finally, all adjuncts are also supposed to be the dependents of the root.

Table 8: The Model's Possible Candidates

Analysis	Score
adjunct(λ) + *subj*(λ) + ***gov*(***adjunct*(?u tond) + *root*(david)**)**	2
adjunct(?u tond) + *subj*(λ) + ***gov*(***adjunct*(λ) + *root*(david)**)**	2
adjunct(λ) + *subj*(?u) + ***gov*(***adjunct*(tond) + *root*(david)**)**	1

3.1 Algorithm Complexity

We have assumed a common situation in which a sentence has k tokens, and n syntactic categories in its HRG base structure. In this case, within a multisubset, the CDG section calculates every single element according to its 2-multicombinations in the complexity time of O(1). All main multisubsets are calculated as below:

$$c = \binom{n+k-1}{k}$$

The availability of m tokens within each element of a multisubset leads to m+1 multisubsets again. In other words, since h syntactic paths existed in a specific syntactic constituent (for example, 'SUBJ' leads to 8 paths as shown in Figure 4), the number of comparisons in CDG unit per each sentence could be calculated as:

$$\sum_{j=1}^{c} h \times \prod_{i=1}^{n} (m_{ij}+1)$$

With $2 \leqslant k \leqslant 7, 2 \leqslant n \leqslant 5$ and as a mean score h=7, with the assumption of $t=n^2+k$, the execution time according to Table 9 could be estimated as $t^{(log(t)+1.7)}$, which means that the algorithm complexity is: $2^{(poly(log_2 n))}$.

Table 9: Amount of Execution

n	k	t	Execution	n	k	t	Execution
2	2	6	70	4	4	20	2310
2	3	7	140	4	5	21	5544
2	4	8	245	4	6	22	12012
3	3	12	392	5	5	30	14014
3	4	13	882	5	6	31	35035
3	5	14	1764	5	7	32	80080

4. Results and Conclusions

The results of the comparison of each sentence in Gold Analysis and the corresponding output of the model with both UAS and LAS have been presented in Table 10. These scores pertain to the entire research corpus and have been calculated based on the average score of each sentence result.

HRG, as indicated in Table 10, is fully capable of representing all base sentence structures as hypergraphs and in a sentence, could generate all possibilities for any desired sequence of POS tags. In our hybrid model, HRG collaborated with CDG on dependency parsing, and with the assistance of the heuristic function, the model could achieve an accuracy of 100% among the research corpus.

Further, the represented model could undertake dependency parsing with no train dataset. Also, we presented a machine-readable verb valency dictionary, based on our proposed HRG model and corpus data. Although the model's time complexity is quasi-polynomial time, the accuracy score in our sample data might be helpful. Other researchers interested in this topic might use the findings of this paper to extend the present model such that it can recognize the internal parts of compound constituents as well as all kinds of sentences in the Persian language.

Table 10: Final Results

UAS	LAS
100	100

List of Symbols

Symbol	Example	Symbol	Example	Symbol	Example
â	car /kâr/	d	desk /desk/	s	six /siks/
a	cat /kat/	f	fact /fakt/	t	two /tu/
e	pen /pen/	h	he /hi/	v	vax /vaks/
i	seat /sit/	k	kick /kik/	y	yes /yes/
u	two /tu/	m	me /mi/	z	zoo /zu/
o	core /kor/	n	nick /nik/	š	She /ši/
b	back / bak/	r	red /red/	?	Cat /ka?t/ in British English
x	Bach / b â x / (German)	l	left /left/		

References

Aduriz, I., Aranzabe, M., Arriola, J., Atutxa, A., Dìaz de Ilarraza, A., Garmendia, A., & Oronoz, M. (2003). Construction of a basque dependencytreebank. Retrieved June 12, 2022 from http://ixa.si.ehu.es/sites/default/files/dokumentuak/3240/TLT2003_pdf.pdf.

Habel, A. (1992). Hyperedge replacement: grammars and languages (Vol. 643). Springer Science & Business Media. Retrieved March 19, 2022 from https://link.springer.com/book/10.1007/BFb0013875.

Kübler, S., McDonald, R., & Nivre, J. (2009). Dependency parsing. Synthesis lectures on human language technologies. Morgan and Claypool Publishers.

Lazemi, S., & Ebrahimpour-Komleh, H. (2019). Feature engineering in Persian dependency parser. Journal of AI and Data Mining, 7(3), 467-474.

Lazemi, S., Ebrahimpour-komleh, H., & Noroozi, N. (2022). Improving Persian dependency-based parser using deep learning. Computer and Knowledge Engineering, 5(1), 13-20.

Maruyama, H. (1990). Structural disambiguation with constraint propagation. In 28th Annual Meeting of the Association for Computational Linguistics, (pp. 31-38).

Nivre, J., De Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., … others (2016). Universal dependencies v1: A multilingual tree-bank collection. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (lrec’16) (pp. 1659–1666).

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. arXiv preprint arXiv:2003.07082.

Rasooli, M. S., Kouhestani, M., & Moloodi, A. (2013). Development of a Persian syntactic dependency

treebank. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 306–314).

Rasooli, M. S., Moloodi, A., Kouhestani, M., & Minaei-Bidgoli, B. (2011). A syntactic valency lexicon for persian verbs: The first steps towards Persian dependency treebank. In 5th Language & Technology Conference (ltc): Human Language Technologies as a Challenge for Computer Science and Linguistics (pp. 227-231).

Rozenberg, G. (1997). Handbook of graph grammars and computing by graph transformation (Vol. 1). World scientific.

Seraji, M. (2015). Morphosyntactic corpora and tools for Persian (Unpublished doctoral dissertation). Acta Universitatis Upsaliensis.

Tabibzadeh, O. (2007). Verb valency and basic sentence structures in modern Persian: A dependency-based approach. Tehran: Nashr-e Markaz.

Tabibzadeh, O. (2012). Persian grammar: A Theory of Autonomous Phrases Based on Dependency Grammar. Tehran: Nashr-e Markaz.