

Reflexive Anaphora Resolution in Pashto Discourse

Rahman Ali, Mohammad Abid Khan and Mushtaq Ali

Department of Computer Science, University of Peshawar, Pakistan

rahmanali.scholar@gmail.com, abid_khan1961@yahoo.com, mushtaqali537@yahoo.com

Abstract

The paper presents a rule-based approach for the reflexive pronouns (ReflPro(s)) resolution problem in Pashto language. Here, first the rules are described and exemplified and then the algorithm for identifying the noun phrase antecedents of reflexive anaphors is developed. Finally, the proposed algorithm is evaluated against a manually annotated Pashto corpus. The algorithm successfully identifies the antecedents of about 87.0% of Pashto reflexive pronouns.

1. Introduction

Anaphora resolution is a major area of research in Natural Language Processing (NLP) because anaphora results in text cohesion. For a machine translation system a cohesion free text is needed. The resolution of pronouns in natural languages is an important process for several NLP applications including information extraction, question answering, and text summarization [7].

The task in this paper is focused on ReflPro(s) resolution in Pashto language. It uses a knowledge-poor approach to resolve these pronouns. There are many recent approaches to anaphora resolution in different languages which are knowledge-poor [6,9,5,11,12,10,8,1,2,3]. The ReflPro(s) resolution in English language has been done by the anaphor binding algorithm [8].

Pashto reflexive anaphora and reciprocal anaphora are two subcategories of lexical anaphora [3]. In Pashto language, some work has recently been done on anaphora resolution. Here the personal and reciprocal pronouns have been worked out computationally [1,2,3,13]. For these pronouns, rule-based approaches have been adopted and algorithms have been developed. Similarly, theoretical work has also been done on Pashto text to define different types of Pashto anaphora [4].

This paper is focused on the resolution of ReflPro(s) in Pashto language. Here, a knowledge-poor rule-based approach is used to perform the task.

The rest of the paper is organized as follows. Section 2 describes reflexive anaphora in Pashto language. Section 3 describes rules with the help of examples that tell that how these pronouns are resolved. Section 4 outlines the rules in the form of an algorithm. Section 5 describes the architecture of the proposed system. Section 6 evaluates the approach while Section 7 concludes the work done.

2. Reflexive anaphora in Pashto

When the Pashto ReflPro(s) point back to some previously mentioned item in the text then this phenomenon is termed as reflexive anaphora. For example:

(2.1)

پټوي خو دا هم د چا د لاسه - خکه چې ¹خان که هر ²خومره او مخ
حسن بل حسن ته ²هم دومره بي وسه شي چې ²خپل خني وخت مخ
وراندې کولې نه شي-

[Hidayat, The poetry of Rahman Baba¹⁴ (د رحمان بابا شاعري سادګي او معنويت) U., H” pp-18]

[aw] [məx] [kə] [hər cúmra] [jān] [paṭawi] [xo] [dā]
[húm] [də] [čā] [də] [lāsa].[jaka] [če] [jne] [waxt]
[məx] [húm] [dúmra] [be wasa] [šI] [če] [xpəl] [húsn]
[bəl] [húsn] [ta] [wřānde kawlay na šay].
[And] [face] [if] [how much] [self] [conceal] [but]
[this] [also] [of] [some one] [of] [reason]. [because]
[that] [some] [time] [face] [also] [so much] [oblige]
[become] [that] [self] [beauty] [another] [beauty] [to]
[show-Neg].

“And as much a face conceals itself, it is due to someone. The reason is that sometimes face is so much helpless that cannot show its beauty in the presence of another beauty.”

In the above example, a ReflPro (خان [jān]-self) that appears in the first sentence refers back to the noun (مخ [məx]-face) in the same sentence. Similarly, another ReflPro (خپل [xpəl]-self) appears in the last sentence and refers to the same noun (مخ [məx]-face) that exists in the same sentence.

Reflexive anaphora is a subcategory of lexical anaphora [3]. Pashto lexical anaphora is composed of:

reciprocal and reflexive [3]. The Pashto reflexives are shown in table 1.

Table1. Pashto reflexive pronouns

Singular (Mas+Fem)	Plural (Mas + Fem)
خان [jān]-Self	خان [jān]-Selves, خانونه [jānona]-Selves
خان [xpəl jān]-Self	خان [xpəl jān]-Selves, خانونه [xpəl jānona]-Selves

In English, the ReflPro(s) always refer to animate noun phrases (NPs) in the text but the Pashto reflexives can refer to inanimate NPs as well. This is clear from example (2.1) in which the reflexive pronoun (خان [jān]-self) and خپل [xpəl]-self refer to inanimate NP (مخ [məx]-face). Similarly, unlike the English reflexives, the Pashto reflexives appear detached from personal pronouns in most of the cases. The Pashto ReflPro(s) are equivalent to the English himself, herself, itself and themselves.

3. Rules for the resolution of reflexive anaphors

There are some rules which are helpful in the resolution process of reflexive pronouns. These are subject preference rule, reflexives connected with personal pronouns, reflexives preceded by a noun or noun phrase, personal pronouns as subject in the same sentence, number agreement and recency.

3.1. Subject preference rule

In Pashto, the ReflPro(s) mostly prefer those NPs as candidates for antecedent which act as subject (SUB). For example:

(3.1.1)

چې څوک د سرو سپینو¹ ادا یو زبردست طنز دے - په هغه ظالمانو
مانی جوړوي- د اولس ویني څکي- د هغوي حقونه تروري - هغوي
اباد وي - خپل خان بربادوي او

[Hidayat, The poetry of Rahman Baba¹⁴ (د رحمان بابا شاعري سادګي او معنويت), pp-30] U., H”

[dā] [yaw] [zabrɔdast] [tənz] [day]. [pə] [háya] [z. ālimano] [čə] [cúk] [də] [sro spino] [māṇəy] [júrɔw]. [də] [aúləs] [wine] [ckI]. [də] [háyúI] [ḥaqona] [tarúrI]. [háyúI] [barbādɔwI] [aw] [xpəl jān] [ábáɔwI] [It] [one] [fírm] [mock-] [is]. [on] [those] [tyrant-pl] [that] [who] [luxurious] [palace-pl] [construct-optative]. [of] [community] [blood] [suck]. [their] [right-pl] [snatch]. [them] [wreck] [and] [selves] [build-].

“It is a great humor on those cruel people who raise luxurious palaces, suck the blood of the nation, snatch their rights, wreck them and build themselves.”

Here, the list of candidate antecedents for the Pashto ReflPro (خان خپل [xpəl jān]-self) is (ظالمانو [z. ālimano]- tyrants, د سرو سپینو مانی [də sro spino māṇəy]- luxurious palaces, د اولس ویني [də aúləs wine]- blood of the nation, حقونه [ḥaqona]-rights). The subject preference rule rules out all the candidates from the list except the SUB (ظالمانو [z. ālimano]- tyrants) and thus select it as the correct antecedent.

3.2. Reflexives connected with personal pronouns

If a Pashto ReflPro is immediately preceded by a personal pronoun (PersPro) then the ReflPro refers to the PersPro for its antecedent. In this case, there is no need to resolve the ReflPro but only the PersPro is resolved by using the algorithm of Ali and his co-authors [1]. For example:

(3.2.1)

سره وارث کاکا روان کړو - خان¹ چې داسې ساختي وخت شو نو ما
[Hydayat, U., H., “Vootona¹⁴ (ووتونه), a Pashto Novel,” Page-53] [čə] [dāse] [sāxtI wəxt] [šo] [no] [mā] [jān] [sará] [wārə kākā] [rawān] [kɔ]. [that] [similar] [before noon] [time] [reach] [then] [I] [self] [along with] [Waris Uncle] [take-perfective]. “When it was before noon, I took uncle Waris with me and went.”

In the example (3.2.1), the ReflPro (خان [jān]-self) is immediately preceded by the first person pronoun (ما [mā]-I). In this case, the antecedent of the ReflPro (خان [jān]-self) is the same as that of the PersPro (ما [mā]-I). Only the PersPro (ما [mā]-I) is resolved by Ali and his Co-authors algorithm [1] and the ReflPro (خان [jān]-self) remains as it is.

3.3. Reflexives preceded by a noun or noun phrase

If a Pashto ReflPro is immediately preceded by a noun (N) or NP that acts as a SUB then the ReflPro refers to that N or NP for its antecedent. For example:

(3.3.1)

ته ببوزي وهل - خان¹ ګرمی کښي به خلقو
[Yar, M., J., “Gulmeena¹⁵ (گل مينه), Pashto Novel,” Page-64] [garməy] [kšə] [bə] [xəlqo] [jān] [tə] [babozI] [wahəl]. [summer] [in] [may] [people] [selves] [to] [hand fan-pl] [used]. “People used hand-fans in summer.”

In the example (3.3.1), the ReflPro (خان [jān]-self) is immediately preceded by the plural (pl) N (خلقو [xəlqo]-people), which acts as a SUB. Hence, it is the correct antecedent.

3.4. Personal pronoun as subject in the same sentence

Sometimes the Pashto PersPro(s) appears in the same sentence in which the ReflPro(s) appear and act as a SUB. In such case, if the above rules fail in identifying the antecedent then the Pashto ReflPro refers to that PersPro. Here, the system calls the algorithm of Ali and his co-authors [1] to resolve that reflexive pronoun of Pashto. For example:

(3.4.1)

لنډي څخه لاس چې ۲ بز ته برابر نه شو - د هغه څخه لاس گلاب
توپک ته رسېدو نو د هغه نه مخکې هغه څو بزه لگېدله و -

[Hidayat, U., H., “Roogha Washawa¹⁴ (روغه اوشوه), a Pashto novel”,
PP-1]

[xo] [gúlāb] [xpəl] [dəz] [tə] [barábər nə šo]. [də
hayá] [lās] [čə] [xpəl] [lnɖI] [topak] [tə] [rsedo] [no]
[də haye] [nə] [mxkʃe] [hayá] [co] [dazá] [lgedalay
wo.]

[but] [Gullab] [self] [fire] [to] [set-Neg]. [his] [hand]
[that] [self] [small] [gun] [to] [reach-] [hence] [of that]
[from] [before] [he] [few] [bullet-pl] [shoot-perfective].
“But Gullab was unable to fire. Before his hand
reached to his small rifle, he was shot by several
bullets.”

Here, in this example the first ReflPro (خپل [xpəl-
self]) that appears in the first sentence is resolved by
applying the rule of section (3.3). But, the ReflPro (خپل
[xpəl-self]) of the second sentence is resolved to the
PersPro (د هغه [də hayá]-his) because it acts as a SUB
in this sentence.

3.5. Number agreement

The Pashto ReflPro(s) (خانونه [jānona]-Selves, خپل
خانونه [xpəl jānona]-Selves) refer to the pl NPs that act
as SUB in the preceding sentence(s). The rest of
reflexive pronouns cannot be resolved by applying this
constraint because they are ambiguous for number
agreement. For example:

(3.5.1)

ته ببوزي وهل - د کوهي يا چيني تازه اوبه لځان ګرمۍ کښې به خلقو
لځانونه به ئې څښلې - څپر د اونې سپوري يا بيا د ولي غارې ته به ئې
رسول -

[Yar, M., J., “Gulmeena¹⁵ (” گل مينه “), Pashto Novel,” Page-64]
[garməy] [kʃe] [bə] [xəlqo] [jān] [tə] [babozI] [wahəl].
[də] [kohI] [yā] [čine] [tāzə] [awba] [bə] [ye] [çʃle].

[cpər] [də] [aone] [sewrI] [yā] [biā] [də] [wale] [yāre]
[tə] [bə] [ye] [jānona] [raswəl].

[summer] [in] [may] [people] [selves] [to] [hand fan-pl]
[blow]. [Of] [well] [or] [spring] [fresh] [water] [may]
[them-clitic] [drink-]. [Wattle] [of] [tree] [or] [then]
[of] [offshoot of canal] [bank] [to] [may] [them-clitic]
[themselves] [reach].

“In the summer, people used hand-fans. They used to
take fresh water from a well or a spring. They would
try to reach to the shade of a wattle tree or the shore of
a canal.”

Here, in this example the pl ReflPro (خان
[jān]-selves) of the first sentence is resolved by applying the
rule of section (3.3). The pl ReflPro (خانونه [jānona]-
Selves) of the last sentence has a list of possible
candidates (د کوهي يا چيني تازه اوبه, خلکو [xəlqo]-people) و
[də kohI yā čine tāzə awba]-fresh water of a well or a
spring, د کوهي يا چيني تازه اوبه [cpər də aone sewrI]-shade of a
wattle tree, د ولي غاري [də wale yāre]-bank of a canal).
But here, the number agreement constraint rules out all
others except (خلکو [xəlqo]-people) because they are
singular (Sg) and do not agree in number with the
anaphor (خانونه [jānona]-Selves).

3.6. Recency

If all the above rules fail then the nearest NP has
the highest preference to be the antecedent of the
Pashto Reflexives. For example:

(3.6.1)

ته او بيلى بيلي لځان ته لځان ۱ حَسَن يا حَسِين وي خو د هر حَسَن
کښې حَسَن لري - څخه لځان په ۲ څپرې وي - هر څيز

[Yar, M., J., “Gulmeena¹⁵ (” گل مينه “), Pashto Novel,” Page-18]

[hūsən] [yā] [ħasin] [wI] [xo] [də] [hər] [hūsən] [jān]
[tə] [jān] [tə] [aw] [bile] [bile] [cehre] [wI]. [hər ciz]
[pə] [xpəl jān] [kʃe] [hūsən] [larI]

[beauty] [or] [beautiful] [is-optative] [but] [every]
[beauty] [self] [to] [self] [to] [and] [separate] [separate]
[face-pl] [are-optative]. [Every] [thing] [at] [self] [in]
[beauty] [possess].

“If it is beauty or beautiful, but every beauty has its
own peculiarity. Everything has its own beauty.”

In this example, the ReflPro (خان [jān]-self) of the
first sentence refers to the nearest NP (هر حَسَن [hər
hūsən]-every beauty) and the ReflPro (خپل خان [xpəl
jān]-self) of the last sentence refers to NP (هر څيز [hər
ciz]-every thing) of the same sentence.

4. Algorithm

1. Search the Pashto text, annotated manually, to identify the ReflPro (s).

2. for each (discourse)_i in the text
 - a. for each ReflPro in the (discourse)_i
 - i. for each rule R_j in the Rule Table
 1. If (R_j.Condition) is satisfied by the (discourse)_i, then goto step 2.
 2. Apply R_j.Rule of the Rule Table to get the antecedent (R_j.Antecedent) or apply (R_i.Action).
 3. end if
 - ii. end for
 - b. end outer for
3. end outermost for

5. Architecture of the system

The architecture of the proposed system consists of three modules which are reflexives identifier, noun phrase extractor and antecedent selector as shown in Figure 1. The reflexives identifier module scans the manually annotated Pashto text and identifies the Pashto reflexive pronouns. When the ReflPro is identified then the noun phrase extractor module takes turn and extracts the N(s) and NP(s) in the same and preceding four sentences. These NPs are stored in a structure for further processing. The last and most important module is the antecedent selector, which takes the list of NPs and applies the ReflPro(s) resolution rules given in the Appendix 1. These rules rule out the inappropriate antecedents from the list and select the correct antecedent. The abbreviations used in the paper, Rule Table 3 and algorithm are shown in Table 2.

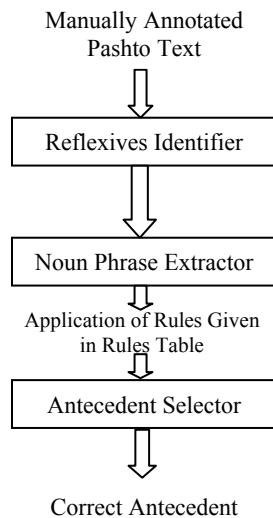


Figure 1: Architecture of the proposed system

Table 2: List of abbreviations

Term	Abbreviation	Term	Abbreviation
NLP	Natural Language Processing	NP	Noun Phrase
pl	Plural	N	Noun
Sg	Singular	SUB	Subject
PersPro	Personal Pronoun	R	Rule #
ReflPro	Reflexive Pronoun	Mas	Masculine
Neg	Negation	Fem	Feminine

6. Evaluation

For Pashto language no annotated corpus is available till now. Therefore, the input text is manually annotated and tagged. The tested input text file is composed of 300 ReflPro(s). The algorithm correctly resolved 261 ReflPro(s) and showed an accuracy of 87.0 %.

7. Conclusion

This paper resolves the reflexive pronouns of Pashto language by applying rules. An algorithm is developed that operates on the Pashto text, which is annotated manually. The proposed algorithm shows 87.0% accuracy.

References

- [1] R. Ali, M.A. Khan and I. Rabbi, "Strong Personal Anaphora Resolution in Pashto Discourse", *In proc. IEEE ICET 3rd International Conference on Emerging Technologies*, Islamabad, Pakistan. 2007, pp 148-154.
- [2] R. Ali, M.A. Khan, R. Ahmad and I. Rabbi, "Rule Based Personal References Resolution in Pashto Discourse for Better Machine Translation", *In proc. IEEE ICEE 2nd International Conference on Electrical Engineering*, UET Lahore, Pakistan. (25-26 Nov 2008), pp 57-62.
- [3] R. Ali, M.A. Khan, M. Bilal and I. Rabbi., "Reciprocal Anaphora Resolution in Pashto Discourse", *In proceeding 4th IEEE International Conference on Emerging Technologies (IEEE-ICET2008)*, College Of Electrical & Mechanical Engineering (E&ME), Rawalpindi 46000, Pakistan, October 18-19, 2008, pp 1-5.
- [4] R. Ali, M.A. Khan, M. Bilal and I. Rabbi., "Emperical Analysis of Pashto Text for Types of Pashto Anaphora", *In proc. International Conference on Information and Communication Technologies (IC-ICT2008)*, Banue University of Science and Technology, Banue, Pakistan. 2008.

- [5] B. Baldwin, Cogniac: “A discourse processing engine”, Ph.D. thesis, Department of Computer and Information Sciences, University of Pennsylvania, Philadelphia, PA. 1995.
- [6] S. E. Brennan, M. W. Friedman and C. Pollard, “A Centering Approach to Pronouns”, *In proc. 25th Annual meeting of the Association for Computational Linguistics (ACL)*. 1987. pp 155–162.
- [7] R. Iida, K. Inui, and Y. Matsumoto, “Anaphora Resolution by Antecedent Identification Followed by Anaphoricity Determination”, *ACM Transaction on Asian Language Information Processing*, Vol. 4, December 2005. pp 417-434.
- [8] M.A. Khan, M. Naveed. Ali and M. Amir. khan, “Pronominal Anaphora Resolution in Urdu Discourse”, *In proc. IEEE ICET 2nd International Conference on Emerging Technologies*. Peshawar, Pakistan. 2006. pp 543-548.
- [9] S. Lappin, and H. J. Leass, “An algorithm for pronominal anaphora resolution”, *Computational Linguistics*, 20, 4. 1994, pp 535–561.
- [10] R. Mitkov, “Factors in anaphora resolution”, *In proc. 36th Annual Meeting of the Association for Computational Linguistics*. 1997.
- [11] H. Nakaiwa and S. Shirai, “Anaphora resolution of Japanese zero pronouns with deictic reference”, *In proc. of the 16th International Conference on Computational Linguistics (COLING)*. 1996, pp 812–817.
- [12] M. Okumura and K. Tamura, “Zero pronoun resolution in Japanese discourse based on centering theory”, *In proc. 16th International Conference on Computational Linguistics (COLING)*. 1996, pp 871–876.
- [13] R. Ali, M.A. Khan and R. Ahmad, “Implementation of the Rule-based approach for the Resolution of Strong Personal Anaphora in Pashto Discourse”, *Accepted in 12th IEEE International Multitopic Conference (IEEE INMIC 2008)*, Bahria University Karachi, Pakistan. 23-24 December 2008, “in press”.
- [14] Hydayat, U., H., “Toor Paroonay, Girl Friend and Roogha Washwa, Pashto Nonels”, Charsada College, N.W.F.P, Pakistan. 2005.
- [15] Yar, M., J., “ Gulmeena, a Pashto Novel”, Pashto Academy, University of Peshawar, N.W.F.P, Pakistan. 2006.

Appendix 1: Rules for the resolution of Pashto reflexive pronouns

Pronoun	Rule #	Rule	Condition	Antecedent/Action	
ReflPro (خان [jān]-Self, خپل خان [xpəl jān]-Self, خانونه [jānona]-Selves, خپل خانونه [xpəl jānona]-Selves)	R1	Reflexive Pronoun Preceded by a Noun/Noun Phrase	... (SUB {N/NP-pl/Sg}), (ReflPro)...	SUB {N/NP-pl/Sg}	
	R2	Reflexives Connected With Personal Pronouns	... (SUB {PersPro-pl/Sg}), (ReflPro)...	SUB {PersPro-pl/Sg} → call strong personal anaphora resolution algorithm (Ali et al, 2007)	
	R3	Subject Preference Rule in the Same Sentence	... Sent ₄ [...], ..., Sent ₁ [... (PersPro-pl/Sg),...],... (ReflPro)....	SUB (N/NP-pl/Sg)	
	R4	Personal Pronoun as SUB in the Same Sentence	... Sent ₄ [...], ..., Sent ₁ [...SUB (PersPro-pl/Sg),...],... (ReflPro)....	SUB(PersPro-pl/Sg) → call strong personal anaphora resolution algorithm (Ali et al, 2007)	
	R5	a	Number agreement for Plural ReflPro (خانونه [jānona]-Selves, خپل خانونه [xpəl jānona]-Selves)	... Sent ₅ [... SUB (N ₅ /NP ₅ -pl)...], Sent ₄ [...SUB (N ₄ /NP ₄ -pl)... SUB (N ₃ /NP ₃ -Sg), ..., Sent ₁ [...(N ₂ /NP ₂ -pl)... SUB (N ₁ /NP ₁ -pl) ... (ReflPro)....]	SUB (N ₄ /NP ₄ -pl), SUB (N ₁ /NP ₁ -pl)
		b	Recency Rule (When There are Subjects Either in the Same or Preceding Four Sentences)	SUB (N ₄ /NP ₄ -pl), ... , SUB (N ₁ /NP ₁ -pl) (ReflPro)	SUB (N ₁ /NP ₁ -pl)
c		Recency Rule (When There is no Subject neither in the Same Sentence nor in the Preceding Four Sentences)	... Sent ₅ [... (N ₅ /NP ₅ -pl)...], Sent ₄ [...(N ₄ /NP ₄ -pl)... (N ₃ /NP ₃ -Sg/pl), ..., Sent ₁ [...(N ₂ /NP ₂ -Sg/pl)... (N ₁ /NP ₁ -Sg/pl) ... (ReflPro)....]	(N ₁ /NP ₁ -Sg/pl)	