19

AMBIGUITY RESOLUTION USING PROBABILISTIC CONTEXT FREE GRAMMAR

Aadil A. Kak

Nazima Mehdi

Introduction

Ambiguity is the property of words, terms, notations and concepts (within a particular context) as being undefined, undefinable or without a clear meaning. A word, phrase, sentence or other communication is "ambiguous" if it can be interpreted in more than one way.

There are various types of ambiguities but for this paper we have taken only two types of ambiguities into consideration viz: Lexical and Syntactic.

In Lexical ambiguity a single word may be interpreted to have more than one interpretation meaning. Normally Lexical ambiguity can be resolved when the context/sentence provides extra information leading to the resolution of the ambiguity. E.g. literally speaking, the word 'bank' has two interpretations viz:

Bank 1 →a financial organisation.

Bank 2→a raised bound adjoining a water body such as river or a lake.

This word can be disambiguated only when appropriate context helps in that matter otherwise the ambiguity cannot be resolved. Consider the sentences 'I went to the bank to withdraw some money' and 'I went to the bank to untie my boat' and compare them with 'I went to the bank'.

Lexical ambiguity is a fundamental characteristic of language and is so vital that any theory of language processing come to grips with the process through which ambiguous meanings are processed and can be resolved.

Another variant of lexical ambiguity is when the two meanings belong to different grammatical categories. Consider the example.

1. I saw her duck

Aux⟶ Tns

VP⟶ Verb

b. pick and eat

S⟶NP Aux VP

NP⟶ ϕ

Aux⟶Tns

VP⟶V,V

c. forget about everything and eat

Advp⟶Adv, v

Adv⟶ Adverb

V⟶Verb

More examples are as follows:

3. The jealous husband read the letter= (the alphabet meaning or the postal meaning).

4. The man started to drill before the truck arrived.

Here drill means either (bore a hole with a drill or repeated exercises).

Examples from Kashmiri

5. asi gatshi tse:r

We want apricot /will be late

5a. we want apricot

5b. we will be late

Here 'tse:r' may be a noun meaning 'apricot' or it can act as an adjective 'late' depending upon the word preceding it.

In the above sentence 'gatshi' is also ambiguous having two interpretations either it can be used as 'will be' or as 'want' depending upon the word following it.

6. yi chu tsok

tsok=(sour or bottom).

I.e. either as an adjective or as a noun.

All the examples given above are lexically ambiguous, having two or more senses. Now the question is how such constructions are processed. The study of lexical ambiguity has raised a number of intriguing questions like, do ambiguous words have more than one node or logogen in semantic memory? Or do we consider multiple meanings of ambiguous word when we hear or see the one?

Various researchers like Foss(1970), Caims and Kamer Man (1975) and Hoga Boam and Perfetti (1975) have worked on lexical ambiguity resolution and from their thorough study and research it has been concluded that lexical ambiguity resolution is influenced by a variety of factors including the frequency of a word, its morphological structure , the presence of semantically related words and the existence of alternative meanings of the word. Common words and meanings appear to be in a state of greater readiness than less often words and meanings. Two main approaches to lexical access have been developed: an active search model, which states that we search through lists of words (or meanings) and a logogen model, which claims that a word's spot in the lexicon is activated when sensory and contextual information reach a threshold for that word.

It appears that when the same sequence of sounds has two meanings which fit equally well people activate both and then decide upon the appropriate meaning or suppress the inappropriate meaning according to the context or according to the word following or preceding the ambiguous word. (Example 5)

Structural Ambiguity

Structural ambiguity arises when a sentence or a phrase can be parsed in more than one way. Consider the phrase 'gullible boys and girls' it can mean either 'gullible boys and girls' or 'gullible boys and gullible girls'.

Gullible boys and girls

NP→NP Conj NP

NP→Adj,N

N→ Noun

Conj→ And

NP→Noun

Gullible boys and gullible girls

NP→ Adjp, NP

Adjp→ Adj

NP→ N Conj N

N→ Noun

This ambiguity of interpretation reflects the fact that the phrase 'gullible boys and girls' has two possible internal organisations for the same linear sequence of words. These internal organisations differ as to whether the adjective gullible modifies boys and girls or just boys. The internal organisation of a linear string of words is known as constituent structure. Same sort of structural ambiguity can occur in Kashmiri as well. Consider the example

7.	asIl	ləDkI	tI	ko:ri
	Adj	Noun	Conj	Noun
	Good	boys	and	girls
	'Good boys and girls'			

8.	tsək'	tsu:nTh'	tI	Tang
	adj	noun	Conj	noun
	sour	apples	and	pears
	'Sour apples and pears'			

Structural ambiguity is not limited to the phrases only, it can also occur at the sentential level e.g.

9. Tim saw a man with the binoculars.

The above sentence can be interpreted in two ways viz:

9a. Tim used binoculars to see a man.

S→ NP Aux VP

NP→ Noun

Aux→ Tns

VP→V,NP,PP

NP→Det,N

PP→P,NP

NP→Det,N

9b. A man was using binoculars and Tim saw him.

S→NPAux VP

NP→Noun

Aux→Tns

VP→V,NP

NP→NP,PP

NP→Det,N

PP→P,NP

NP→Det,N

Kashmiri examples

10. me vuch su tamo:k Ceva:n

I saw him hookah drinking

'I saw him smoking a hookah'

Again two senses of the sentence are perceived, one being the hookah smoker is me and the other is that the hookah smoker is the other person.

Regarding the sense of

(10a) A man was smoking and I saw him. The PS rules in Kashmiri will be as follows

S→NPAux VP

NP→Pro

Aux→Tns,Asp

VP→V,S

V→Verb

S→NP,Aux,VP

NP→Pro

Aux→Tns,Asp

VP→NP,V

NP→Noun

V→Verb

And regarding the sense of (10b) 'I was smoking and I saw the man'. The PS rules in Kashmiri will be as follows

S→NP,Aux,VP

NP→Pro

Aux→Tns,Asp

VP→VP,VP

VP→V,NP

V→Verb

NP→Pro

VP→NP,V

NP→Noun

V→Verb

Research indicates that whenever an ambiguous word or sense is encountered people momentarily access it and activate both the senses and then rule out their irrelevant sense and decide upon the relevant one, even if they are not aware of this process going on.

ANALYSIS

The above discussion was all about how humans process ambiguous word or a structure. Now the issue is how a machine can resolve ambiguity both lexical and structural and then how a disambiguated text can be translated from one language to another. At a lower level it is believed that all that is required for machine translation is a bilingual dictionary and rules for reordering words in a sentence. E.g. to translate the following sentence from Kashmiri to English

1.	ra:m-	an	kh'ov	batI
	noun	erg	verb	noun
	raam ate rice			

2.	əs	chi	tsu:nTh	kheva:n
	noun	verb be	noun	verb prog
	we are eating an apple			

Here the Kashmiri words are replaced by English equivalent words and word order is also reordered, (SOV to SVO e.g. 2). But the number of problems comes in the way of machine translation like, choice of

sentence structure, pronoun reference, and noun- noun modification, identification of tense and modality, ambiguous words / sentences. Here we will talk in terms of ambiguity only as the present paper focuses on ambiguity resolution.

For the present paper we have used a probabilistic / stochastic context free grammar which was first proposed by Booth (1969). This paper describes how the probabilistic approach is used to resolve ambiguity.

A probabilistic context free grammar (PCFG) is a context free grammar in which every rule is annotated with the probability of choosing the rule . Each PCFG rule is treated as if it were conditionally independent, thus the probability of a sentence is computed by multiplying the probabilities of each rule ir the parse of a sentence.

For this paper three contexts were analysed.

Context 1

əs' chi dargah basa:n - 2. tati chu akh boD ba:g - 3. ath ba:gas manz chi va:riyah lakIT ləDkI tI ko:ri ginda:n - 4. me chi yim sə:ri: lakIT baCi khara:n - 5. yim sə:ri: lakIT ləDkI tI ko:ri chi ath ba:gas manz sakh So:r kara:n-

TAG:

1. əs'	chi	dargah	basa:n -	2. tati	chu	akh
Pro	verb *be*	noun	verb prog	pro	verb *be*	det

boD	ba:g -	3. ath	ba:gas	manz	chi	va:riyah
adj	noun	det	nuon	pp	verb *be*	deg

lakIT	ləDkI	tI	ko:ri	ginda:n -	4. me	chi
adj	noun	conj	noun	verb	pro	verb *be*

yim	sə:ri:	lakIT	baCi	khara:n -	5. yim	sə:ri:	lakIT
pro	det	adj	noun	verb	pro	det	adj

ləDkI	tI	ko:ri	chi	ath	ba:gas	manz	sakh
noun	conj	noun	verb *be*	det	noun	pp	deg

So:r	kara:n-
adj	verb

The English equivalent of the context is as under

We live in dargah. There is a big field. In this field small boys and girls play. I hate all these small kids. All these small girls and boys make lot of noise in this field.

Context 2

m'o:n	do:s	chu	vana:n	zi	su	chu	va:riyah
pro	noun	verb be	verb	comp	pro	verb be	deg

tamo:k	Ceva:n-	yeli	bI	university	dra:yas
noun	verb	adv	pro	noun	verb

me	vuch	su	tamo:k	Ceva:n-	me
pro	verb	pro	noun	verb	pro

vonmas	tsI	kya	yetinas	kara:n?	asi
verb	pro	Q.word	adv	verb	pro

pazi	jaldi:	ne:run	natI	ne:ri	bas
modal	adv	verb	conj	verb	noun

tI	asi	gatshi	tse:r-

conj pro modal adj

Context 3

me	chu	petras	boD	mevi	ba:g-	As'	gəyi
pro	verb *be*	noun	deg	noun	noun	pro	verb

ra:th	tinhdis	ba:gas	manz	sa:rivly	kheyi	
adv	pro	noun	pp	det	verb	

tsu:nTh'-	me	vonmakh	me	gatshi	tse:r-	me	chi
noun	pro	verb	pro	verb	noun	pro	verb *be*

tse:rl	sakh	khOS	kara:n-	asi	chu	gari
noun	deg	adj	verb	pro	verb *be*	noun

ti	akh	tse:rl	kul	vovmut-	magar	su
conj	det	noun	noun	verb	conj	pro

chu	vlni	va:riyah	lakuT-	me:n'	do:s	ru:d'
verb *be*	adv	deg	adj	pro	noun	verb

me:nis	petrisInde:	me	vonmakh	bl	ne:rl	natl
pro	noun	pro	verb	pro	verb	conj

gatsi	me	tse:r-

modal pro adj/ adv

First context number 1 was taken into consideration. In this context there are 5 sentences, out of these, sentence 3 is ambiguous I.e.it has got two senses but only one sense is appropriate as per the given context. For each and every sentence PSG rules are given as follows:

Sentence 1

S→NP AUX VP

NP→ PRO,N

AUX→TNS,ASP

TNS→PRT

ASP→PROG

VP→NP,V

NP→NOUN

V→VERB

Sentence 2

S→NP Aux VP

NP→ Pro

Aux→Tns

Tns→ Prt

VP→ V,NP

NP→Det, Adjp,N

Adjp→Adj

V→be

Sentence 3

S→PP Aux VP

PP→NP,P

NP→Det,N

Aux→Tns,Asp

Tns→Prt

Asp→ Prog

VP→ NP,V

NP→Adjp.NP

Adjp→Deg, Adj

NP→NP,Conj.NP

V→Verb

NP→Noun

Sentence 4

S→NP,AuxVP

NP→Pro

Aux→Tns,Asp

Tns→ Prt

Asp→ Prog

VP→NP,V

NP→Pro,NP

NP→Det, Adjp,N

Adjp→Adj

V→Verb

N→Noun

Sentence 5

S→NP Aux VP

NP→Pro,NP

NP-→Adjp,NP

Adjp→Det,Adj

NP-→NP,Conj,NP

NP→Noun

Aux→Tns,Asp

Tns-→Prt

Asp→Prog

VP-→PP Adjp,V

PP→NP,P

NP→Det,N

Adjp→Deg,Adj

V→Verb

Taken context into consideration probability was given for each rule as under :

Rules	Count	Total for LHS	MLE /Probability
S→NP,Aux,VP	4	6	0.67
S→PP,Aux,VP	2	6	0.34
Aux→Tns,Asp	5	6	0.84
Aux→Tns	1	6	0.17
NP→Pro	3	22	0.14
NP→Det,N	3	22	0.14
NP→Det,Adjp,N	2	22	0.91
NP→Adjp,NP	2	22	0.91
NP→Pro,NP	1	22	0.046
NP→Noun	6	22	0.28

NP→NP,Conj,NP	3	22	0.14
NP→Adjp,	3	3	1
VP→V,NP	1	6	0 .17
V→Verb	5	6	0.84
VP→NP,	1	6	0.17
V→Aux verb be	1	6	0.17
Adjp→Deg,Adj	3	6	0.5
Adjp→Adj	2	6	0.34

Then the above given probability was used for both the senses of the ambiguous sentence 3 in the context 1 for ambiguity resolution to check which sense is more appropriate for a given context

PS rules for the sentence 3are as follows:

Sentence3

ath	ba:gas	manz	chi	va:riyah	lakIT	ləDkI
det	noun	pp	aux verb	deg	adj	noun
this	field	in	are	very	small	boys

tI	ko:ri	ginda:n
conj	noun	verb
and	girls	play

In this field various small boys and girls play

3a

Small boys and small girls

Rules	MLE
S→PP,Aux,VP	0.34
PP→NP,P	1

NP→Det,N	0.14
Aux→Tns,Asp	0.84
VP→NP,V	0.67
NP→Adjp,NP	0.091
Adjp→Deg,Adj	0.5
NP→NP,conj,NP	0.14
NP→Noun	0.28
V→Verb	0.84

3b

Only boys are small not girls .

Rules	MLE
S→PP,Aux,VP	0.34
PP→NP,P	1
NP→Det,N	0.14
Aux→Tns,Asp	0.84
VP→NP,V	0.67
NP→NP,Conj,NP	0.14
NP→Adjp,N	0.046
Adjp→Deg,Adj	0.5
NP→N	0.28
V→Verb	0.84

The probability of each of the structure can be computed by multiplying together each of the rules used in the derivation. I.e. the probability of 3a and 3b can be computed as follows:

3a

0.34*1* 0.14* 0.84*0.67*0.091*0.5*0.14*0.84*0.28=0.0000401

3b

0.34*1*0.14*0.84*0.14*0.67*0.046*0.5*0.28*0.84=0.0000202

We can see that 3a has a higher probability. Based on this it can be chosen as the more appropriate sense for the context 1. And the same sentence can be chosen by the disambiguation algorithm because it selects the parse with the highest PCFG probability.

Conclusion

The present paper focuses on Probabilistic Context Free Grammar to resolve the ambiguity for Kashmiri annotated data. And also an algorithm is proposed for filtering out the least improbable candidate and thus reducing the level of ambiguity. The solution assumes annotated data and applies PCFG formalism given by Booth (1969) to compute constituent probabilities of a parse tree to obtain the total probability of a sentence. The winning candidate will be with the highest probability. We have tested this on a significant data/ contexts and it worked fine with the simple sentences only. This study is in a preliminary stage and in future we will test this with other languages also.

Reference

Bharti, A., Chatanya, V. and Sangal, R. 1995. *Natural Language Processing: A Paninian Perspective.* New Delhi: Prentice Hall.

Booth,T.L. 1969. Probabilistic Representation for Formal Languages.*In IEEE Conference Record Of the 1969 Ten Annual Symposium on Switching and Automata Theory*, pp.74-81.

Cairns,H.S.and Kamerman,J. 1975. Lexical Information Processing during Sentence Comprehension.*Journal of Verbal Learning and Verbal behaviour*,14,170-179.

Chomsky, N. 1965. *Aspects of the theory of sntax.* Cambridge, Mass: MITPress.

Dorr,B. 1993. *Machine Translation: A view from the lexicon.* Cambridge, Mass:MIT Press.

Foss, D. 1970. Some effects of Ambiguity upon Sentence Comprehension. *Journal of Verbal Learning and Verbal Behaviour*, 9,699-706.

Harnish,R.M. 2002. *Minds.Brains,Computers: A Historical Introduction to the Foundation of Cognitive Science.* Malden,Massachussets,U.S.A: Blackwell publishers.

Hogaboam,T.W.and Perfetti,C.A. 1975. Lexical Ambiguity and Sentence Comprehension.*Journal of Verbal Learning and Verbal Behaviour*,14,265-274.

Jurafsky,Daniel.and Martin,H.James. 2000. *Speech and Language Processing: An Introduction to natural Language Processing, Computational Linguistics, and Speech Recognition.* UK: Printice Hall InternationalLimited.

Kak, A.A. and Talashi, R. 2002. *Kashmiri Language: A Grammatical Description.* Nihar Publications: Jammu.

Miller,A.Goerge and Jhonson-Laird,N.Philip. 1976. *Language and Perception.* London: Cambridge University Press.

Wali,K. and O.N. Koul. 1997. *A Cognitive- Descriptive Grammar.* London: Routledge.