

Hindi Generation from Interlingua (UNL)

Om P. Damani, IIT Bombay
(Joint work with S. Singh, M. Dalal,
V. Vachhani, P. Bhattacharya)

Cross Lingual Information Access: Language Independent Representation



AgroExplorer
a meaning based multilingual search engine

Meaning Concept

moneylenders exploit farmers Search

Found: 1 results



1. INDIAN AGRICULTURE

- The farmers are still being exploited by moneylenders who provide finance at exorbitant interest rates and there are cartels of traders who pay very little for their produce even in the well recognized mandies in the country.

View in: [Original](#) [Hindi](#) [Marathi](#) [UNL](#)

Resources | Recommend a site | Help

What is Agro-Explorer? CSE, IITB | Media Lab Asia

Best Viewed at 800 * 600 | IE 5 & above



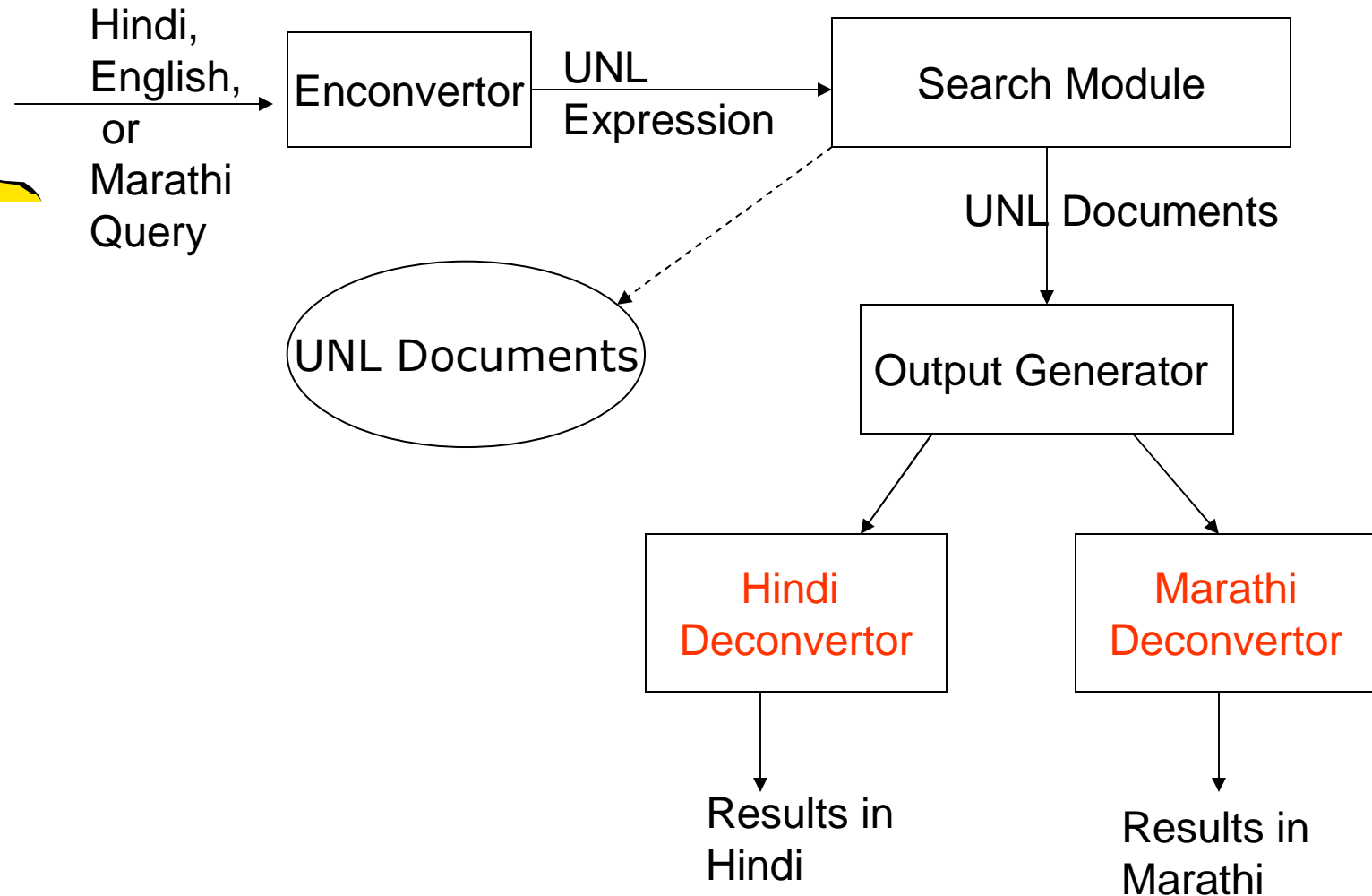
Reference with Encarta



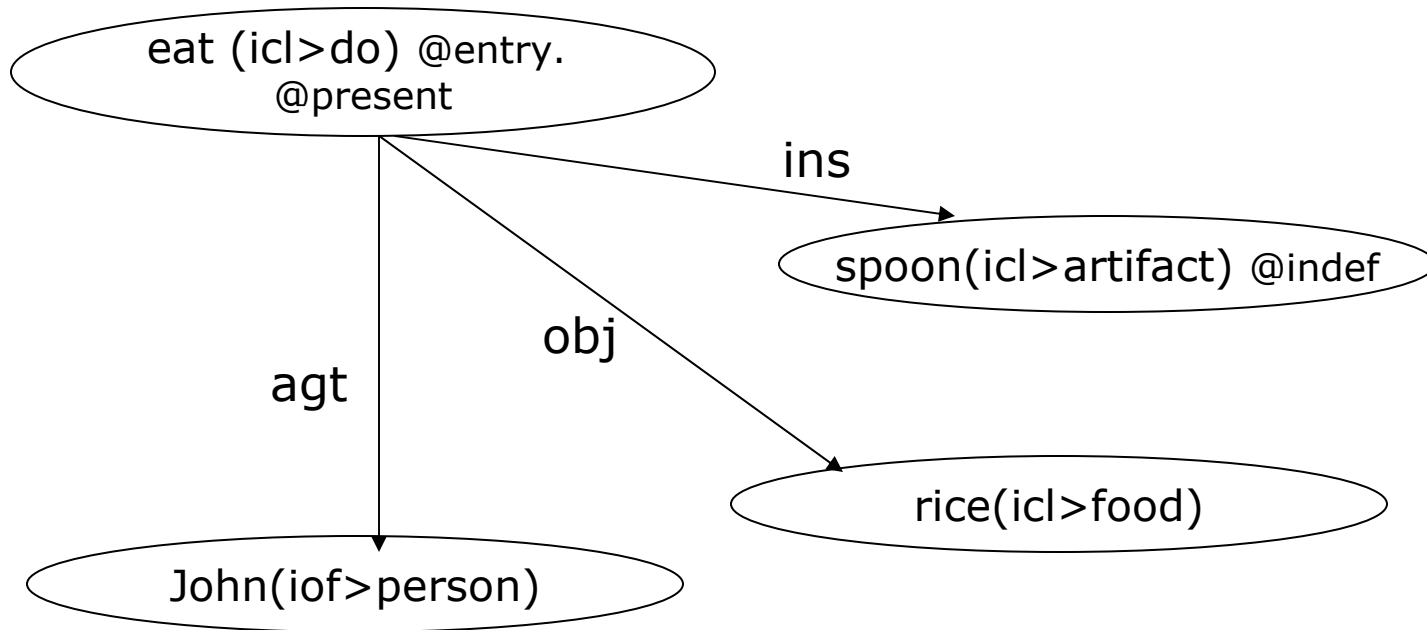
उत्पाद, पैदावार, और विभिन्न फसलों के लिए नरिधारति भूमि की वृद्धि की दृष्टि से भारत ने स्वतंत्रता के बाद से कृषि में काफी उन्नत की है। वह हरति करांत, श्वेत करांत, पीत करांत और नील करांत से गुजर चुका है। आज विश्व में भारत दूध, फल, काजू, नारियल और चाय का सबसे बड़ा, गेहूँ, सब्जियों, शक्कर और मछली का दूसरा सबसे बड़ा और तम्बाकू और चावल का तीसरा सबसे बड़ा उत्पादक है। जनसंख्या में ३५ करोड़ से ९५ करोड़ की वृद्धि के बावजूद देश में खाद्यान्न की परतियकृत उपलब्धता १९५१ में ३५० ग्राम से अब लगभग ५०० ग्राम परतदिनि, दूध की १२५ ग्राम से २१० ग्राम परतदिनि और अंडों की ५ से ३० परत वर्ष बढ़ गयी है। फ्रि भी अगर हमारे योजनाकारों ने स्वतंत्रता के प्रारंभिक वर्षों में भारत के उस विकास को अपेक्षति महत्व दिया होता जिसका वह पात्र था, तो कृषीय उत्पादन और आर्थिक विकास कई दृष्टि से आज वह और भी बेहतर स्थिति में होता। आज भी भारत में कृषक बैंकों से अपनी आवश्यकताओं का सरिफ़ १५% कृषि ऋण पाने में समर्थ है। विभिन्न राज्य बीजनगिम हमारे किसानों द्वारा अपेक्षति बीजों का मात्र १०% ही उत्पादन कर पाते हैं। फ्रलिहाल, मात्र २३.३०% कृषक विभिन्न सरकारी एजेन्सियों द्वारा प्रदत्त वसितार सेवाओं का थोड़ा बहुत लाभ उठा पाते हैं और हर साल यत्न रख रखाव, गरि जाने, वाद, सूखे, कीड़ों और रोगों के कारण लगभग २०% फसल का नुकसान होता है। फलों और सब्जियों में नुकसान लगभग ३०% है। कृषक अमी भी ऐसे महाजनों द्वारा शोषति किये जाते हैं जो बेहिसाब बयाज दर पर वसित प्रदान करते हैं और वयपारियों के ऐसे उत्पादक संघ हैं जो देश की संप्रतषितति मंडियों में भी उनके उत्पादों की बहुत कम कीमत देते हैं।

कृषि में कुल पूँजी वरिचन

System Architecture for Achieving Cross Lingual Information Access

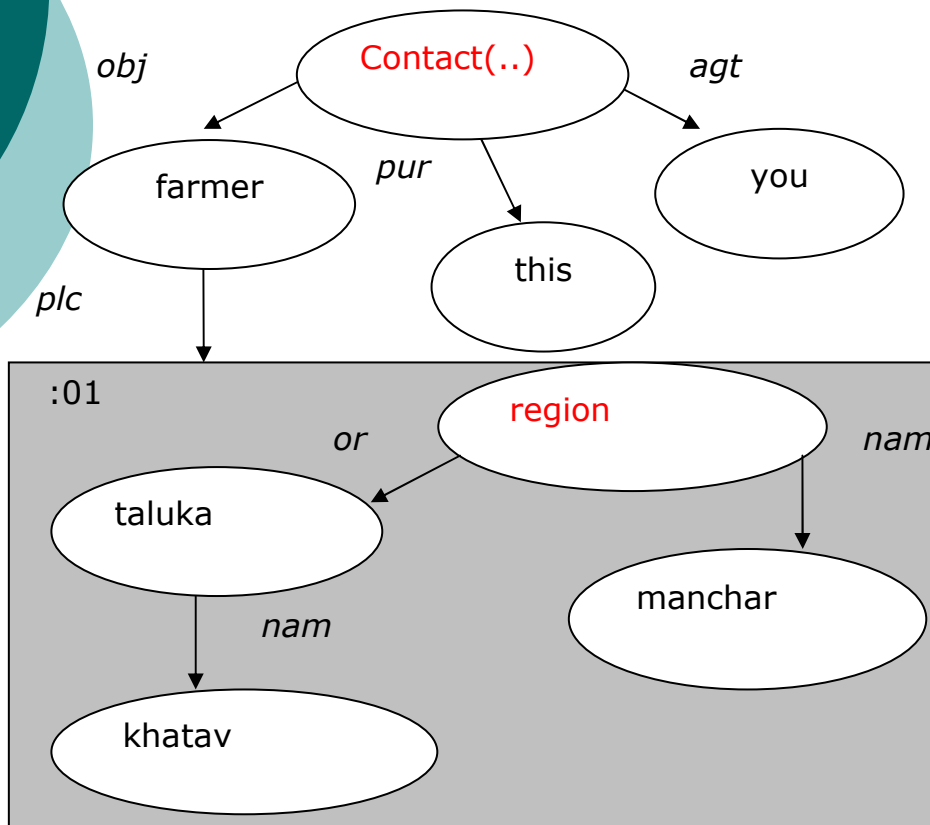


Universal Networking Language (UNL)



```
[UNL:1]
agt(eat(icl>do).@entry.@present, John(iof>person))
obj(eat(icl>do).@entry.@present, rice(icl>food))
ins(eat(icl>do).@entry.@present, spoon(icl>artifact))
[\\UNL]
```

UNL Scopes: Representing Embeddings



```

obj(contact(icl>communicate(agt>person,
  obj>person)):0W.@imperative.@entry
  ,farmer(icl>creator):1T.@pl.@def)
pur(contact(icl>communicate(agt>person,
  obj>person)):0W.@imperative.@entry
  ,this:04)
agt(contact(icl>communicate(agt>person,
  obj>person)):0W.@imperative.@entry
  ,you(icl>persons):0J)
plc(farmer(icl>creator):1T.@pl.@def,:01)
or:01(region(icl>location):38.@entry,talu
  ka(icl>geographicalarea):4A)
nam:01(region(icl>location):38.@entry,M
  anchar(icl>geographical place):2R)
nam:01(taluka(icl>geographical
  area):4A,Khatav(icl>geographical
  area):3U)
  
```

For this, you contact the farmers of [Manchar region](#) or of [Khatav taluka](#)

HinD (Hindi Deconverter): Yet Another Deconversion Engine



Output Quality

Manati

Deco

HinD

	agt	aoj	obj	ins
Agt		L	L	L
Aoj			L	L
Obj				R
Ins				

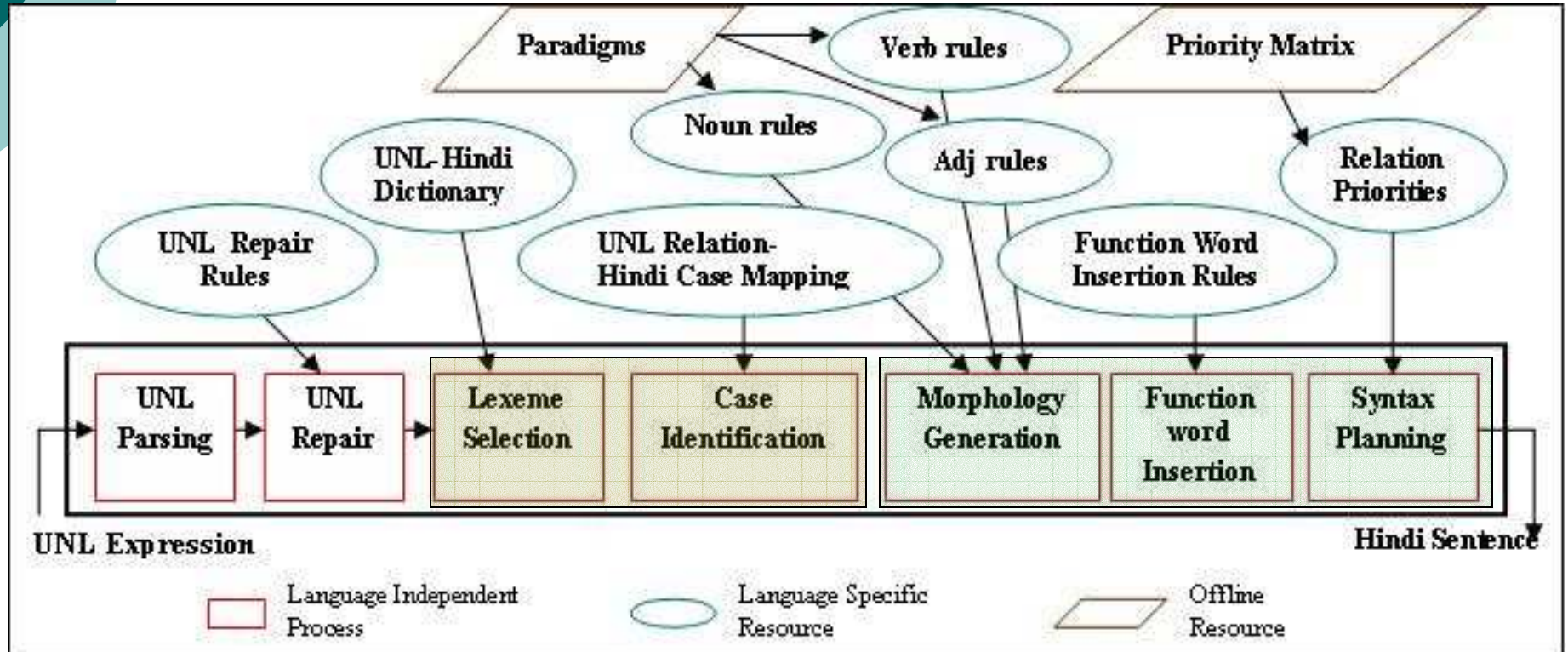
Suffix	Attribute values
<i>uM</i>	@N@N,@M@l,@blique
<i>U</i>	@N@N,@M@sg@blique
<i>I</i>	@N@N,@F,@sg,@blique
<i>iyM</i>	@N@N,@F,@l,@blique
<i>aM</i>	@N@N,@NOICH@F,@l,@blique

Suffix	Tense	Aspect	Mood	N	Gen	P	V E
-e rahaa thaa	@past	@progress	-	@sg	@male	3 rd	e
-taa hai	@present	@custom	-	@sg	@male	3 rd	-
-iyaa thaa	@past	@complete	-	@sg	@male	3 rd	I
saktii hain	@present	-	@ability	@pl	@female	3 rd	A

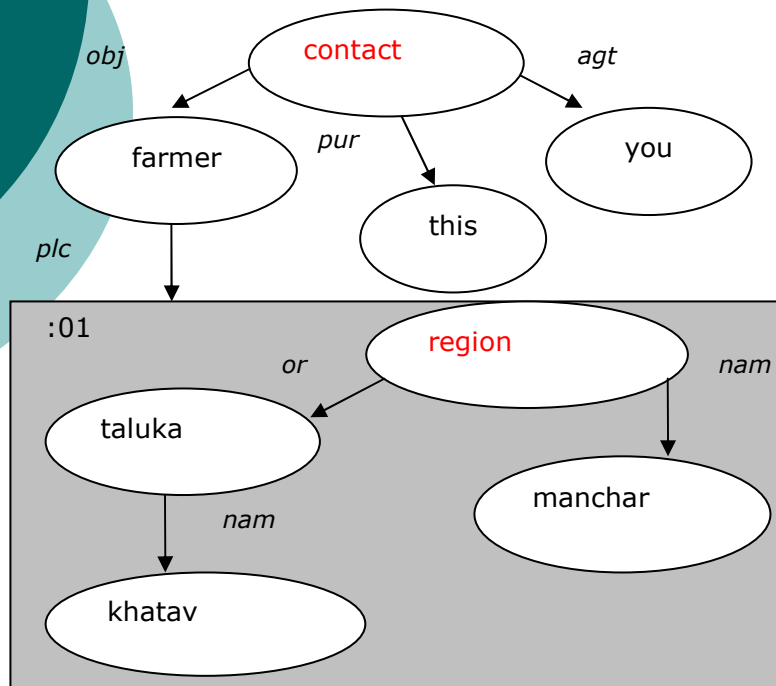
Simplicity of Specification

HinD Architecture

Deconversion = Transfer + Generation



Step-through the Deconverter



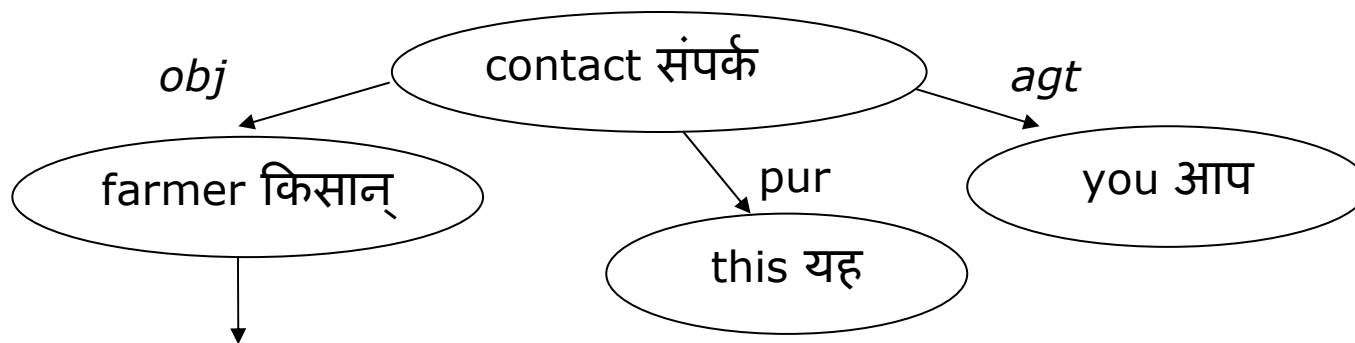
Module	Output
UNL Expression	obj(contact(...
Lexeme Selection	संप्रक कसिन् यह आप क्षेत्र तलुक मंचर खटव contact farmer this you region taluka manchar khatav
Case Identification	संप्रककसिन्* यह आप क्षेत्र* तलुक* मंचर खटव contact farmer* this you region* taluka* manchar khatav
Morphology Generation	संप्रक कीजिए कसिनें इस आप क्षेत्र contact .@imperative farmer.@pl this you region तलुके मंचर खटव taluka manchar Khatav
Function Word Insertion	संप्रककीजिए कसिनें के इसके लिए आप क्षेत्र contact farmers this for you region य तलुके के मंचर खटव or taluka of Manchar Khatav
Linearization	इसके लिए आप मंचर क्षेत्र य खटव This for you manchar region or khatav तलुके के कसिनें के संप्रककीजिए taluka of farmers contact

Lexeme Selection

[सं पर क]{ } "contact(icl>communicate(agt>person,obj>person))"
(V,VOA,VOA-ACT,VOA-COMM,VLTN,TMP,CJNCT,N-V,link,Va)

[पहचा न का व् यक् त्ति]{ } "contact(icl>representative)"
(N,ANIMT,FAUNA,MML,PRSN,Na)

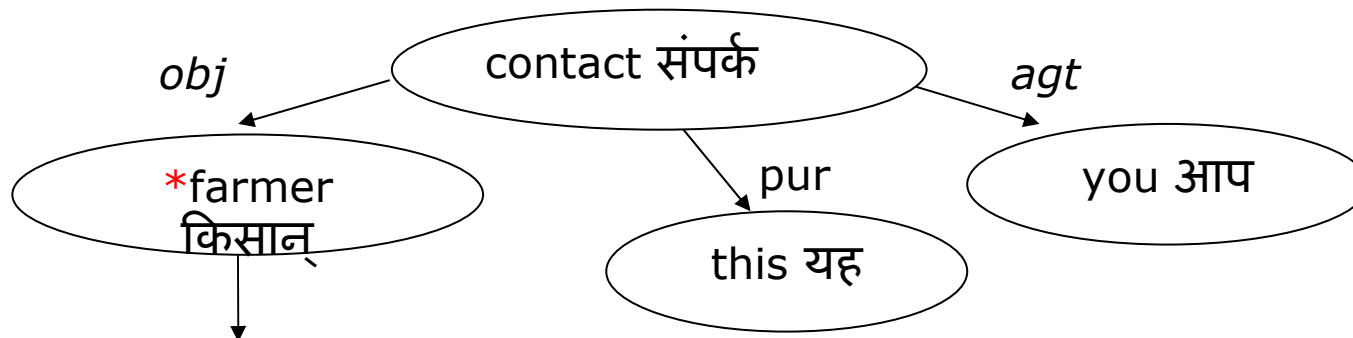
Lexical Choice is unambiguous



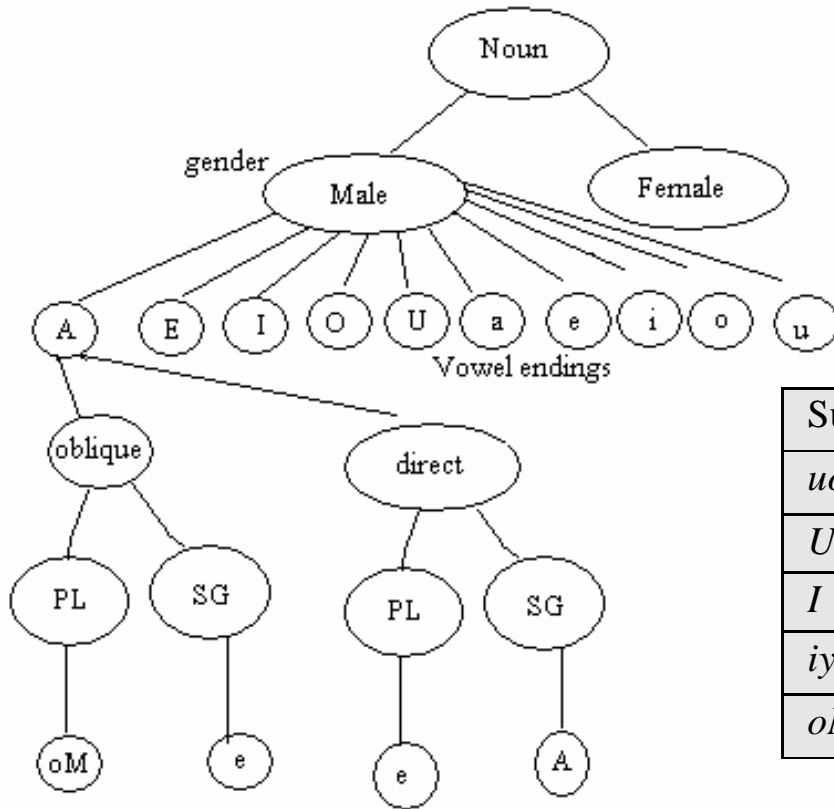
Case Marking

Relation	Parent+	Parent-	Child+	Child-
Obj	V	VINT	N	
Agt	V	@past	N	

- Depends on UNL Relation and the properties of the nodes
- Case get transferred from head to modifiers



Morphology Generation: Nouns



The boy saw me.

लड़के ने मुझे देखा ।

Boys saw me.

लड़कों ने मुझे देखा ।

The King saw me.

राजा ने मुझे देखा ।

Kings saw me.

राजाओं ने मुझे देखा ।

Suffix	Attribute values
<i>uoM</i>	@N, @NU, @M, @pl, @oblique
<i>U</i>	@N, @NU, @M, @sg, @oblique
<i>I</i>	@N, @NI, @F, @sg, @oblique
<i>iyom</i>	@N, @NI, @F, @pl, @oblique
<i>oM</i>	@N, @NA, @NOTCH, @F, @pl, @oblique

Verb Morphology

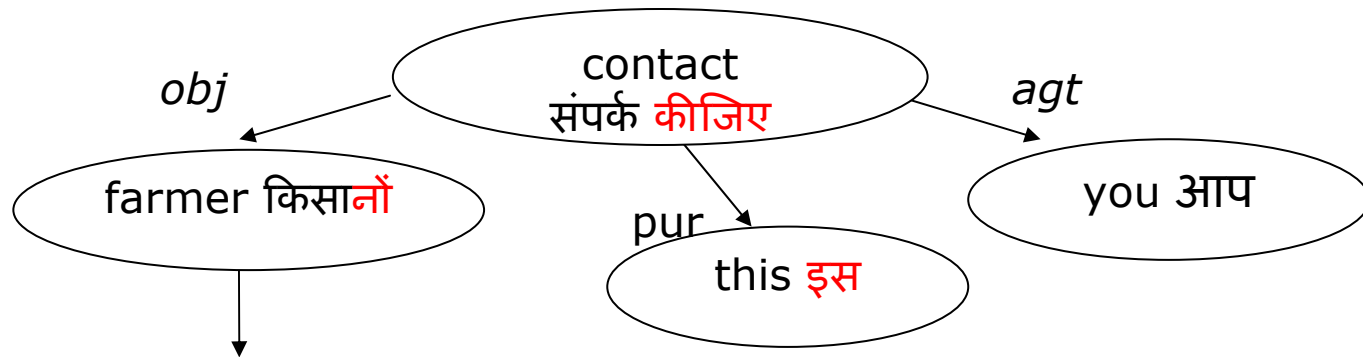
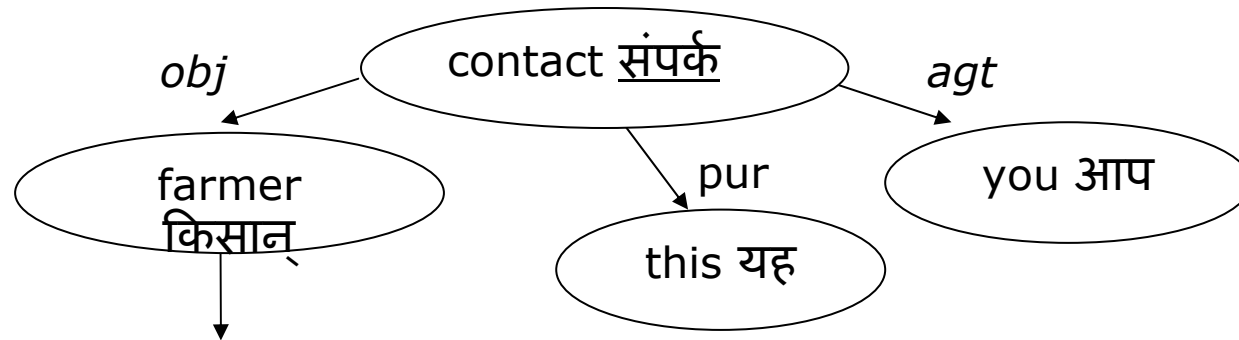
<i>Suffix</i>	<i>Tense</i>	<i>Aspect</i>	<i>Mood</i>	<i>N</i>	<i>Gen</i>	<i>P</i>	<i>V</i> <i>E</i>
-e rahaa thaa	@past	@progress	-	@sg	@male	3 rd	e
-taa hai	@present	@custom	-	@sg	@male	3 rd	-
-iyaa thaa	@past	@complete	-	@sg	@male	3 rd	I
saktii hain	@present	-	@ability	@pl	@female	3 rd	A

लड़का किताब दे रहा था

(The boy was giving the book)

Suffixes mentioned in the rules get attached to the stems selected from the Stem Dictionary

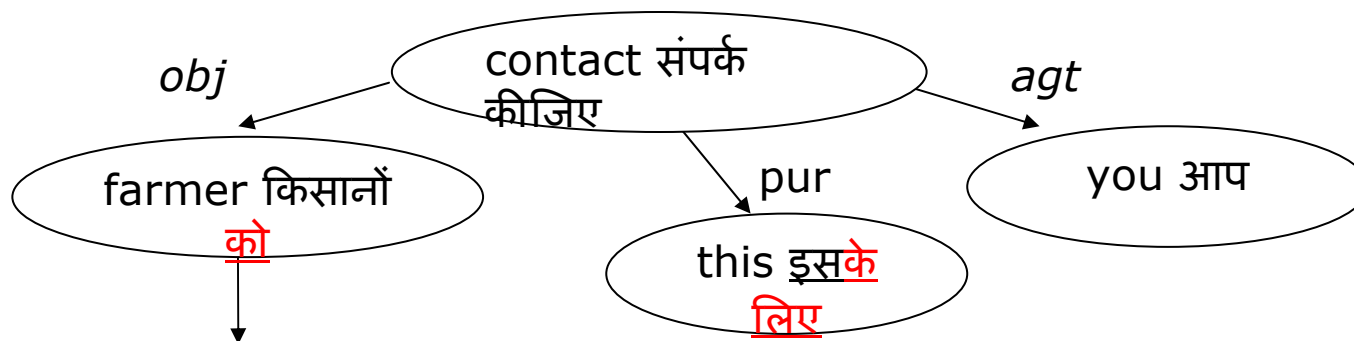
After Morphology Generation



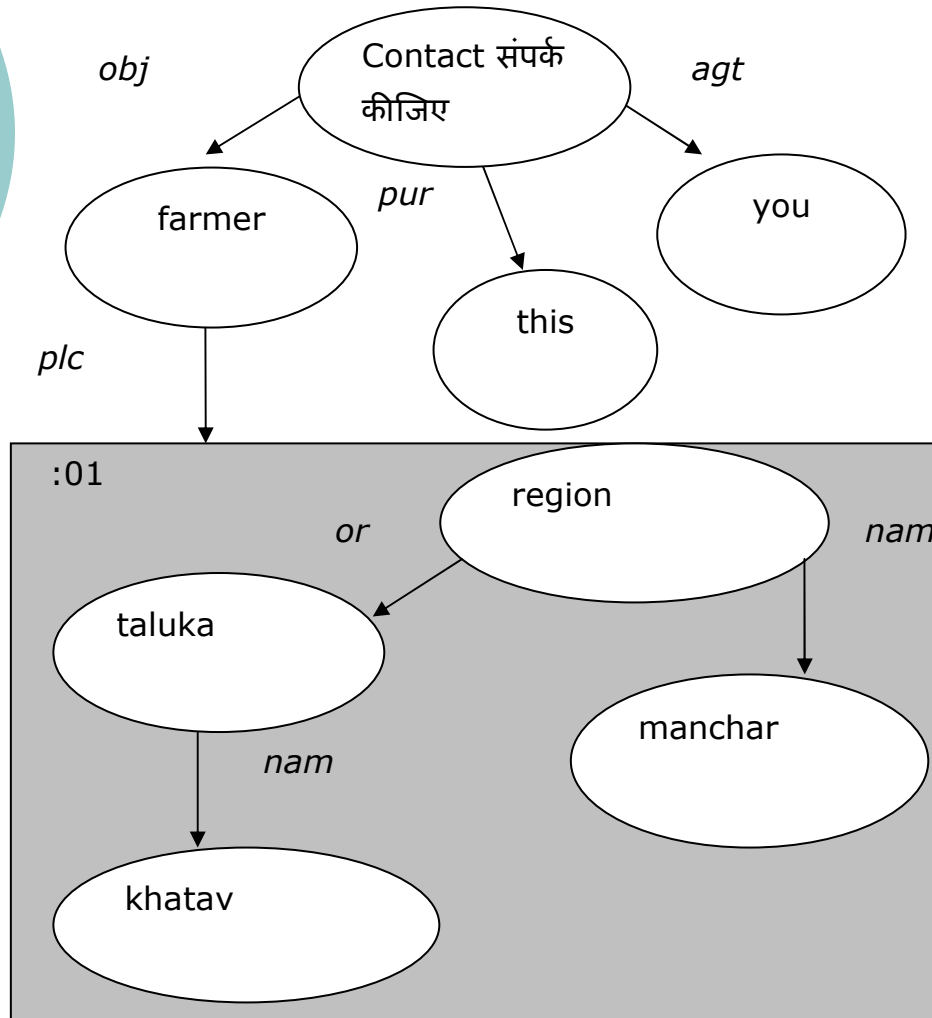
Function Word Insertion

संपर्क कीजिए किसानों यह आप क्षेत्र तालुके मंचर खटाव
संपर्क कीजिए किसानों को इसके लिए आप क्षेत्र या तालुके के मंचर खटाव

Rel	Par+	Par-	Chi+	Chi-	Ch/FW
Obj	V	VINT	N#ANIMT	@topic	<u>को</u>



Linearization



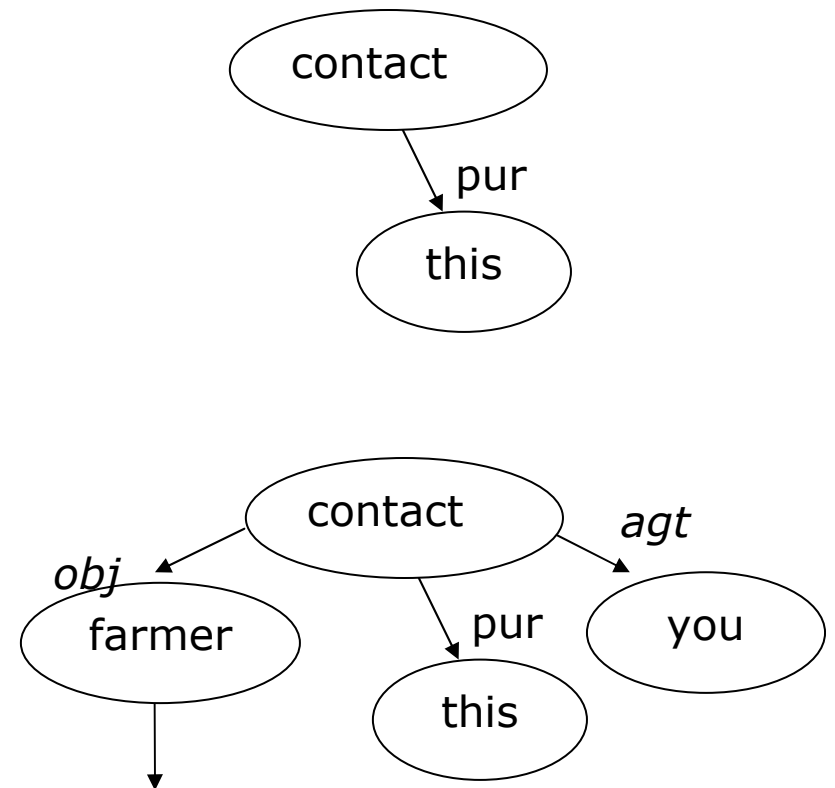
इस आप मंचर क्षेत्र
This you manchar region

खटाव तालुके किसानों
khatav taluka farmers

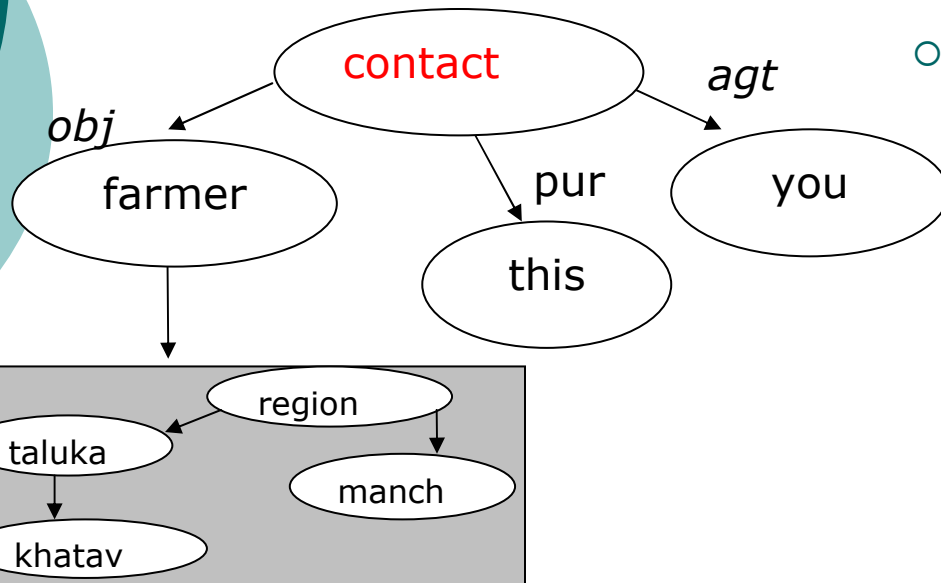
संपर्क
contact

Syntax Planning: Assumptions

- The relative word order of a UNL relation's relata does not depend on:
 - Semantic Independence: the semantic properties of the relata.
 - Context Independence: the rest of the expression.
- The relative word order of various relations sharing a relatum does not depend on
 - Local Ordering: the rest of the expression.

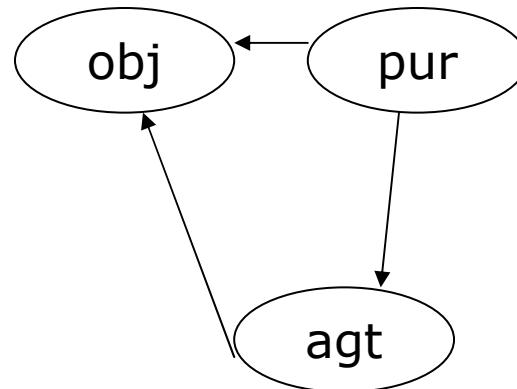


Syntax Planning: Strategy



- Divide a nodes relations in
 Before_set = {obj,pur,agt}
 After_set = {}

	agt	aoj	obj	ins
Agt		L	L	L
Aoj			L	L
Obj				R
Ins				

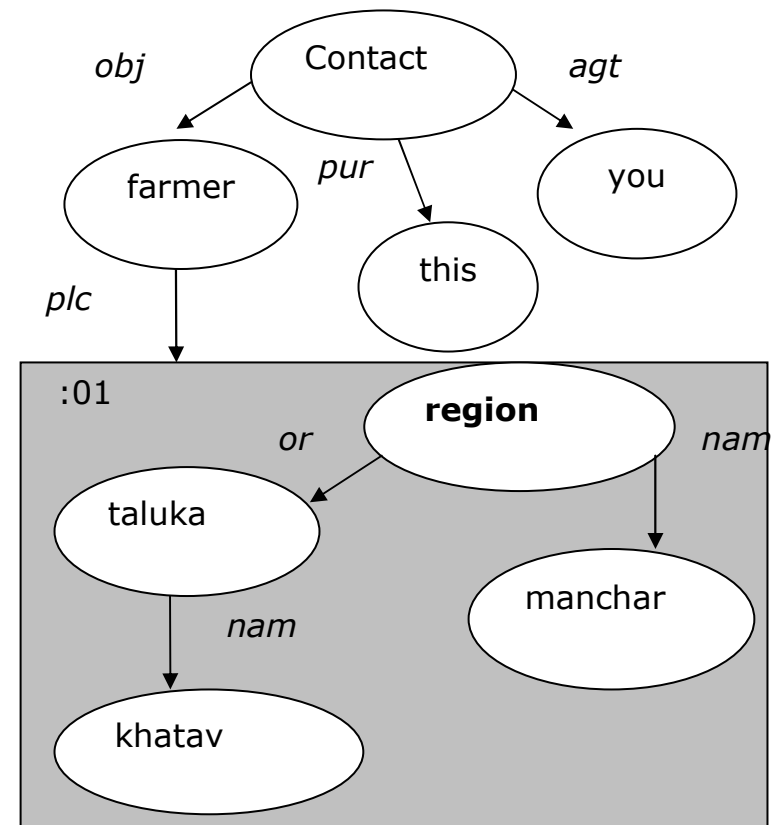


Topo Sort each group:
 pur agt obj
 this you farmer

Final order: this you farmer contact

Syntax Planning Algo

Stack	Before Current	After Current	Output
region farmer contact			this you
	manchar	taluka	
manchar region taluka farmer contact			
region taluka farmer Contact			this you manchar
taluka farmer contact			this you manchar region



All Together (UNL -> Hindi)

Module	Output
UNL Expression	obj(contact(..
Lexeme Selection	संपर्क किसान् यह आप क्षेत्र तालुक् मंचर खटाव contact farmer this you region taluka manchar khatav
Case Identification	संपर्क किसान्* यह आप क्षेत्र* तालुक्* मंचर खटाव contact farmer* this you region* taluka* manchar khatav
Morphology Generation	<u>संपर्क कीजिए</u> किसानों यह आप क्षेत्र contact.@imperative farmer.@pl this you region तालुके मंचर खटाव taluka manchar Khatav
Function Word Insertion	<u>संपर्क कीजिए</u> किसानों <u>को</u> <u>इसके लिए</u> आप क्षेत्र contact farmers this for you region <u>या</u> <u>तालुके के</u> मंचर खटाव or taluka of Manchar Khatav
Syntax Linearization	इसके लिए आप मंचर क्षेत्र या खटाव This for you manchar region or khatav तालुके के किसानों को संपर्क कीजिए taluka of farmers contact



How to Evaluate UNL Deconversion

- UNL -> Hindi
- Reference Translation needs to be generated from UNL
 - Needs expertise
- Compromise: Generate reference translation from original English sentence from which UNL was generated
 - Works if you assume that UNL generation was perfect
- Note that fidelity is not an issue



Input Detail

- Marathi -> English -> UNL
- 901 Sentences from Agriculture Domain
- English: Median length 14 word, std dev 7.5
- Hindi Reference translation generated from English



Manual Evaluation Guidelines

Fluency of the given translation is:

- (4) Perfect: Good grammar*
- (3) Fair: Easy-to-understand but flawed grammar*
- (2) Acceptable: Broken - understandable with effort*
- (1) Nonsense: Incomprehensible*

Adequacy: How much meaning of the reference sentence is conveyed in the translation:

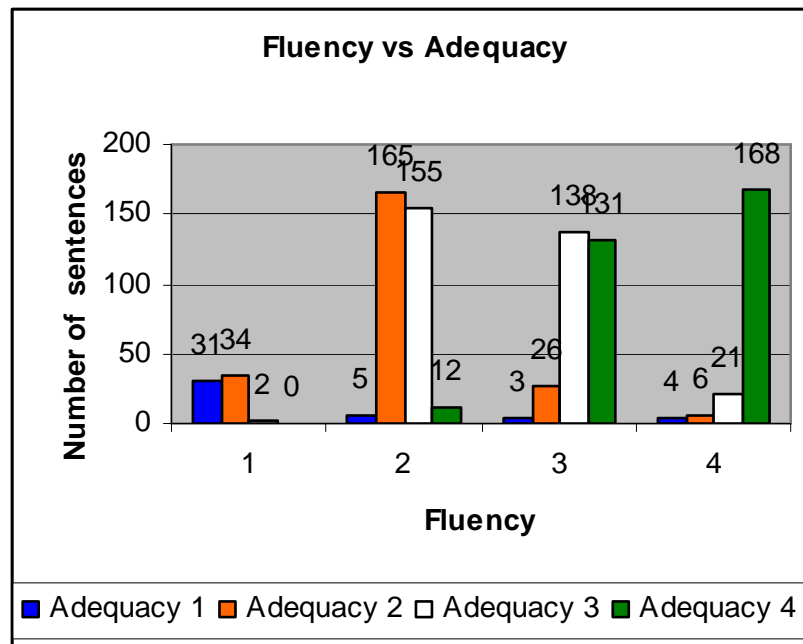
- (4) All: No loss of meaning*
- (3) Most: Most of the meaning is conveyed*
- (2) Some: Some of the meaning is conveyed*
- (1) None: Hardly any meaning is conveyed*

Sample Translations

Hindi Output	Fluency	Adequacy
कवक के कारण आम की नाजुक पत्तियां यदि झुलसे रहे हैं 0.5 प्रतिशत का बोर्पो मिश्रण 10 लीटर पानी के साथ तो पिड़का जाना चािहए	2	3
परीक्षण के अनुरूप नियमित रूप से फलों की अच्छी वृद्धि के लिए खादों की खुराकें दी जानी चाहिए	3	4
हमें मेथी का फसल के बाद बोता ह। मेथी और धनिया फसल को अच्छे बढ़ने या अच्छे बढ़ने बता	1	1
जीवाण्विक संक्रमण से इसकी जड़ें प्रभावित होती हैं	4	4
इमु का पक्षी रेटाइट का परिवार को संबंधित होता ह।और थोड़ा यह शतुरमुर्ग के साथ समान दिखती हैं	2	3

Results

	<i>BLEU</i>	Fluency	Adequacy
Geometric Average	0.34	2.54	2.84
Arithmetic Average	0.41	2.71	3.00
Standard Deviation	0.25	0.89	0.89
Pearson Cor. <i>BLEU</i>	1.00	0.59	0.50
Pearson Cor. Fluency	0.59	1.00	0.68



- Good Correlation between Fluency and BLEU
- Strong Correlation between Fluency and Adequacy
- Can do large scale evaluation using Fluency alone

Caution: Domain diversity,
Speaker diversity



Future Work

- Extend to Marathi and other languages
- Test in domains other than agriculture



Thanks !!