

# TOWARDS A PSYCHOLINGUISTICALLY MOTIVATED DEPENDENCY GRAMMAR FOR HINDI

Samar Husain<sup>†</sup>, Rajesh Bhatt<sup>§</sup>, Shravan Vasishth<sup>†</sup>

<sup>†</sup>Universität Potsdam, Germany; <sup>§</sup>University of Massachusetts, Amherst, USA

August 29, 2013

# Motivation

- The overall goal of our work is to build a dependency grammar-based human sentence processor for Hindi.
- As a first step towards this end, we present a dependency grammar that is motivated by psycholinguistic concerns.

- Introduction
- Relevant experimental work
- Grammar induction
- Processing concerns
- (An outline of a) dependency-based human sentence processor
- Issues and challenges

# Introduction

- Most of the human sentence processing proposals and modeling work generally employ a constituent-based representation
  - ▶ mainly because of how modern linguistics has evolved
- Dependency grammar has been quite popular in computational linguistics
- Related to lexicalized grammars such as LTAG, CCG, etc. (eg. Kuhlmann (2007))
  - ▶ Categorical Grammar: to handle processing of empty categories (Pickering & Barry, 1991)
  - ▶ Dependency categorial grammar: processing both local and non-local dependencies (Pickering, 1994)
  - ▶ LTAG: (Kim, Srinivas, & Trueswell, 1998), Demberg (2010) has proposed a psycholinguistically motivated P-LTAG

# Introduction

- Dependency-based paradigm remains mostly unexplored in psycholinguistics
  - ▶ To our knowledge, the work of Boston and colleagues (Boston et al., 2011), (Boston et al., 2008) is the only such attempt.
- Can a processing model based on dependency paradigm account for classic psycholinguistic phenomena?
- Can one adapt a high performance dependency parser for psycholinguistic research? If yes, then how?
- How will the differences in different dependency parsing paradigms affect the predictive capacity of the models based on them?
- ...

# Relevant experimental work

- Locality constraints
- Expectation-based/Predictive sentence processing
- Processing word order variation

# Background: Hindi

- One of the official languages of India, an Indo-European language
- Free-word order, head final, relatively rich morphology
- Agreement: Verb-Subject, Noun-Adjective

- (1)
- a. malaya ne abhiisheka ko kitaaba dii  
Malaya ERG Abhishek DAT book gave  
'Malaya gave a book to Abhishek.' (S-IO-O-V)
  - b. malaya ne kitaaba abhiisheka ko dii (S-O-IO-V)
  - c. abhiisheka ko malaya ne kitaaba dii (IO-S-O-V)
  - d. abhiisheka ko kitaaba malaya ne dii (IO-O-S-V)
  - e. kitaaba abhiisheka ko malaya ne dii (O-IO-S-V)
  - f. kitaaba malaya ne abhiisheka ko dii (O-S-IO-V)

## Relevant experimental work: Locality constraints

- A great deal of experimental research has shown that working-memory limitations play a major role in sentence comprehension difficulty (e.g., (Lewis & Vasishth, 2005), (Gibson, 2000))

## Need for temporary (working) memory

- Sentence processing is immediate and incremental
- Some words must be retained for future processing

(2) John gave Emma a gift that she liked very much.

- like → gift
- *gift* must be interpreted as an object of *liked*, but in order to make such an interpretation, *gift* must be retained in memory until *like* is encountered
- Such memory usage is not unique to the above sentence, but is commonplace.

## Dependency locality theory (Gibson, 2000)

- (3) a. (DISTANCE = 1) The administrator who the nurse<sub>1</sub> *supervised* scolded the medic while. . .
- b. (DISTANCE = 2) The administrator who the nurse<sub>1</sub> from the clinic<sub>2</sub> *supervised* scolded the medic while. . .
- c. (DISTANCE = 3) The administrator who the nurse<sub>1</sub> who was<sub>2</sub> from the clinic<sub>3</sub> *supervised* scolded the medic while. . .

Reading time at *supervised* should be a function of distance

# Relevant experimental work: Expectation-based/Predictive processing

- Considerable empirical evidence for predictive parsing (eg. Konieczny (2000), Staub and Clifton (2006), Kamide, Scheepers, and Altmann (2003))
- Different accounts of how this unfolds:
  - ▶ (Konieczny, 2000), (Grodner & Gibson, 2005), (Vasishth & Lewis, 2006), (Levy, 2008), (also see, Demberg (2010))

## Empirical evidence: Antilocality effects (Konieczny, 2000)

(4) a. (DISTANCE = 2)

Er hat das Buch, das Lisa gestern gekauft hatte, hingelegt  
He has the book, that Lisa yesterday bought had, laid down  
'The has laid down the book that Lisa had bought yesterday

b. (DISTANCE = 0)

Er hat das Buch hingelegt, das Lisa gestern gekauft hatte  
He has the book laid down, that Lisa yesterday bought had  
'He has laid down the book that Lisa had bought yesterday'

Reading time at *hingelegt* are faster when distance is more

# Argument structure order processing

- Processing word order variation is costly (Hyona & Hujanen, 1997), (Bader & Meng, 1999), (Kaiser & Trueswell, 2004), (Sekerina, 2003), (Vasishth, 2004)
- Processing costs could be due to variety of reasons (such as, syntactic complexity, frequency, information structure, prosody, memory constraints, etc).

- Processing cost due to:
  - ▶ Locality constraints → resource limitation
  - ▶ Expectation-based/Predictive processing → predictive parsing
  - ▶ Processing word order variation → predictive parsing

- ✓ Introduction
- ✓ Relevant experimental work
  - Grammar induction
  - Processing concerns
  - (An outline of a) dependency-based human sentence processor
  - Issues and challenges

# Grammar induction

- The task of automatic grammar induction from a treebank can be thought of as making explicit the implicit grammar present in the treebank
- Can be beneficial for a variety of tasks, such as, complementing traditional hand-written grammars, comparing grammars of different languages, building parsers, etc. (Xia, 2001), (Kolachina et al., 2010)
- Our task is much more focussed: we want to bootstrap a grammar from a Hindi dependency treebank (Bhatt et al., 2009) that can be used for a dependency-based human sentence processor for Hindi.

# Inducing a (psycholinguistically motivated) dependency grammar for Hindi from a dependency treebank

- Main components:
  - ▶ Lexicon,
  - ▶ Frame variations,
  - ▶ Probability distribution of dependency types,
  - ▶ Prediction rules

- Syntactic properties of various heads (e.g. verbs)
- Apriori selection of the argument relations in the Hindi dependency treebank
  1. Subject
  2. Object
  3. Indirect object
  4. Experiencer verb subject
  5. Goal
  6. Noun complements of subject (for copula verbs)

# Lexicon: Verbal heads

- Based on these we formed around 13 clusters
- These clusters were then merged into 6 super-clusters based on the previously mentioned relations (this time acting as discriminators)
- These super-clusters correspond to:
  1. Intransitive verbs (e.g. *so* 'sleep', *gira* 'fall')
  2. Transitive verbs (e.g. *khaa* 'eat')
  3. Ditransitives (e.g. *de* 'give')
  4. Experiencer verbs (e.g. *dikha* 'to appear')
  5. Copula (e.g. *hai* 'is')
  6. Goal verbs (e.g. *jaa* 'go')

# Lexicon: Verbal heads

- The 6 verb classes can be thought of as tree-templates and can be associated with various class specific constraints such as:
  1. number of mandatory arguments,
  2. part-of-speech category of the arguments,
  3. canonical order of the arguments,
  4. relative position of the argument with respect to the verb,
  5. agreement (but for now won't talk about this)
  6. etc.

# Lexicon: Tree-templates

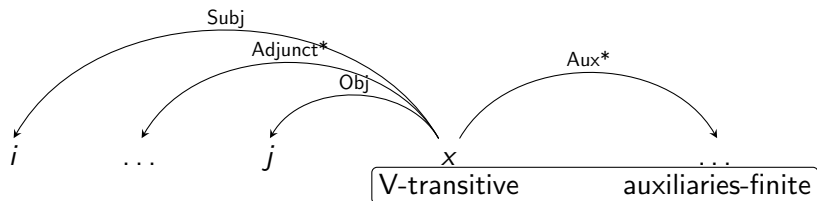


Figure 1: A simplified transitive tree-template. Aux = Auxiliaries. \* signifies 0 or more instances.

- $(i, j, x)$  will be instantiated by those lexical items that are of a particular type (eg.  $x$  with transitive verbs)
- encodes the arity of the verbal head
- canonical word order of its dependents

## Lexicon: Tree-templates encode argument order

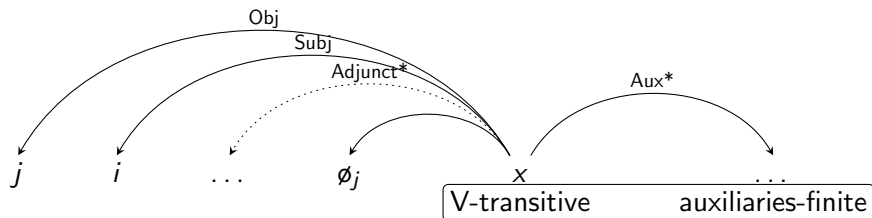


Figure 2: A simplified transitive tree template showing object fronting. Aux = Auxiliaries,  $\emptyset_j$  = canonical position of object  $j$

- the arc coming into the empty node ( $\emptyset_j$ ) is not a dependency relation (just a representational strategy to show order variation)

# Frame variations

- Tree-templates have been induced using finite verb occurrences,
- But, finite-templates cannot be used for non-finite verbs
  - ▶ because the surface requirements of non-finite verbs are different from that of finite verbs,
  - ▶ when *khaa* “eat” occurs as *khaakara* “having eaten”, its original requirement changes: no subject,
  - ▶ In addition, it requires another finite or a non-finite verb as its head

# Frame variations

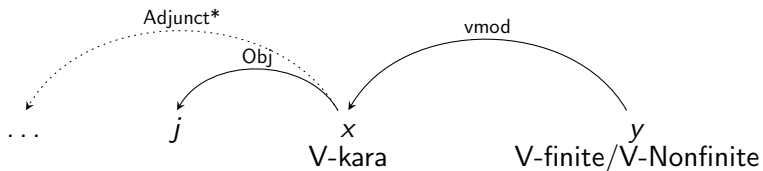


Figure 3: A -kara tree-template. fin = Finite.

# Inducing a (psycholinguistically motivated) dependency grammar for Hindi from a dependency treebank

- Main components:
  - ✓ Lexicon,
  - ✓ Frame variations,
  - ☞ Probability distribution of dependency types,
    - ▶ Prediction rules

# Prediction rules

- As each incoming word is incorporated into the existing structure, predictions are made about upcoming words based on current information
- In order for the parser to proceed in such a fashion it must have ready access to such information
- The grammar that we propose, provides this information in the form of prediction rules
- We begin with one simple cue, case-marker of the arguments
  - ▶ Note: for illustrative purpose, while gathering statistics to formulate predictions shown here, we consider only verbal arguments. The presence of adjuncts has been neglected.

# Prediction rules

- Arg1-Case → Verb cluster/template
- *ne* (ERG) → transitive
- *ko* (ACC) → transitive
- *se* (INST) → ditransitive
- *o* (NOM) → intransitive
  
- The verb classes that we get for *ne*, *se*, *o* reflect the default distribution of ERG, INST and NOM case-markers vis-à-vis the type of verbs a they tend to occur with
- Of course, predictions will become more precise as more words are processed

# Prediction rules

- Arg1-Case Arg2-Case → Verb cluster
  
- *0 0* → copula
- *0 -* → intransitive
- *0 se* → transitive
- *0 ko* → transitive
- *0 ne* → transitive
  
- *ne 0* → transitive
- *ne ko* → transitive/ditransitive
- *ne se* → ditransitive

# Prediction rules

- as we get more information, we might have to revise our previous predictions and make necessary structural changes (or rerank in case of multiple parses)
- one can also use other features to make these predictions more realistic. For example, we could use features such as:
  - ▶ Position in sentence,
  - ▶ Animacy feature, etc.

# Predictions: Processing concerns

- Given the observation that predictions will go wrong and the parser will have to make revisions (or rerank), we need to ask:
  - ▶ What is predicted?
  - ▶ What are the different cues that are pooled to make a prediction?
  - ▶ What is the processing cost when a prediction is incorrect?
  - ▶ How can we quantify this cost?
  - ▶ How does the prediction system interact with other aspects of the comprehension process?

# Predictions: Processing concerns

| Prediction   | CO  | NCO  |
|--|---|--|
| <b>Correct prediction</b>  | Predicted → <i>0 0</i> : copula<br>Correct → <i>0 0</i> : copula                          | Predicted → <i>ko ne</i> : transitive<br>Correct → <i>ko ne</i> : transitive   |
| <b>Incorrect prediction</b><br>(incorrect class)                   | Predicted → <i>0 0</i> : copula<br>Correct → <i>0 0</i> : transitive                      | Predicted → <i>ko ne</i> : transitive<br>Correct → <i>ko ne</i> : ditransitive |
| <b>Incorrect prediction</b><br>(incorrect word order)              | Predicted → <i>ko ko</i> : ditransitive<br>Correct → <i>ko ko</i> : ditransitive<br>(NCO) | Predicted → ?<br>Correct → ?   |
| <b>Incorrect prediction</b><br>(incorrect class<br>and word order) | Predicted → <i>ko 0</i> : ditransitive<br>Correct → <i>ko 0</i> : transitive (NCO)        | Predicted → ?<br>Correct → ?   |

Table 1: Different prediction scenarios. Canonical order: CO, Non-canonical order: NCO

- Prediction: A verb template (Verb class, argument structure order)
- for example, after the 1st *ko* is seen, a canonical transitive-template is predicted, this prediction changes to non-canonical transitive template in case *ne* happens to be the next case-marker; on the other hand if a *0* case-marker was encountered instead, the parser revises its prediction to a canonical ditransitive-template.

## Predictions: Processing concerns

- There are 2 factors that will influence the processing cost of a prediction:
  - 1 Correct/Incorrect verb-template prediction,
  - 2 Correct/Incorrect argstr-order prediction,
- Based on these two factors, the processing hypothesis about the cost of such a prediction is:

*Correct prediction < Incorrect prediction (argstr order or verb class) < Incorrect prediction (argstr and class)*
- This will have to be evaluated experimentally.

- ✓ Introduction
- ✓ Relevant experimental work
- ✓ Grammar induction
- ✓ Processing concerns
  - (An outline of a) dependency-based human sentence processor
  - Issues and challenges

# Dependency-based human sentence processor: An outline

- Boston, Hale, Patil, Kliegl, and Vasishth (2008) have previously used a transition-based dependency parser to model human sentence processing difficulty
  - ▶ But, their parser will not be able to correctly analyse crossing/discontiguous dependencies
  - ▶ In addition, they have no notion of prediction explicitly built into their system

# Dependency-based human sentence processor: An outline

- We plan to adapt the graph-based dependency parsing paradigm
- The formulation proposed by McDonald, Crammer, and Pereira (2005), McDonald, Pereira, Ribarov, and Hajič (2005) needs to be modified in order to adapt it for the goals of this paper
  - a. the algorithm needs to be incremental
    - ★ instead of starting with a complete sentence, one needs to form complete graphs out of all the available words at any given time
  - b. prediction rules need to part of the parsing process
    - ★ forming complete graphs using unlexicalized tree-template (that will be predicted by already seen tokens), and extracting MST out of it
  - c. use the parser to compute costs due to
    - ★ memory-based constraints, expectation, word order variation

- Dependency grammar based human sentence processing system presents itself as an attractive alternative to phrase structure based models currently dominant in the psycholinguistic literature
  - ▶ Representational simplicity
  - ▶ Efficient parsing paradigms
  - ▶ cf. talks at Depling2013

# Issues and challenges

- Parser adaptation
  - ▶ the prediction system needs to be seamlessly integrated within the parsing process
- Configurational constraints:
  - ▶ considerable evidence that while processing some dependencies, for example, filler-gap dependencies, anaphora resolution, etc., human sentence comprehension system uses certain grammatical constraints (Phillips, Wagers, & Lau, 2011)
  - ▶ these constraints (e.g. c-command) have been traditionally formalized using phrase-structure representation
  - ▶ If it is true that the parser does employ configurational constraints such as c-command then it will be imperative to formulate a functionally equivalent definition of c-command within the dependency framework

# Final remarks

- An attempt to build a dependency-based human sentence processor
- Dependency grammar induction is the first step towards that goal
- Grammar induction also points to various patterns that can be used to design informed experiments (which will in turn help the grammar/parser development)
  - ✓ Predictive parsing (and how it interacts with working memory constraints),
  - 👉 Non-contiguous dependencies (e.g. Relative clause, Genitives),
  - 👉 Processing non-canonical argument structure order
- Graph-based dependency parser needs to be adapted to incorporate a predictive component, and to reflect memory-based constraints

- Bader, M., & Meng, M. (1999). Subject-Object Ambiguities in German Embedded Clauses: An Across-the-Board Comparison. In *Journal of Psycholinguistic Research*. (Vol. 28, pp. 121–143).
- Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D., & Xia, F. (2009). A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu. In *Proceedings of the Third Linguistic Annotation Workshop* (pp. 186–189). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Boston, M. F., Hale, J. T., Patil, U., Kliegl, R., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2(1), 1–12.
- Butt, M., & King, T. C. (1996). Structural topic and focus without movement. In M. Butt and T. H. King, eds., *The First LFG Conference*. CSLI Publications.
- Demberg, V. (2010). *A broad-coverage model of prediction in human sentence processing*. Unpublished doctoral dissertation, The University of Edinburgh.
- Gambhir, V. (1981). *Syntactic restrictions and discourse functions of word order in standard Hindi*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.
- Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita W. O'neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium*. Cambridge, MA: MIT Press.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29(2), 261–290.
- Husain, S., Vasishth, S., & Srinivasan, N. (2013). Expectation and Locality in Hindi. In *Proceedings of AMLaP Conference*. Marseille.
- Hyona, J., & Hujanen, H. (1997). Effects of Case Marking and Word Order on Sentence Parsing in Finnish: An Eye Fixation Analysis. In *The Quarterly Journal of Experimental Psychology* (Vol. 50A, pp. 841–858).
- Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. In *Cognition* (Vol. 94, pp. 113–147).
- Kamide, Y., Scheepers, C., & Altmann, G. (2003). Integration of Syntactic and Semantic Information in Predictive Processing: Cross-Linguistic Evidence from German and English. In *Journal of Psycholinguistic Research* (Vol. 32, pp. 37–55).
- Kidwai, A. (2000). *XP-Adjunction in universal grammar: Scrambling and Binding in Hindi-Urdu*. Oxford University Press, New York.
- Kim, A., Srinivas, B., & Trueswell, J. (1998). The convergence of lexicalist perspectives in psycholinguistics and computational linguistics. In P. Merlo and S. Stevenson (eds.), *Papers from the Special Section on the Lexicalist Basis of Syntactic Processing, CUNY Conference*.
- Kolachina, P., Kolachina, S., Singh, A. K., Husain, S., Viswanatha, N., Sangal, R., et al. (2010). Grammar Extraction from Treebanks for Hindi and Telugu. In *Proceedings of The 7th International Conference on Language Resources and Evaluation (LREC)*. Valletta. Malta.
- Konieczny, L. (2000). Locality and Parsing Complexity. In *Journal of Psycholinguistic Research* (Vol. 29, pp. 627–645).
- Kuhlmann, M. (2007). *Dependency Structures and Lexicalized Grammars*. Unpublished doctoral dissertation, Saarland University, Saarbrücken, Germany.
- Levy, R. (2008). Expectation-based syntactic comprehension. In *Cognition* (Vol. 106, pp. 1126–1177).
- Levy, R., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, 122(1), 12–36.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 1–45.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.
- McDonald, R., Crammer, K., & Pereira, F. (2005). Online large-margin training of dependency parsers. In *Proceedings of the 43rd ACL*.
- McDonald, R., Pereira, F., Ribarow, K., & Hajič, J. (2005). Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of HLT/EMNLP*.
- Nivre, J., & Nilsson, J. (2005). Pseudo-projective dependency parsing. In *Proc. of acl 2005*.
- Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (ed.), *Experiments at the Interfaces, Syntax and Semantics* (Vol. 37, pp. 153–186).
- Pickering, M. (1994). Processing Local and Unbounded Dependencies: A Unified Account. In *Journal of Psycholinguistic Research* (Vol. 23, p. 323–352).
- Pickering, M., & Barry, G. (1991). Sentence Processing without Empty Categories. In *Language and Cognitive Processes* (Vol. 6, p. 229–259).
- Sekera, I. A. (2003). Scrambling and Processing: Dependencies, Complexity, and Constraints. In S. Karimi (Ed.), *Word Order and Scrambling* (pp. 301–324). Blackwell Publishers.
- Staub, A., & Clifton, J. C. (2006). Syntactic prediction in language comprehension: Evidence from either ... or. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Vol. 32, p. 425–436).
- Vasishth, S. (2004). Discourse Context and Word Order Preferences in Hindi. In R. Singh (ed.), *The Yearbook of South Asian Languages and Linguistics* (pp. 113–127).
- Vasishth, S., & Drenhaus, H. (2011). Locality effects in German. *Dialogue and Discourse*, 1(2), 59–82.
- Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 82(4), 767–794.
- Vasishth, S., Shahar, R., & Srinivasan, N. (2012). The role of clefting, word order and given-new ordering in sentence comprehension: Evidence from Hindi. In *Journal of South Asian Linguistics* (Vol. 5).
- Xia, F. (2001). *Automatic Grammar Generation from Two Different Perspectives*. Unpublished doctoral dissertation, University of Pennsylvania.

Thanks!