Categorial Proof Nets and Dependency Locality

A New Metric for Linguistic Complexity

Mehdi Mirzapour
Jean-Philippe Prost
Christian Retoré

August 31, 2018
LAComLing2018
Department of Mathematics
University of Stockholm

Content:

PART I : The Problem of Linguistic Difficulty Measurement

PART II : Review of Gibson's Psycholinguistic Theories

PART III: Linguistic Difficulty Metrics using Categorial Proof Nets

PART I:

The Problem of Linguistic Difficulty Measurement

The Problem

A (quantitative) computational linguistic account of why a sentence is harder to be comprehended (by human) than some other one?

Examples: [Gibson, 91]

- The reporter disliked the editor.
- The reporter [who the senator attacked] disliked the editor
- The reporter [who the senator [who John met] attacked] disliked the editor].

PART II:

Review of Gibson's Psycholinguistic Theories

Gibson’s Psycholinguistic Theories

- Incomplete Dependency Theory [Gibson, 1991]
- Dependency Locality Theory [Gibson, 2000]

Incomplete Dependency Theory [Gibson, 1991]

- IDT is based on the idea of counting missing incomplete dependencies during the incremental processing of a sentence when a new word attaches to the current linguistic structure.

- The main parameter in IDT is the number of incomplete dependencies when the new word integrates to the existing structure.

Incomplete Dependency Theory [Gibson, 1991]

Example: The reporter [who the senator [who John met] attacked] disliked the editor].

- Five incomplete dependencies at the point of processing “John”.

1. the NP the reporter is dependent on a verb that should follow it;
2. the NP the senator is dependent on a different verb to follow;
3. the pronoun who (before the senator) is dependent on a verb to follow
4. the NP John is dependent on another verb to follow
5. the pronoun who (before John) is dependent on a verb to follow.

- These are five unsaturated or incomplete or unresolved dependencies.

Dependency Locality Theory [Gibson, 2000]

- DLT is a distance-based referent-sensitive linguistic complexity measurement put forward by Gibson to supersede the predictive limitations of the incomplete dependency theory.

- The linguistic complexity is interpreted as the locality-based cost of the integration of a new word to the dependent word in the current linguistic structure which is the number of the intervened new discourse-referents.

Dependency Locality Theory [Gibson, 2000]

Example:

- The reporter [who the senator [who John met] attacked] disliked the editor].
- The reporter [who the senator [who I met] attacked] disliked the editor].

PART III:

Linguistic Difficulty Metrics using Categorial Proof Nets

Lambek Categorial Grammar [Lambek,1958]

Definition 2.1. The category formulas (L_p) are freely generated from a set of usual syntactical primitive types $P = \{S, np, n, pp, \cdots\}$ by directional divisions, namely the binary infix connectives \ (over), / (under) and $\bullet$ (product) as follows:

$$Lp ::= P \mid (L_p \backslash L_p) \mid (L_p / L_p) \mid (L_p \bullet L_p)$$

	Introduction rules	Elimination rules
Intuitionistic	$\dfrac{\begin{array}{c}[A]^n\\ \vdots\\ B\end{array}}{A \to B} \to I_n$	$\dfrac{A \quad A \to B}{B} \to E$
Lambek	$\dfrac{\begin{array}{c}[A]^n \ldots\\ \vdots\\ B\end{array}}{A \backslash B} \backslash I_n$ $\dfrac{\begin{array}{c}\ldots [A]^n\\ \vdots\\ B\end{array}}{B / A} / I_n$	$\dfrac{A \quad A \backslash B}{B} \backslash E$ $\dfrac{B / A \quad A}{B} / E$

Examples:

Relevant Lambek Proof:

$$\dfrac{\dfrac{\dfrac{\dfrac{(S\,/\,(np\,\backslash\,S))\,/\,n \quad n}{(S\,/\,(np\,\backslash\,S))}\,/_e \qquad \dfrac{(np\,\backslash\,S)\,/\,np \quad [np]^1}{(np\,\backslash\,S)}\,/_e}{S}\,/_e}{S\,/\,np}\,/_i(1) \qquad \dfrac{((S\,/\,np)\,\backslash\,S)\,/\,n \quad n}{(S\,/\,np)\,\backslash\,S}\,\backslash_e}{S}\,\backslash_e$$

Corresponding Intuitionistic Proof:

$$\dfrac{\dfrac{\dfrac{\dfrac{\overset{\textit{every}}{(\mathbf{e}\to\mathbf{t})\to(\mathbf{e}\to\mathbf{t})\to\mathbf{t}} \quad \overset{\textit{child}}{(\mathbf{e}\to\mathbf{t})}}{(\mathbf{e}\to\mathbf{t})\to\mathbf{t}}\to_e \qquad \dfrac{\overset{\textit{ate}}{\mathbf{e}\to\mathbf{e}\to\mathbf{t}} \quad \overset{\textit{o}}{[\mathbf{e}]^1}}{\mathbf{e}\to\mathbf{t}}\to_e}{\mathbf{t}}\to_e}{\mathbf{e}\to\mathbf{t}}\to_i(1) \qquad \dfrac{\overset{\textit{a}}{(\mathbf{e}\to\mathbf{t})\to(\mathbf{e}\to\mathbf{t})\to\mathbf{t}} \quad \overset{\textit{pizza}}{(\mathbf{e}\to\mathbf{t})}}{(\mathbf{e}\to\mathbf{t})\to\mathbf{t}}\to_e}{\mathbf{t}}\to_e$$

Sequent Calculus Rules for LC

$$\frac{\Gamma, B, \Gamma' \vdash C \quad \Delta \vdash A}{\Gamma, \Delta, A \backslash B, \Gamma' \vdash C} \backslash h \qquad \frac{A, \Gamma \vdash C}{\Gamma \vdash A \backslash C} \backslash i \quad \Gamma \neq \epsilon$$

$$\frac{\Gamma, B, \Gamma' \vdash C \quad \Delta \vdash A}{\Gamma, B/A, \Delta, \Gamma' \vdash C} /h \qquad \frac{\Gamma, A \vdash C}{\Gamma \vdash C/A} /i \quad \Gamma \neq \epsilon$$

$$\frac{\Gamma, A, B, \Gamma' \vdash C}{\Gamma, A \bullet B, \Gamma' \vdash C} \bullet h \qquad \frac{\Delta \vdash A \quad \Gamma \vdash B}{\Delta, \Gamma \vdash A \bullet B} \bullet i$$

$$\frac{\Gamma \vdash A \quad \Delta_1, A, \Delta_2 \vdash B}{\Delta_1, \Gamma, \Delta_2 \vdash B} cut \qquad \frac{}{A \vdash A} axiom$$

Examples

$$\dfrac{\dfrac{\dfrac{}{S \vdash S}\,axiom \qquad \dfrac{}{np \vdash np}\,axiom}{np, np\backslash S \vdash S}\,\backslash h \qquad \dfrac{}{np \vdash np}\,axiom}{np, (np\backslash S)/np, np \vdash S}\,/h$$

$$\dfrac{\dfrac{\dfrac{C \vdash C \quad B \vdash B}{B, B\backslash C \vdash C}\,\backslash h \qquad A \vdash A}{A, A\backslash B, B\backslash C \vdash C}\,\backslash h}{A\backslash B, B\backslash C \vdash A\backslash C}\,\backslash i$$

$$\dfrac{\dfrac{\dfrac{B \vdash B \quad A \vdash A}{A, A\backslash B \vdash B}\,\backslash h \qquad C \vdash C}{A, A\backslash B, B\backslash C \vdash C}\,\backslash h}{A\backslash B, B\backslash C \vdash A\backslash C}\,\backslash i$$

Definitions:

Definition of \ and /: $A\backslash B \equiv A^{\perp} \wp B \qquad B/A \equiv B \parr A^{\perp}$

De Morgan equivalences $(A^{\perp})^{\perp} \equiv A \qquad (A \parr B)^{\perp} \equiv B^{\perp} \otimes A^{\perp}$

$$(A \otimes B)^{\perp} \equiv B^{\perp} \parr A^{\perp}$$

A polar category formula is a Lambek categorial type labeled with positive (°) or negative (•) polarity recursively definable as follows:

$$L^{\circ} ::= P \mid (L^{\bullet} \parr L^{\circ}) \mid (L^{\circ} \parr L^{\bullet}) \mid (L^{\circ} \otimes L^{\circ})$$
$$L^{\bullet} ::= P^{\perp} \mid (L^{\circ} \otimes L^{\bullet}) \mid (L^{\circ} \otimes L^{\bullet}) \mid (L^{\bullet} \parr L^{\bullet})$$

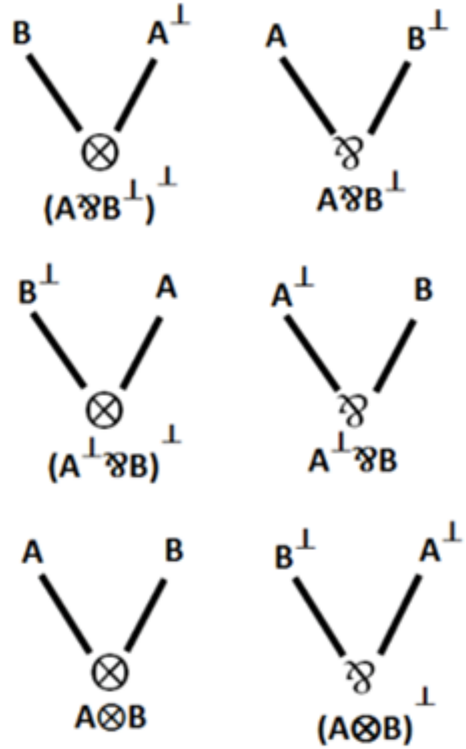
Based on the previous inductive definitions, we can have an easy decision procedure to check whether a formula F is in L° or $L^{\bullet}$:

$\otimes$	$\bullet$	$\circ$
$\bullet$	*undefined*	$\bullet$
$\circ$	$\bullet$	$\circ$

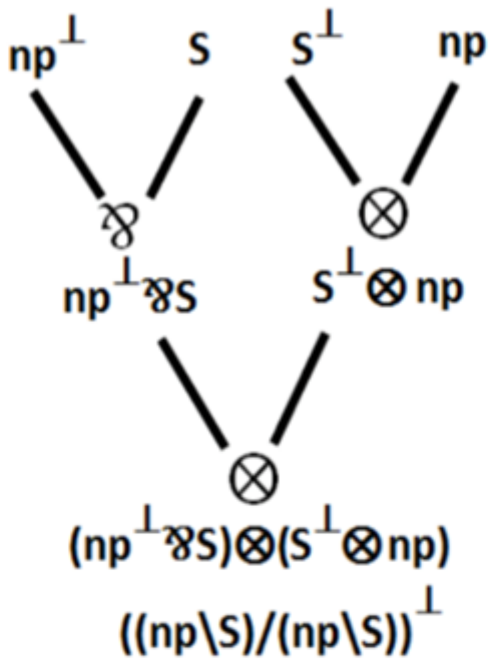
$\parr$	$\bullet$	$\circ$
$\bullet$	$\bullet$	$\circ$
$\circ$	$\circ$	*undefined*

Definition:

A polar category formula tree is a binary ordered tree in which the leaves are labeled with polar atoms and each local tree is one of the following logical links:



Example



The polar categorial tree of $((np\backslash S)/(np\backslash S))^{\perp}$

Categorial Proof Nets [Moot, Retoré, 2012]

Definition A proof frame is a finite sequence of polar category formula trees with one positive polarity corresponding to the unique succedent of sequent.

Definition A proof structure is a proof frame with axiom linking which corresponds to the axiom rule in the sequence calculus. Axioms are a set of pairwise disjoint edges connecting a leaf z to a leaf $z^{\perp}$, in such a way that every leaf is incident to some axiom link.

Definition A proof net is a proof structure satisfying the following conditions:

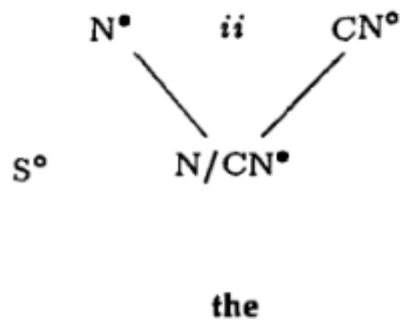
Acyclicity: every cycle contains the two edges of the same $\wp$ branching.

Enumerate: there is a path not involving the two edges of the same $\wp$ branching between any two vertices.

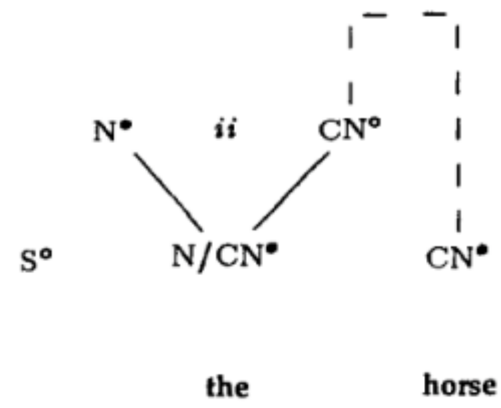
Intuitionism: every conclusion can be assigned some polarity.

Non commutativity: the axioms do not cross (are well bracketed).

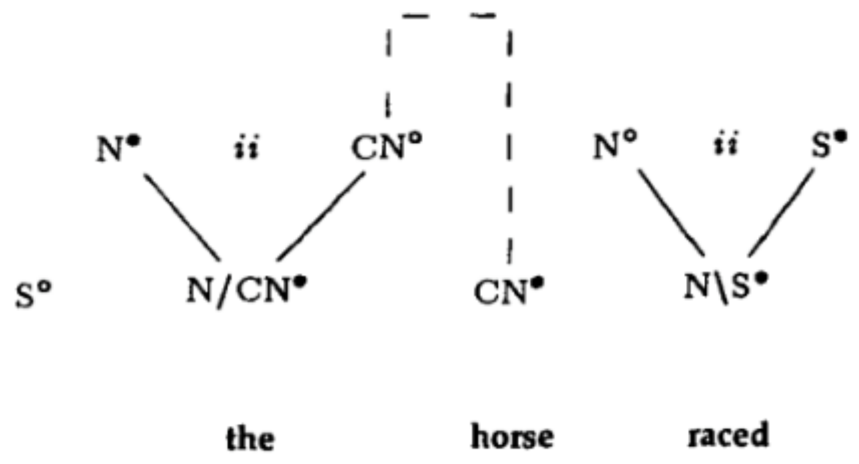
Incremental Processing with CPN [Morrill, 2000]



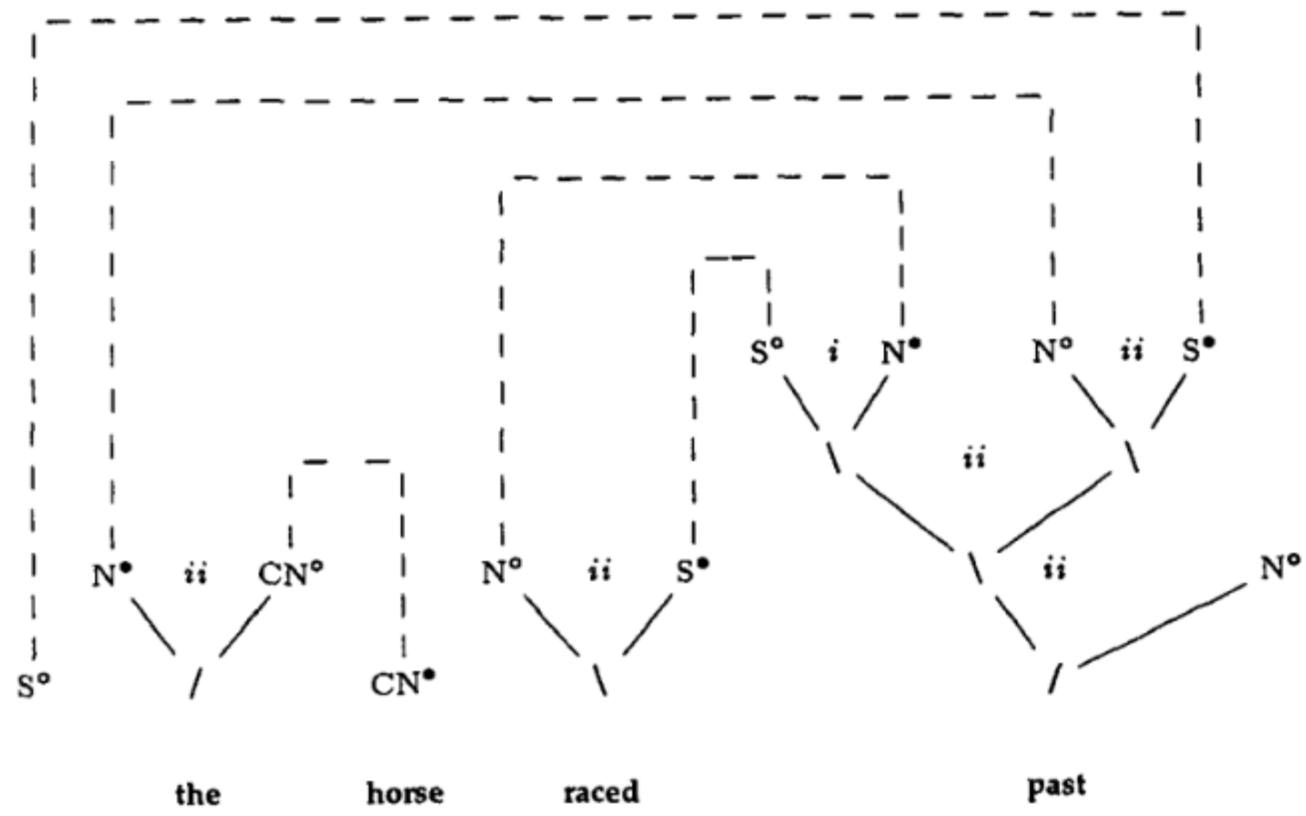
Incremental Processing with CPN [Morrill, 2000]



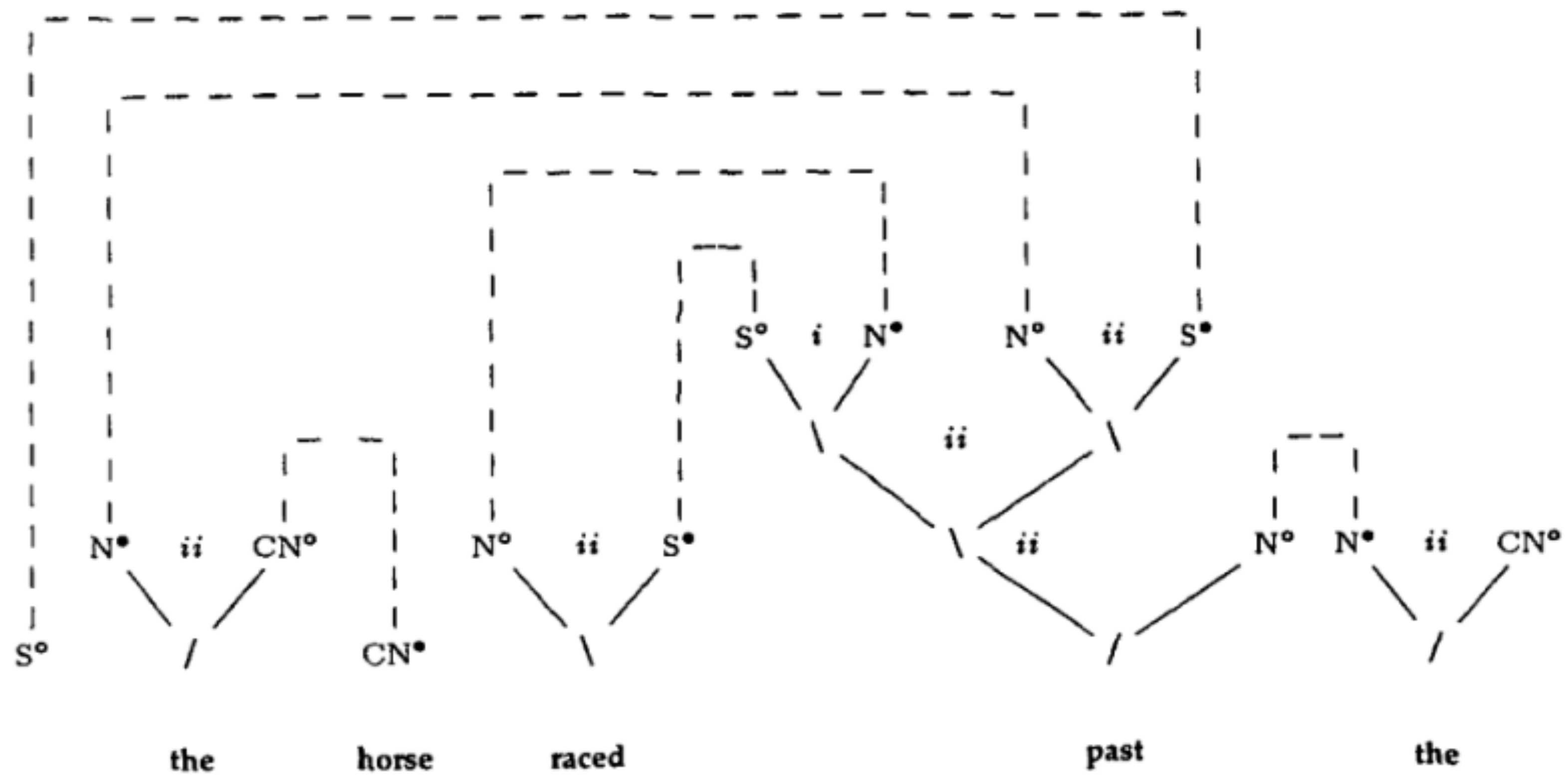
Incremental Processing with CPN [Morrill, 2000]



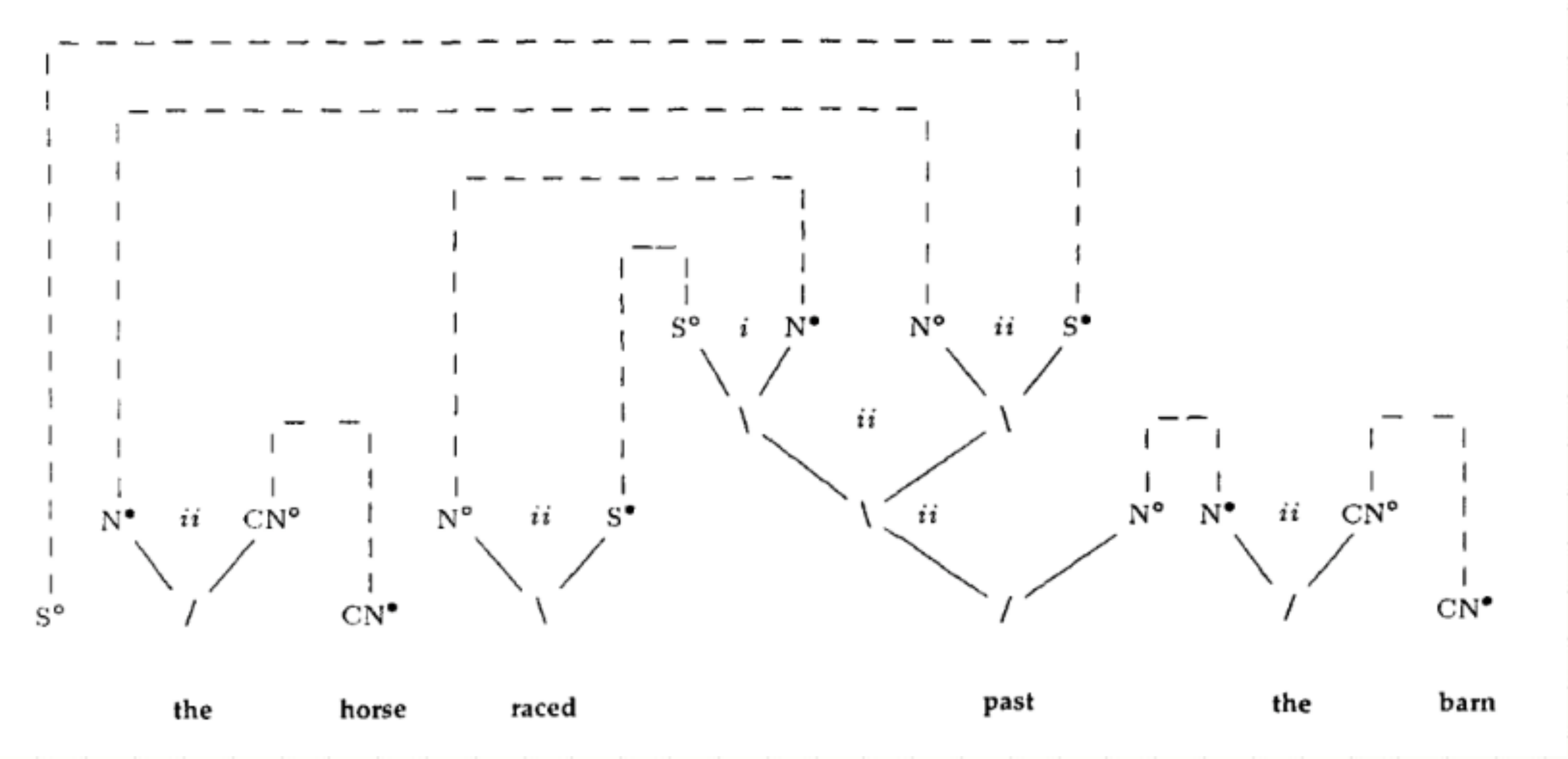
Incremental Processing with CPN [Morrill, 2000]



Incremental Processing with CPN [Morrill, 2000]



Incremental Processing with CPN [Morrill, 2000]



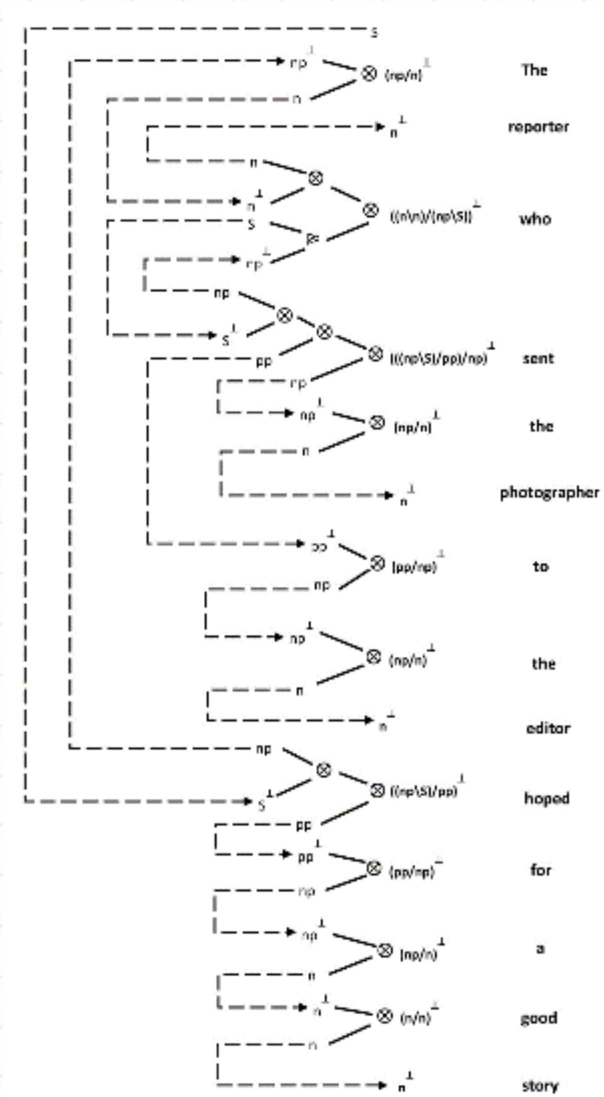
IDT-based Complexity Profiling [Morrill, 2000]

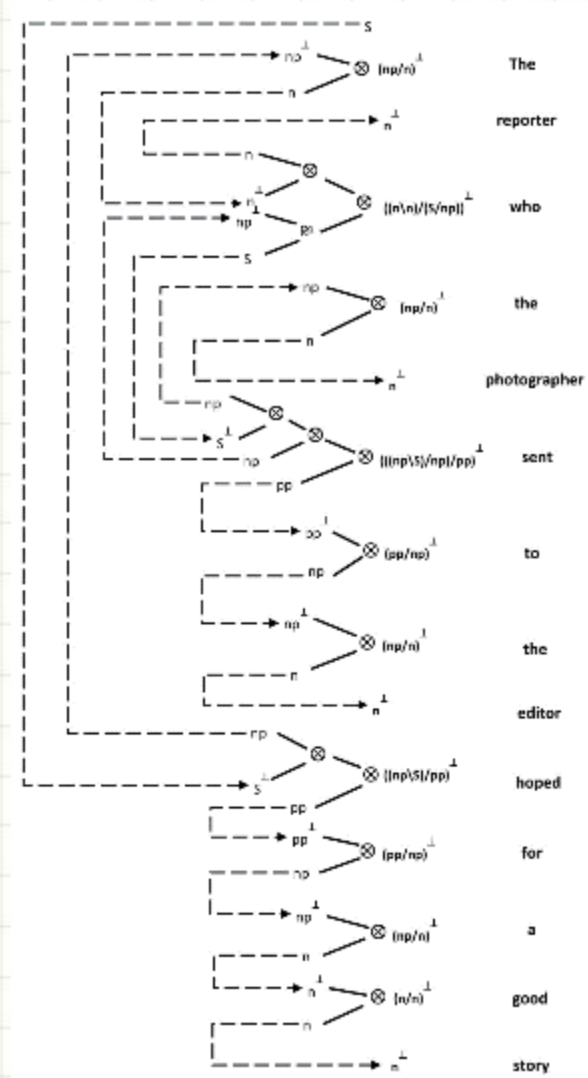
Definition 1: Let π be a a syntactic analysis of $w_1, \cdots, w_n$ with categories $C_1, \ldots,$ C_n — that is a categorial proof net with conclusions $(C_n)^\perp$, ..., $(C_1)^\perp, S$. Let C_{i_0} be one of the C_i ($i \in [1, n]$). The incomplete dependency number of C_{i_0} in π, written as $ID_\pi(C_{i_0})$, is the count of axioms $c - c'$ in π such that $c \in (C_{i_0 - m} \cup S)$ $(m \geq 0)$ and $c' \in C_{i_0+n+1}$ $(n \geq 0)$.

Definition 2: Let π be a a syntactic analysis of $w_1, \cdots, w_n$ with categories $C_1, \ldots,$ C_n — that is a categorial proof net with conclusions $(C_n)^\perp$, ..., $(C_1)^\perp, S$. We define the IDT-based linguistic complexity of π, written $f_{idt}(\pi)$ by $(1 + \sum_{i=1}^{n} ID_\pi(C_i))^{-1}$.

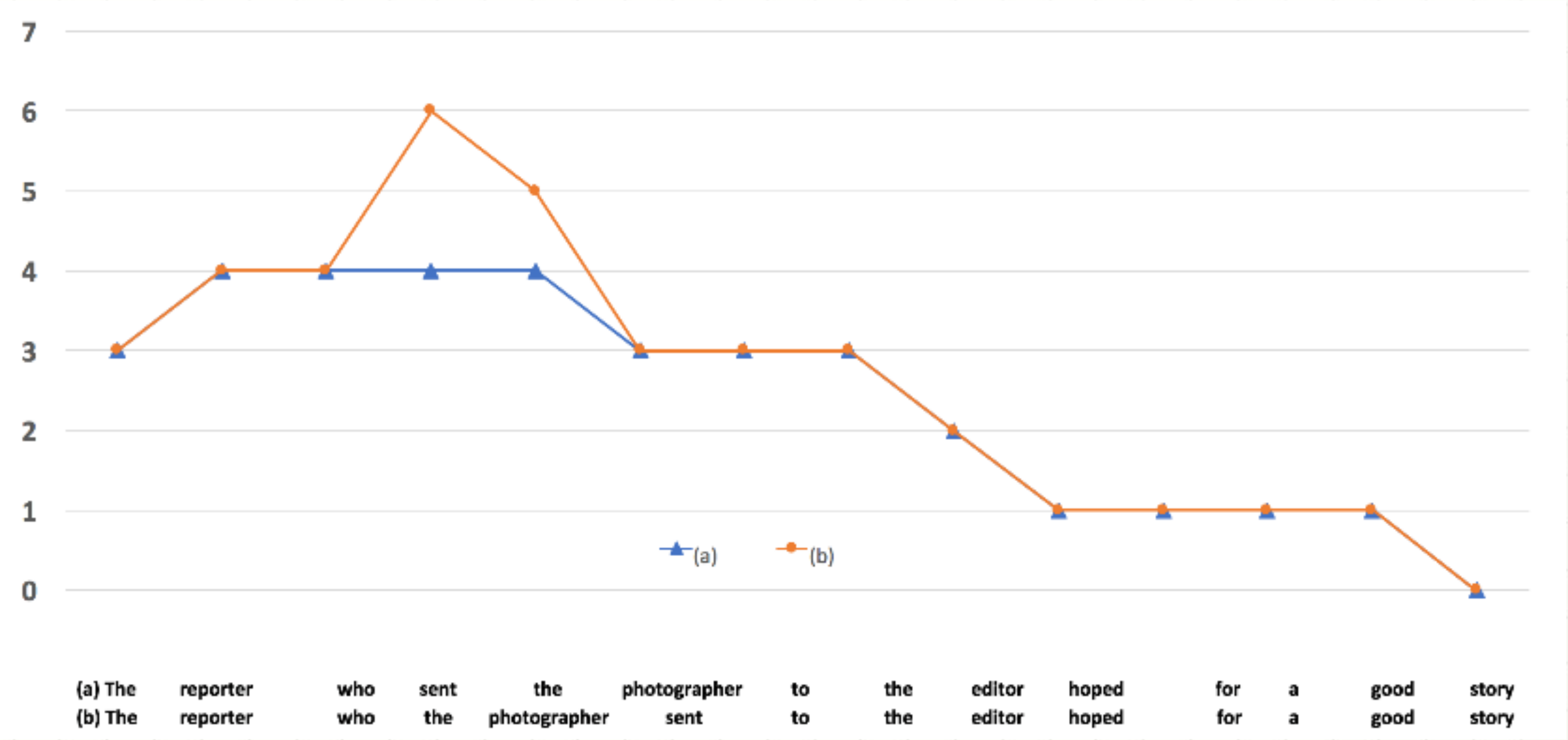
Definition 3: Given two syntactic analyses π_i and π_j, not necessarily of the same words and categories, we say that π_i is IDT-preferred to π_j whenever $f_{idt}(\pi_i) > f_{idt}(\pi_j)$.

Subject/Object-extracted Relative Clauses





Subject/Object-extracted Relative Clauses



DLT-based Complexity Profiling

Definition 4: A word w is said to be a discourse referent whenever it is a *proper noun*, *common noun* or *verb*.

Definition 5: Let π be a a syntactic analysis of $w_1, \cdots, w_n$ with categories $C_1, \ldots,$ C_n — that is a categorial proof net with conclusions $(C_n)^\perp$, ..., $(C_1)^\perp, S$. Let $c - c'$ be an axiom in π such that $c \in C_i$ and $c' \in C_j$ $(i, j \in [1, n])$. We define the **length** of axiom $c - c'$ as the integer $i + 1 - j$.

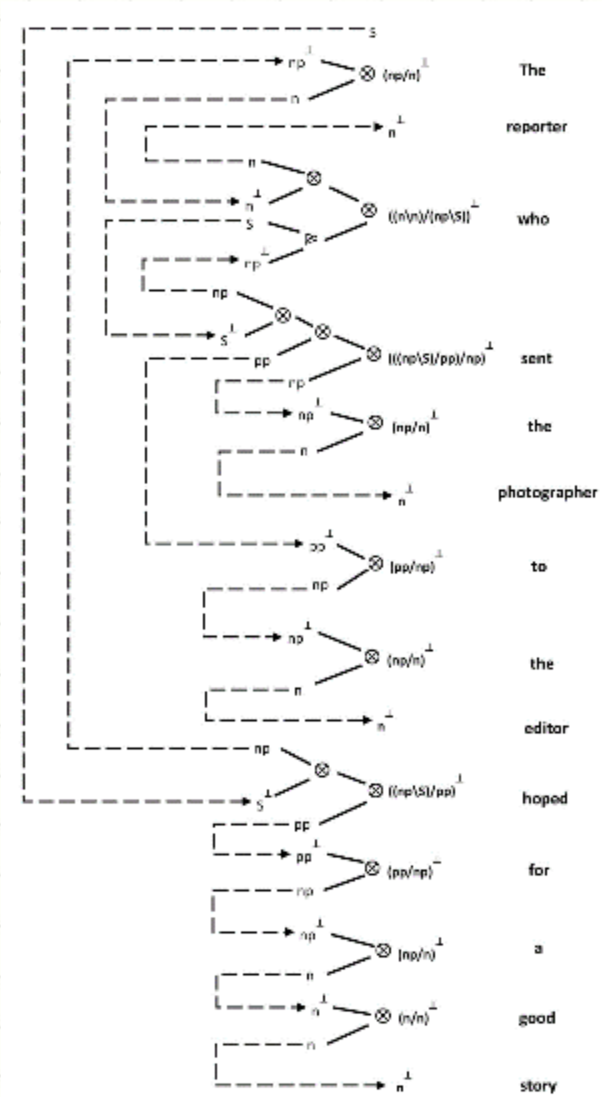
Definition 6: Let π be a a syntactic analysis of $w_1, \cdots, w_n$ with categories $C_1, \ldots,$ C_n — that is a categorial proof net with conclusions $(C_n)^\perp$, ..., $(C_1)^\perp, S$. Let C_{i_0} be one of the C_i, and let consider axioms $c - c'$ with c in C_{i_0} and c' in some C_{i_0-k}. Let us consider the largest k for which such an axiom exists — this is the longest axiom starting from C_{i_0} with the previous definition. The dependency locality number of C_{i_0} in π, written $DL_\pi(C_{i_0})$ is the number of discourse referent words between $w_{i_0} : C_{i_0}$ and $w_{i_0-k} : C_{i_0-k}$. The boundary words, i.e. $w_{i_0} : C_{i_0}$ and $w_{i_0-k} : C_{i_0-k}$ should also be counted. Alternatively, it may be viewed as $k+1$ minus the number of non-discourse references among those $k + 1$ words.

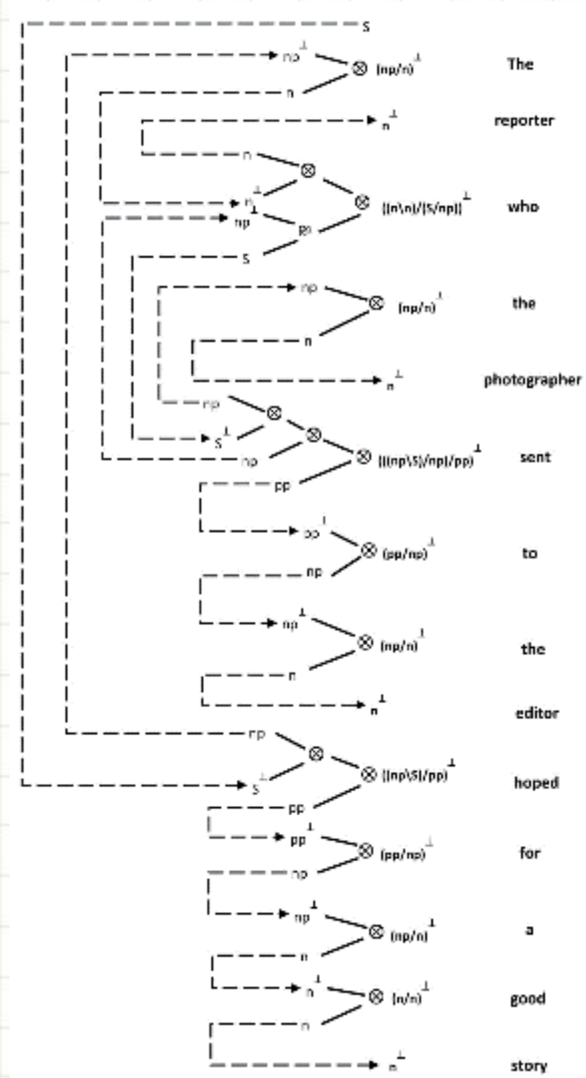
DLT-based Complexity Profiling

Definition 7: Let π be a a syntactic analysis of $w_1, \cdots, w_n$ with categories $C_1, \ldots,$ C_n — that is a categorial proof net with conclusions $(C_n)^\perp$, ..., $(C_1)^\perp$, S. We define the DLT-based linguistic complexity of π, written $f_{dlt}(\pi)$ by $(1 + \sum_{i=1}^{n} DL_\pi(C_i))^{-1}$.

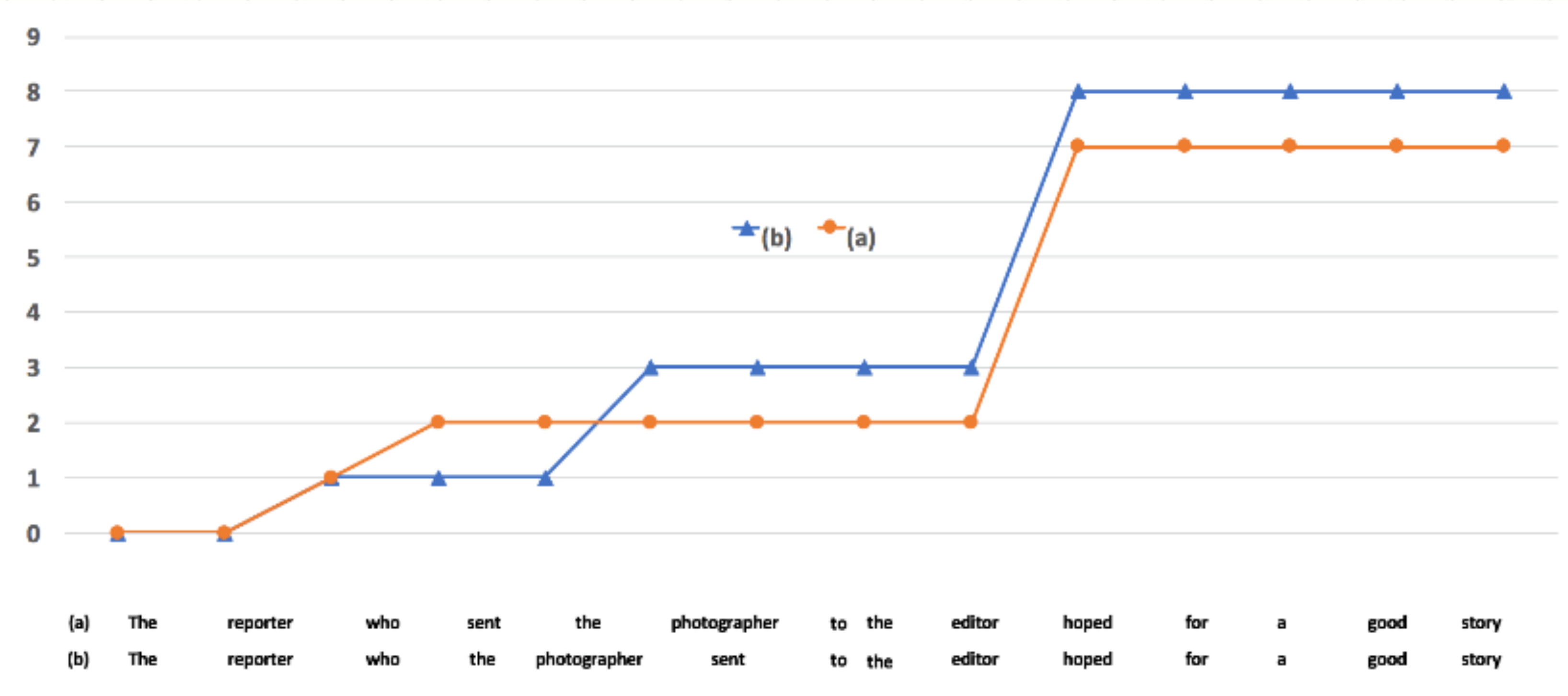
Definition 8: Given two syntactic analyses π_i and π_j, not necessarily of the same words and categories, we say that π_i is DLT-preferred to π_j whenever $f_{dlt}(\pi_i) > f_{dlt}(\pi_j)$.

Subject/Object-extracted Relative Clauses [Gibson, 2000]

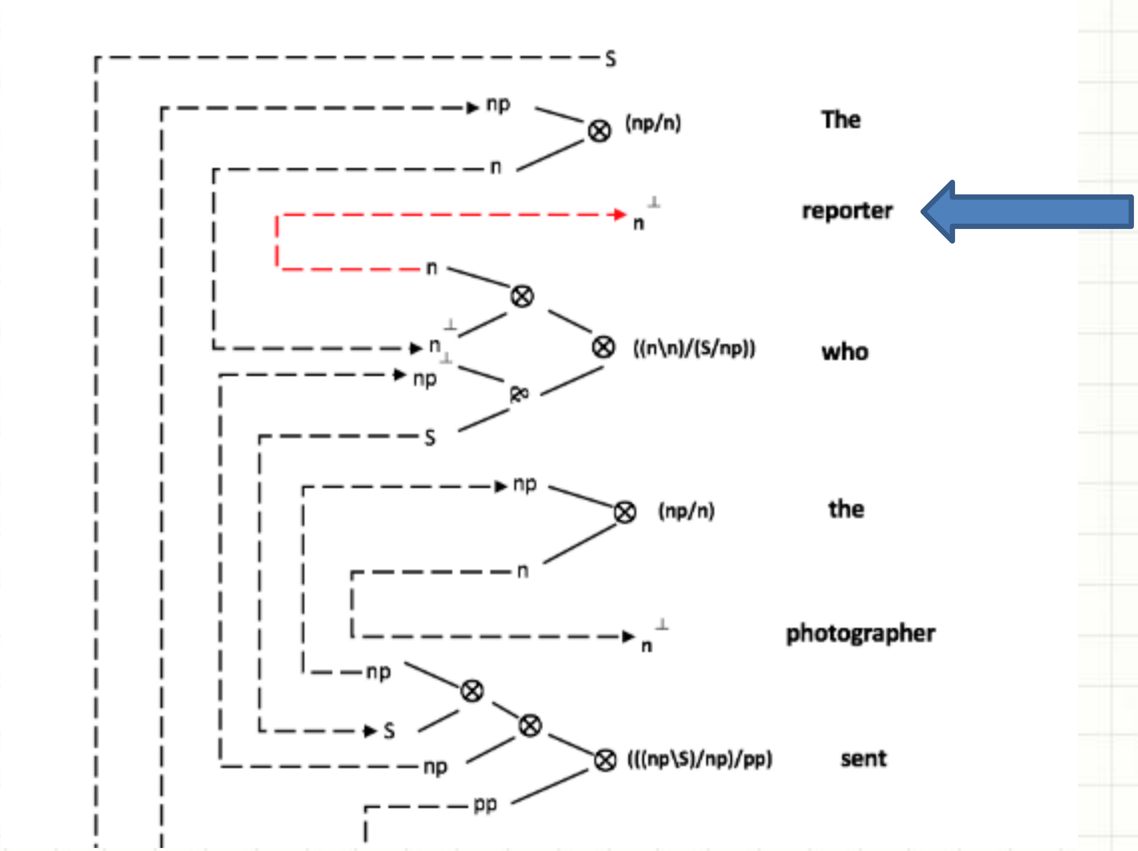




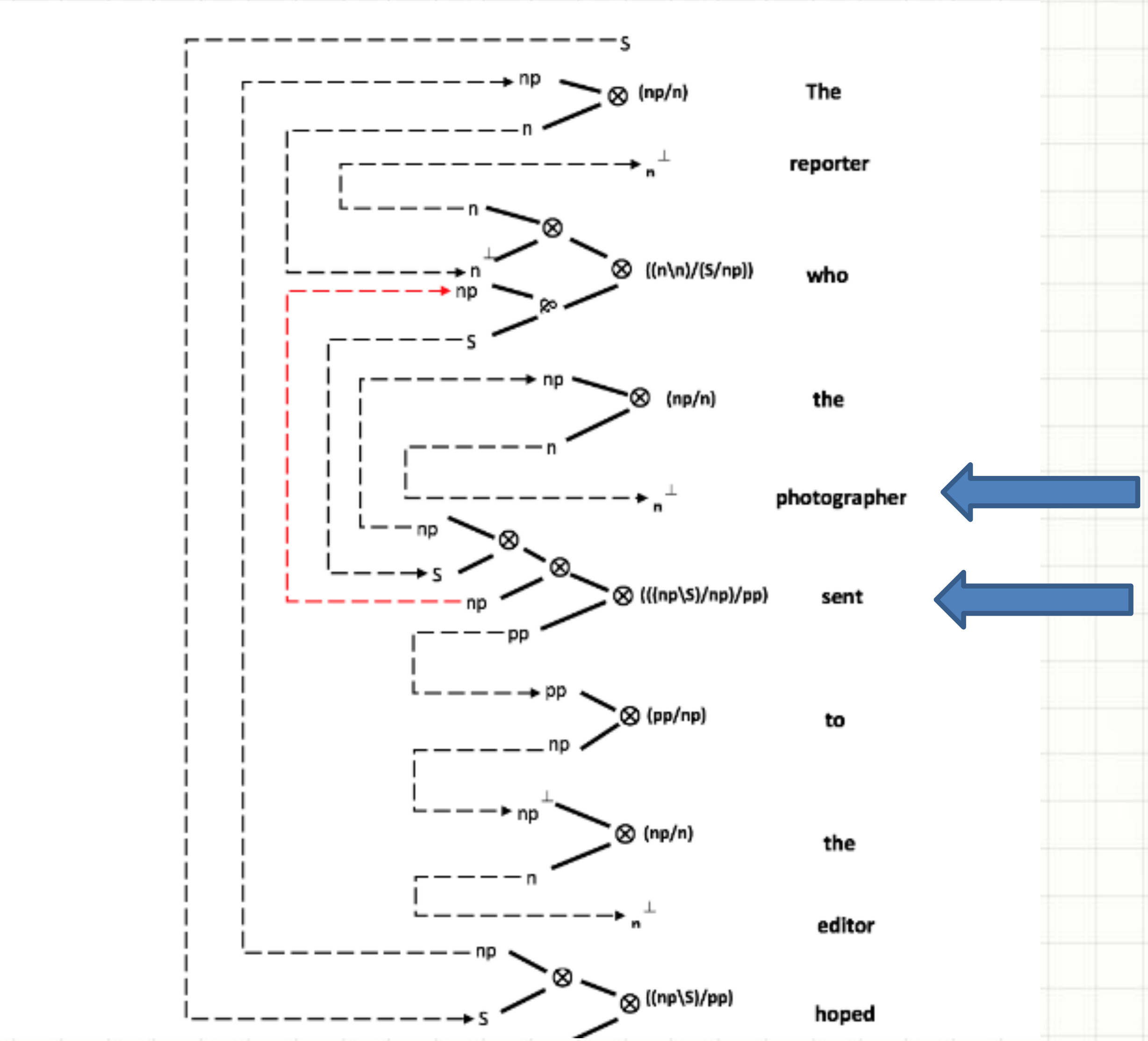
Subject/Object-extracted Relative Clauses



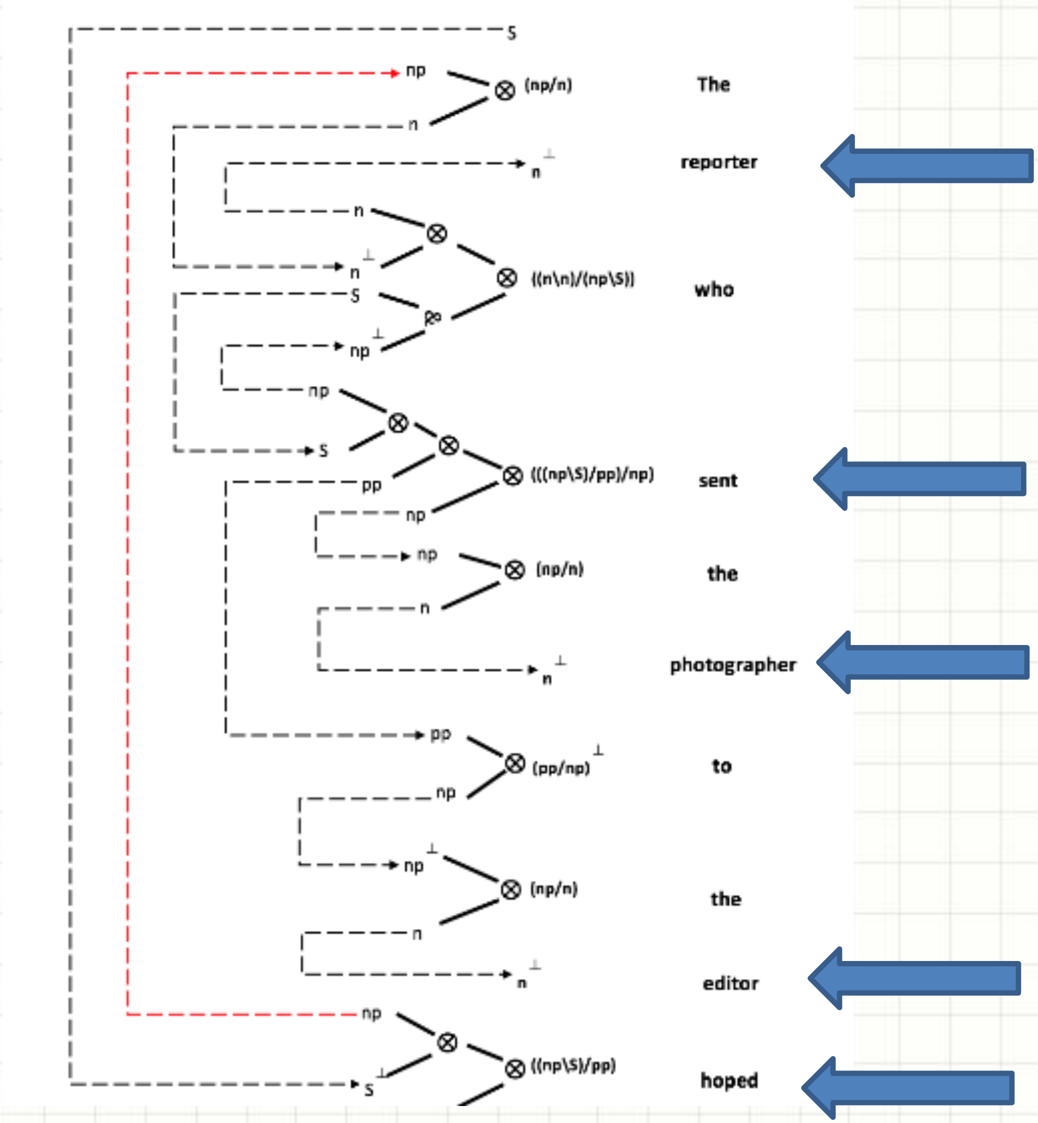
Subject/Object-extracted Relative Clauses



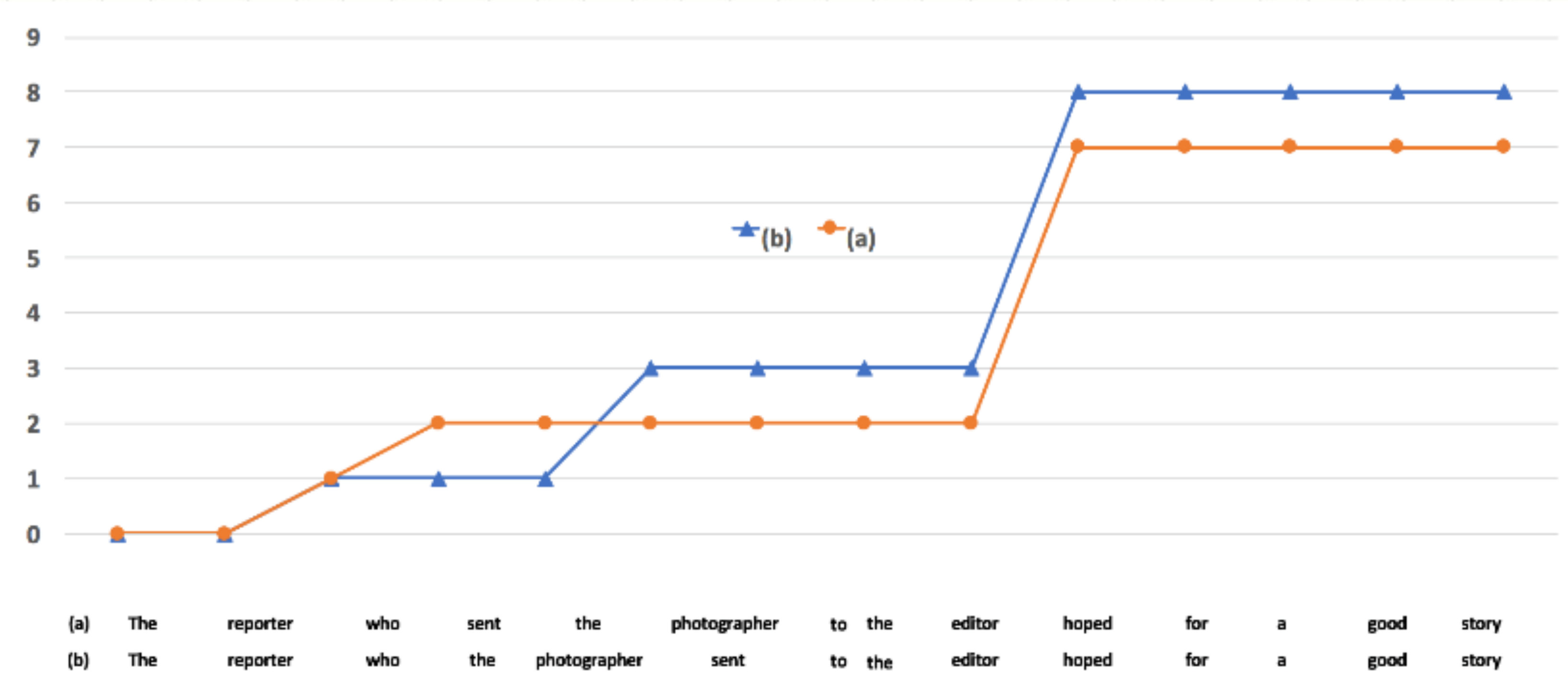
Subject/Object-extracted Relative Clauses



Subject/Object-extracted Relative Clauses

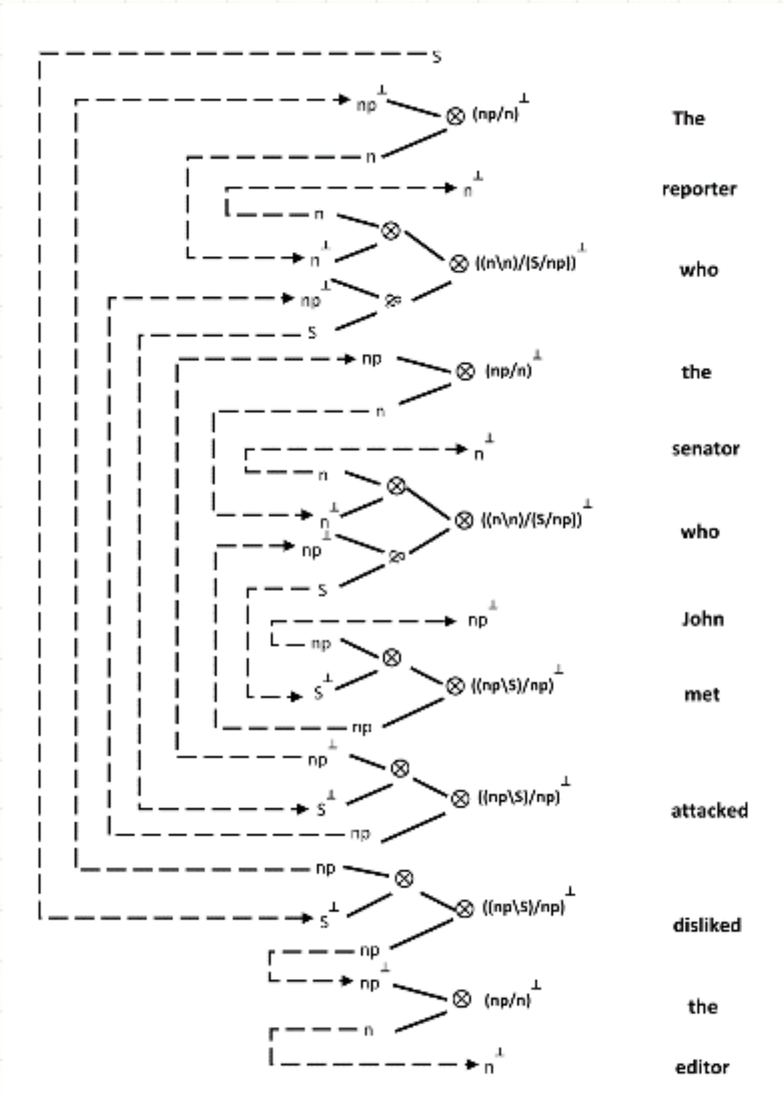


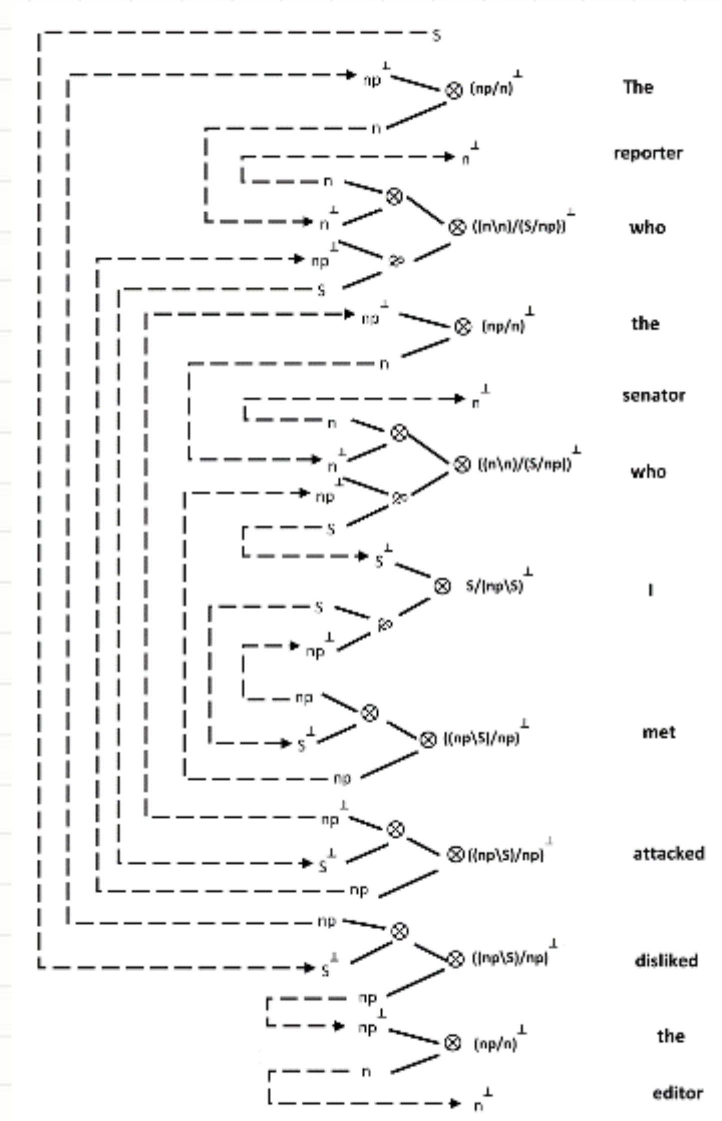
Subject/Object-extracted Relative Clauses



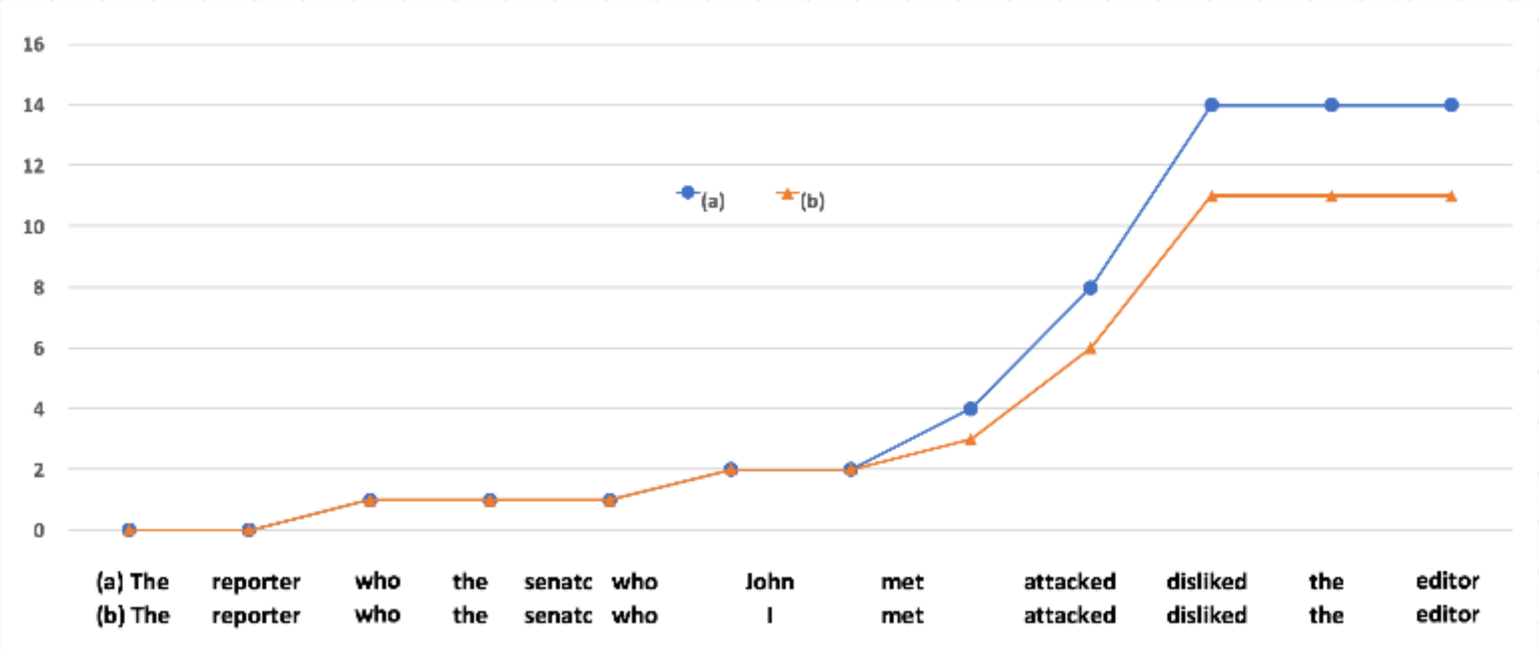
5.3a	x	The	reporter	who	sent	the	photographer	to	the	editor	hoped	for	a	good	story
	$ID(x)$	0	0	1	1	0	0	0	0	0	5	0	0	0	0
	AccSum(x)	0	0	1	2	2	2	2	2	2	7	7	7	7	7
5.3b	y	The	reporter	who	the	photographer	sent	to	the	editor	hoped	for	a	good	story
	$ID(y)$	0	0	1	0	0	2	0	0	0	5	0	0	0	0
	AccSum(y)	0	0	1	1	1	3	3	3	3	8	8	8	8	8

Center Embedding Clauses [Johnson, 1998]



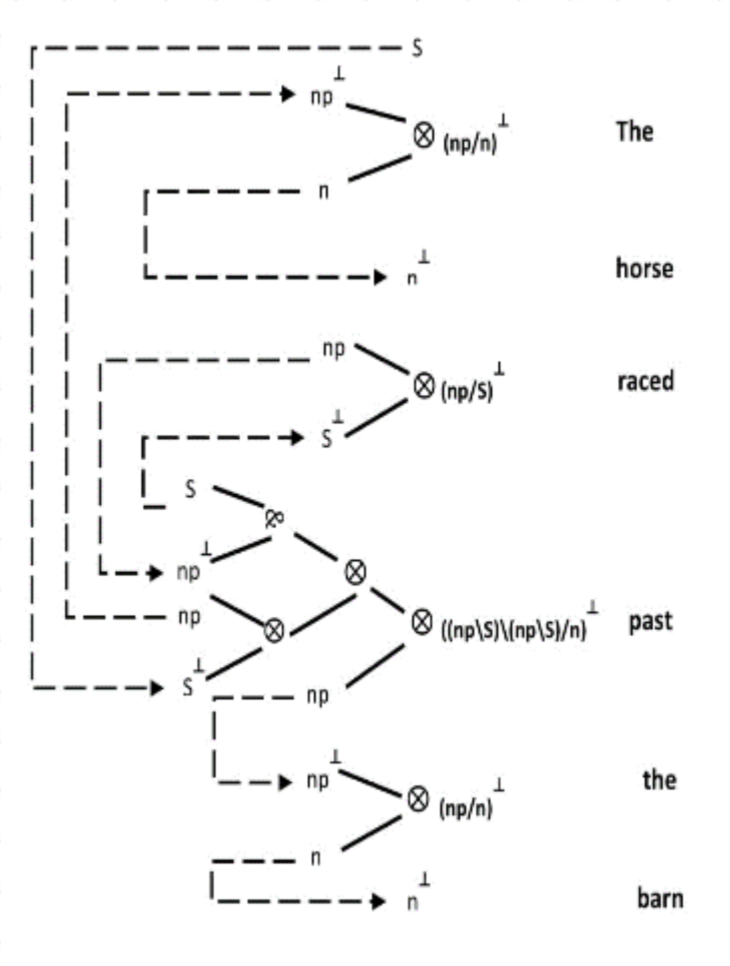


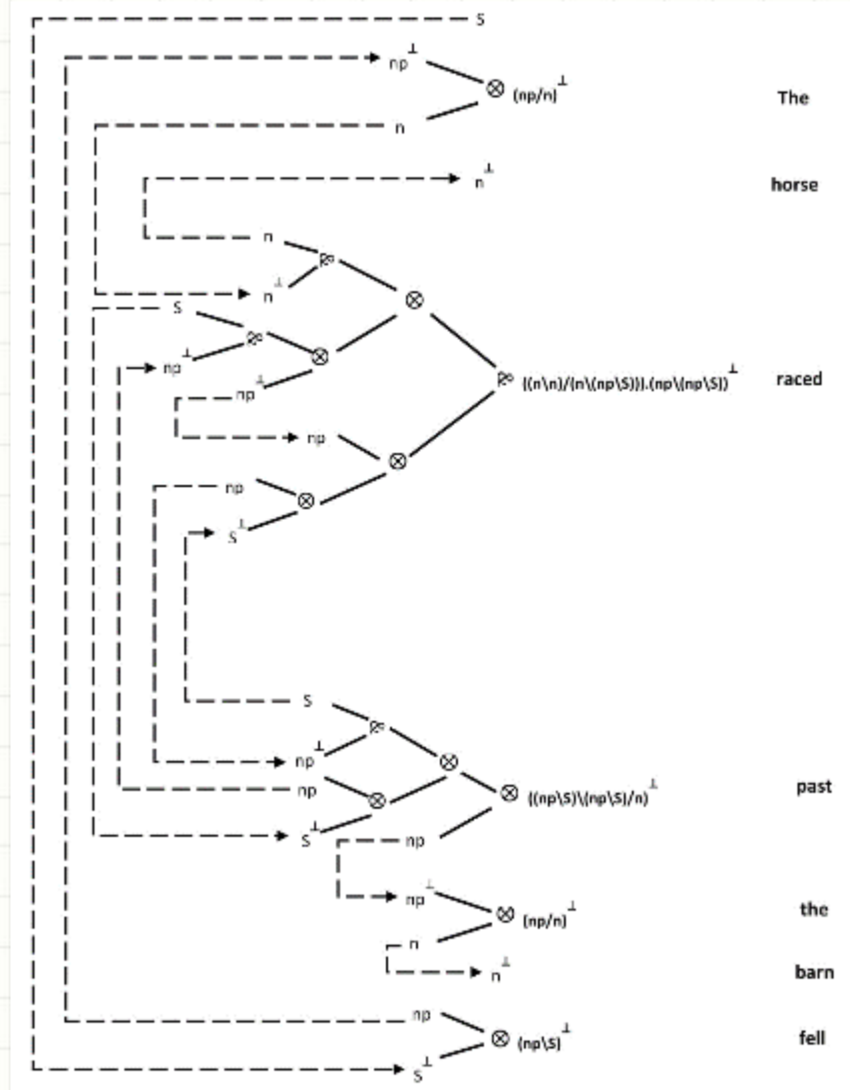
Center Embedding Clauses



5.4a	x	The	reporter	who	the	senator	who	John	met	attacked	disliked	the	editor
	$DL(x)$	0	0	1	0	0	1	0	2	4	6	0	0
	AccSum(x)	0	0	1	1	1	2	2	4	8	14	14	14
5.4b	y	The	reporter	who	the	senator	who	I	met	attacked	disliked	the	editor
	$DL(y)$	0	0	1	0	0	1	0	1	3	5	0	0
	AccSum(y)	0	0	1	1	1	2	2	3	6	11	11	11

Garden Path [Bever, 1997]

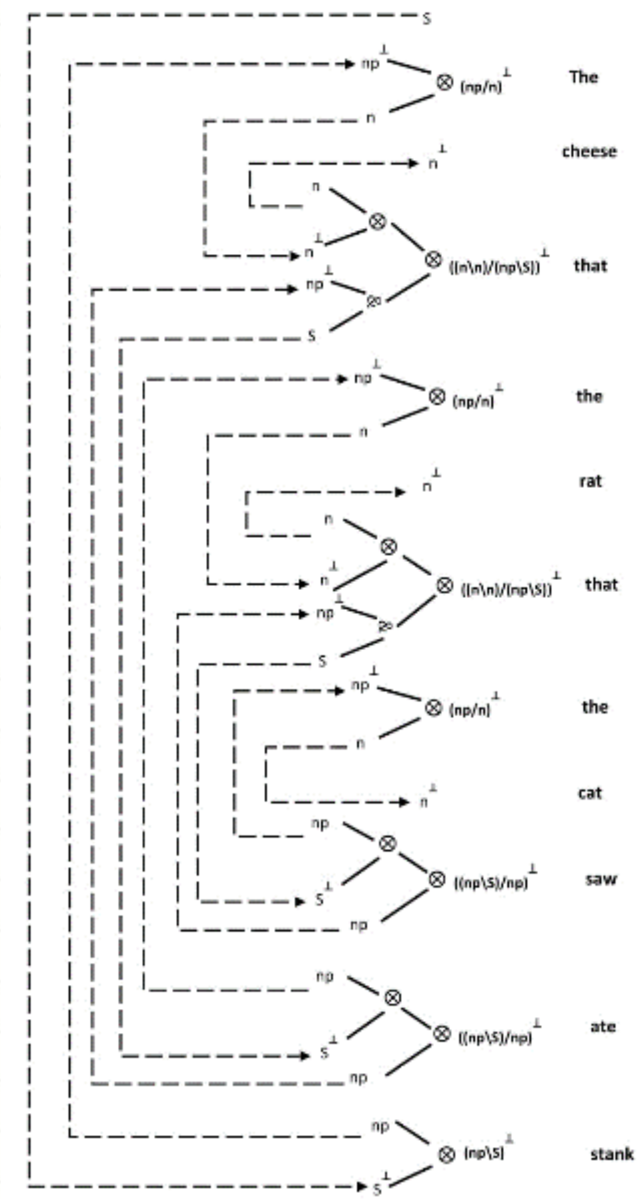


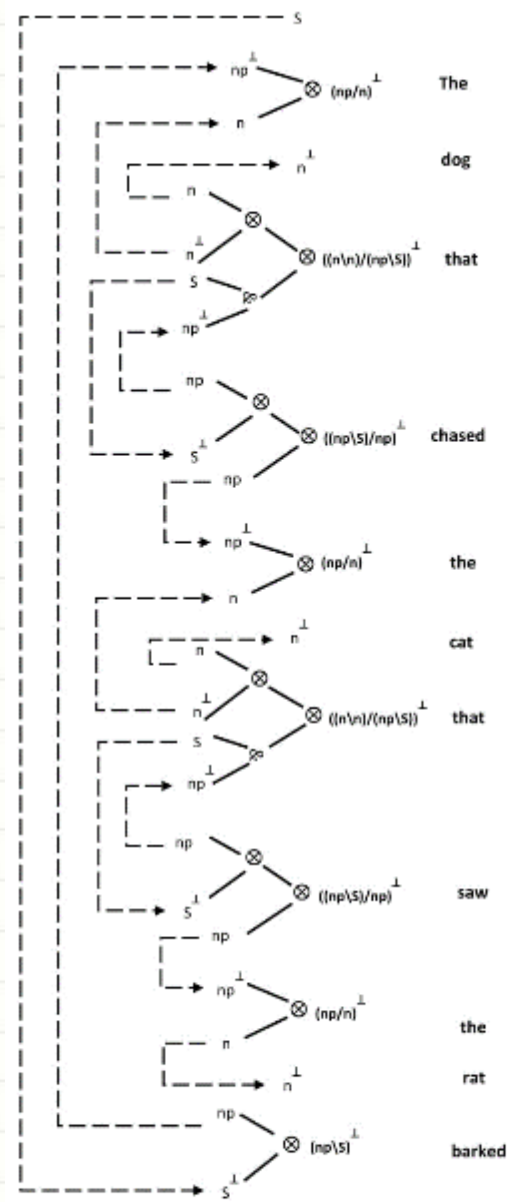


Garden Path

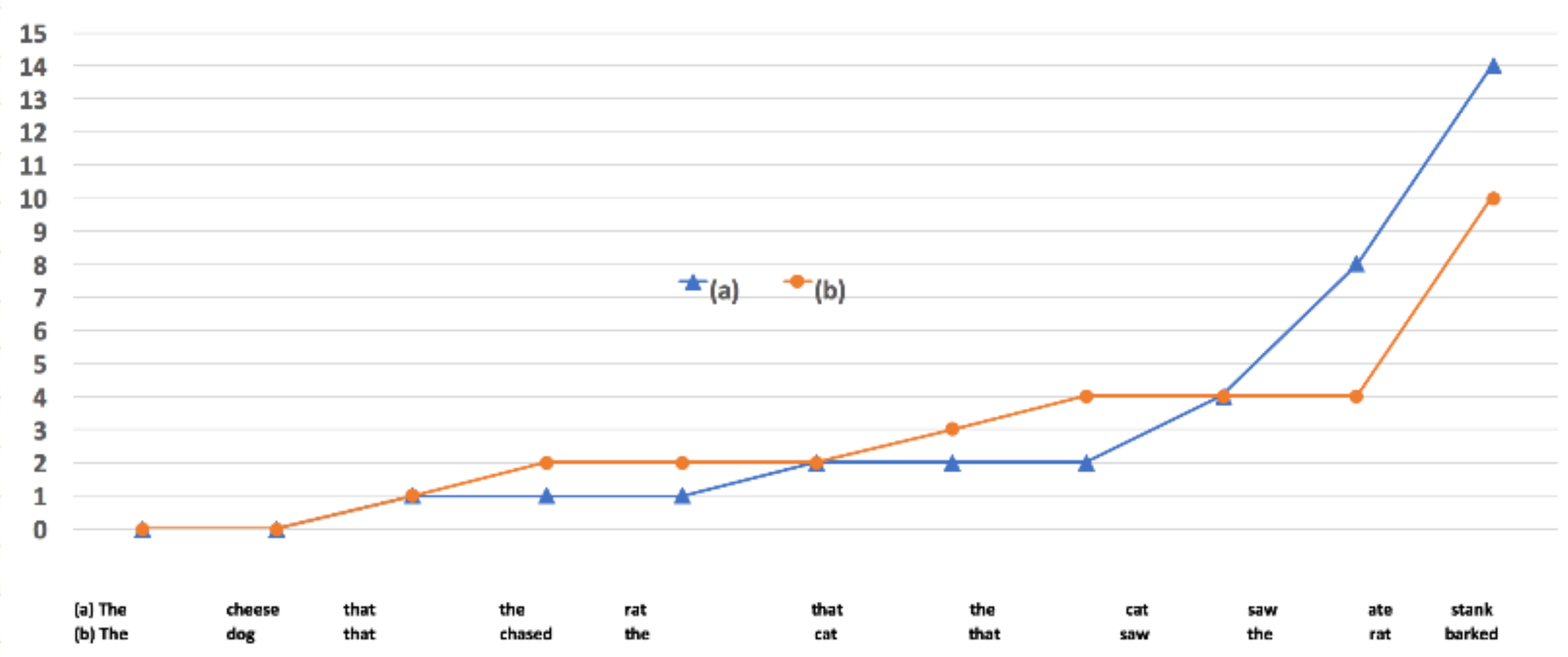
8
7
6
5
4
3
2
1
0
(a)
(b)
(a) The horse raced past the barn
(b) The horse raced past the barn fell

Nested Subject/Object Relativization [Chomsky, 1965]

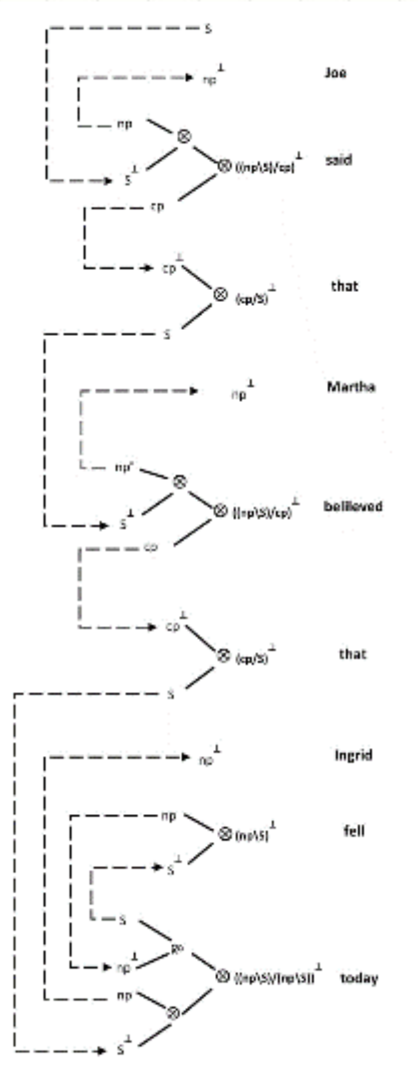


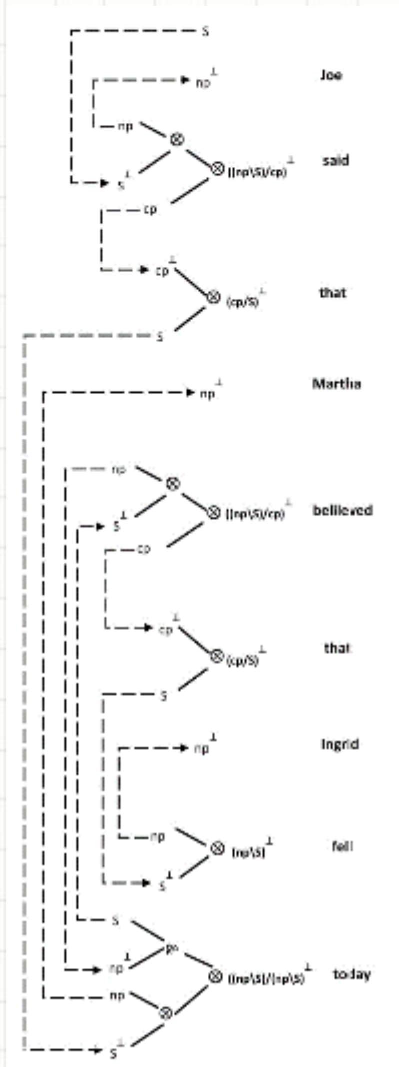


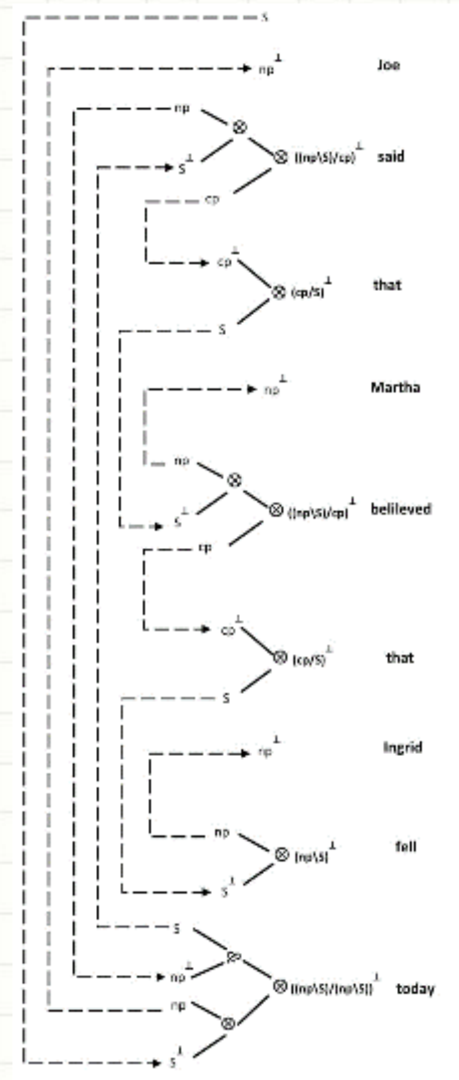
Nested Subject/Object Relativization



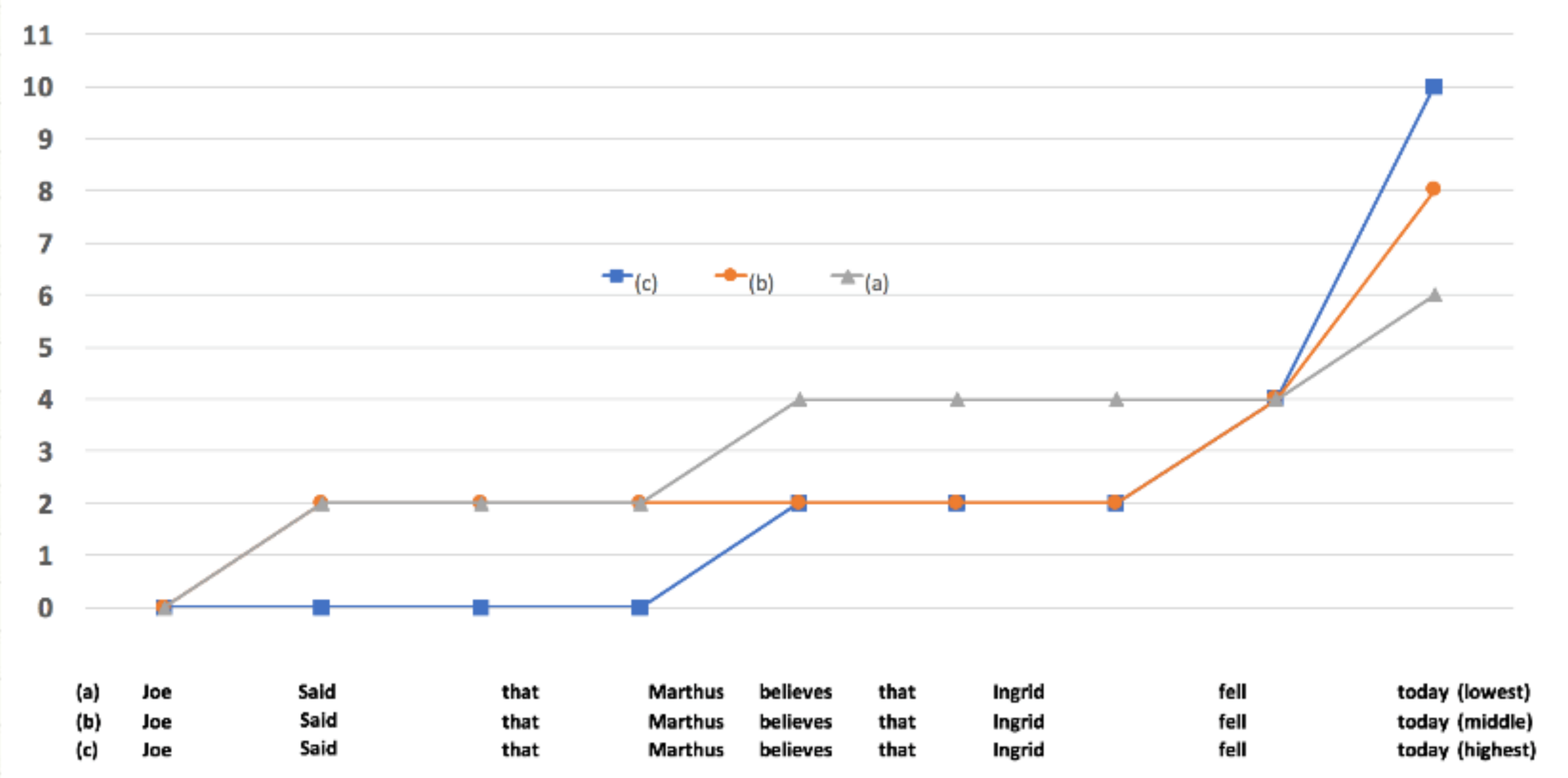
Adverbial Attachment [Kimball, 1973]



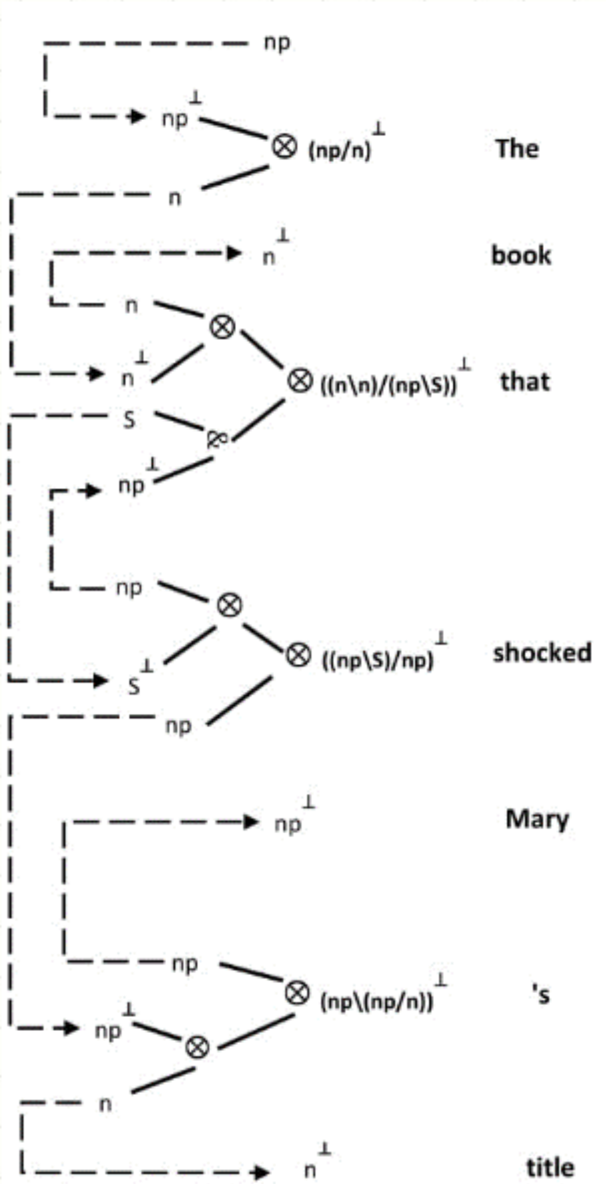


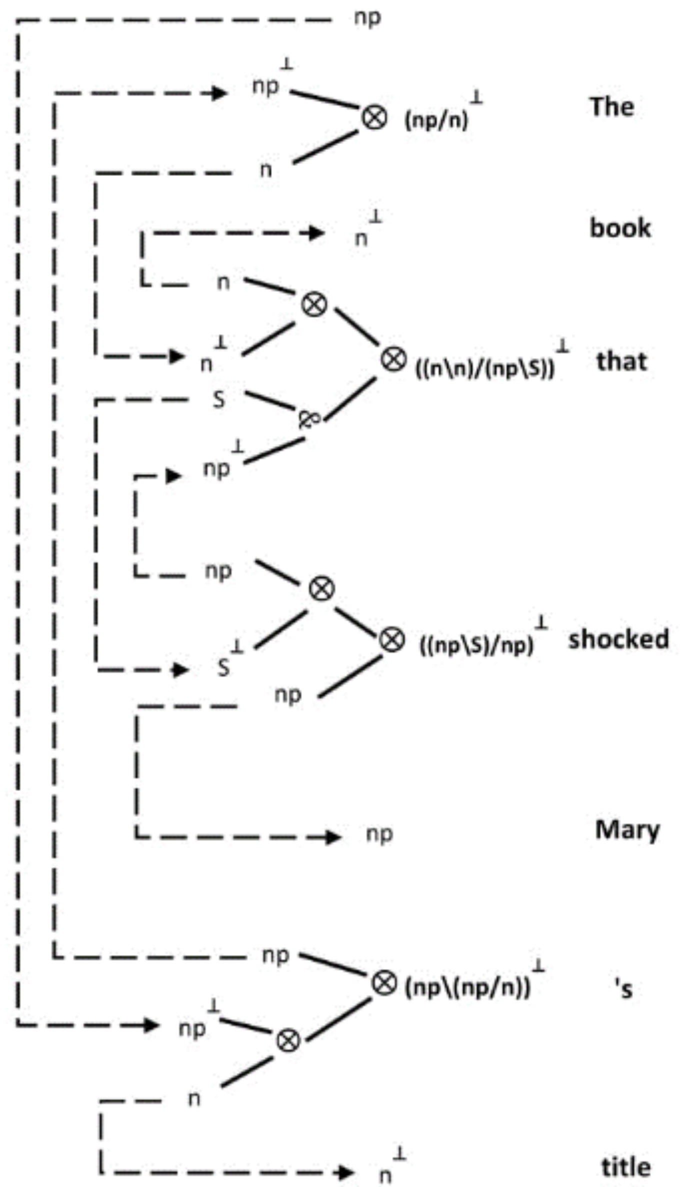


Adverbial Attachment

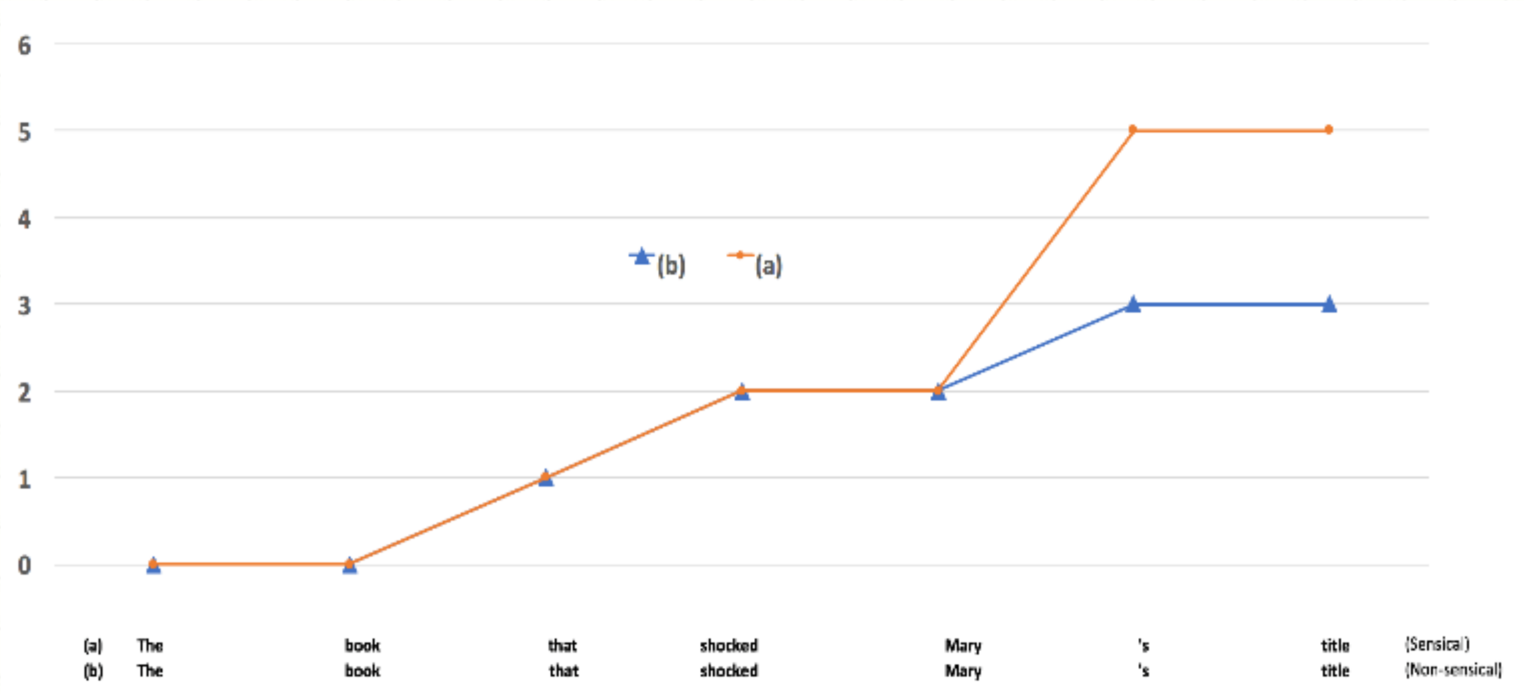


Wrong Parse Preference [Morrill, 2000]

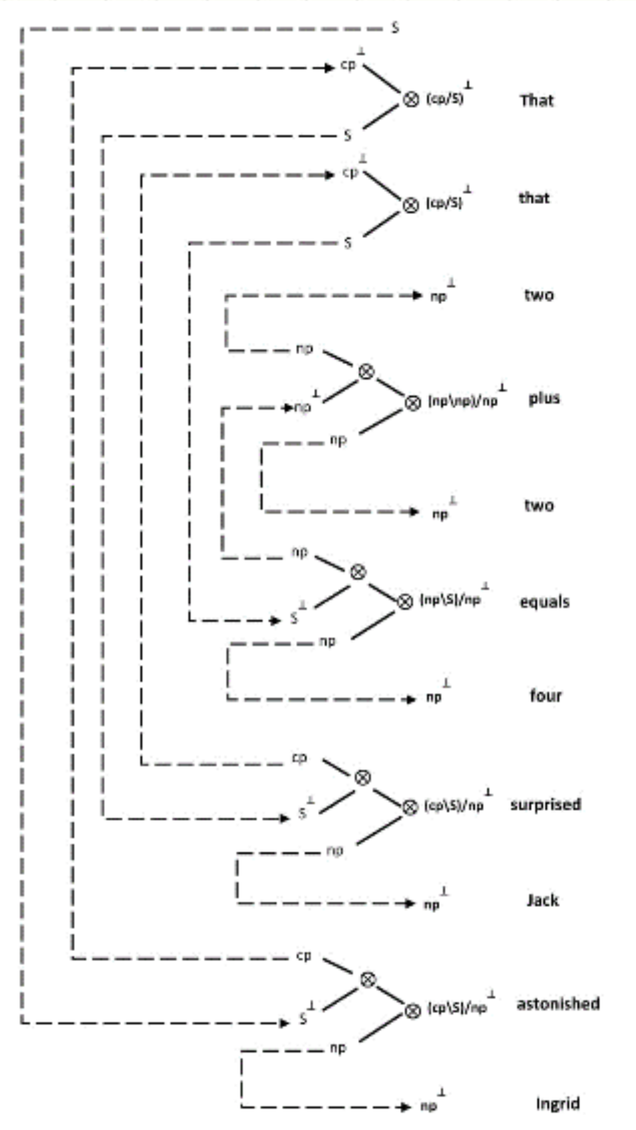


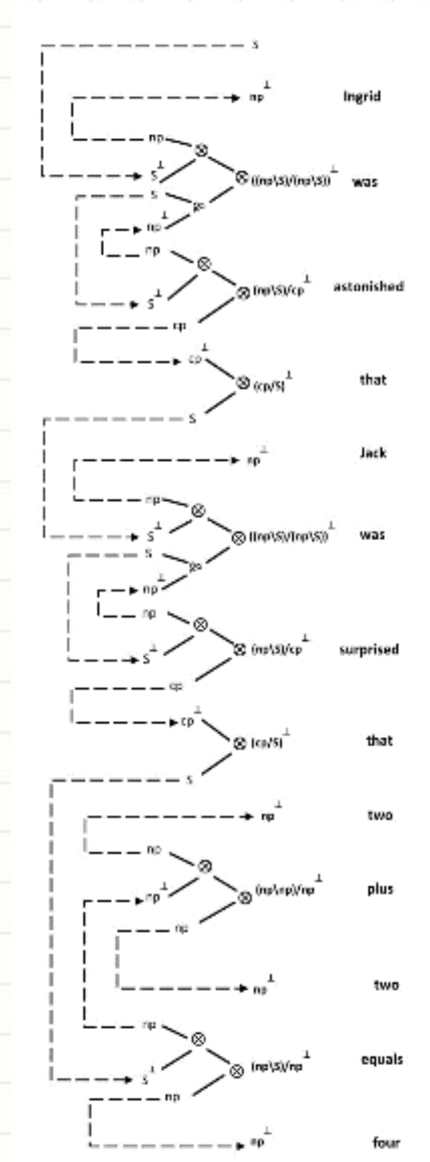


Wrong Parse Preference

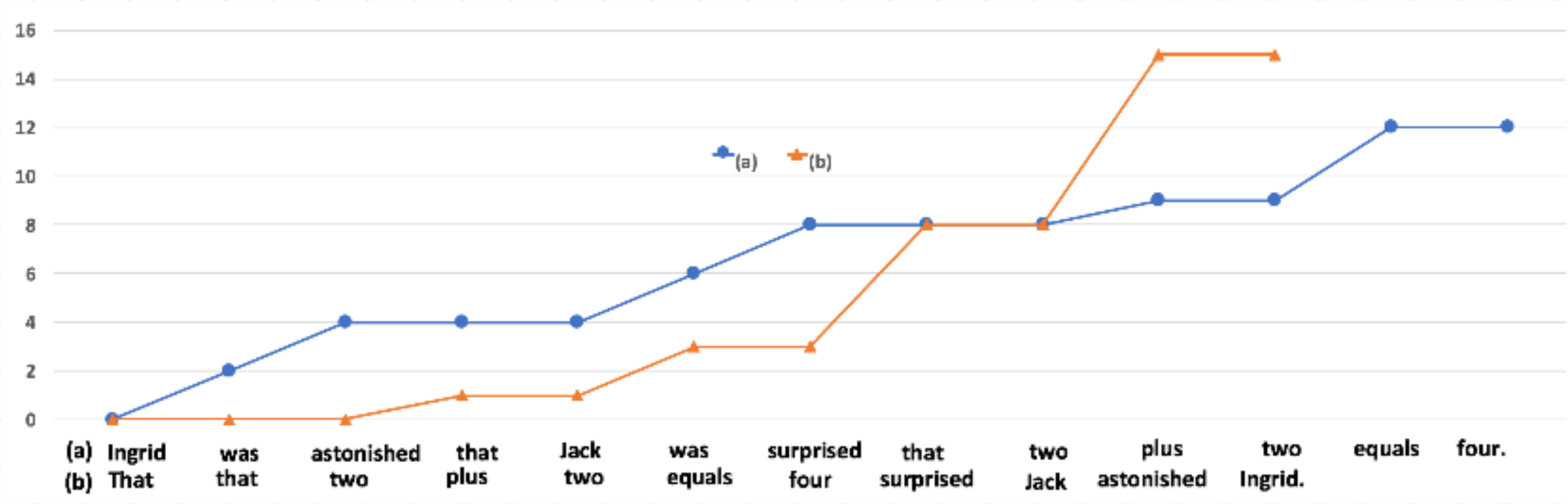


Passive Paraphrases [Morrill, 2000]

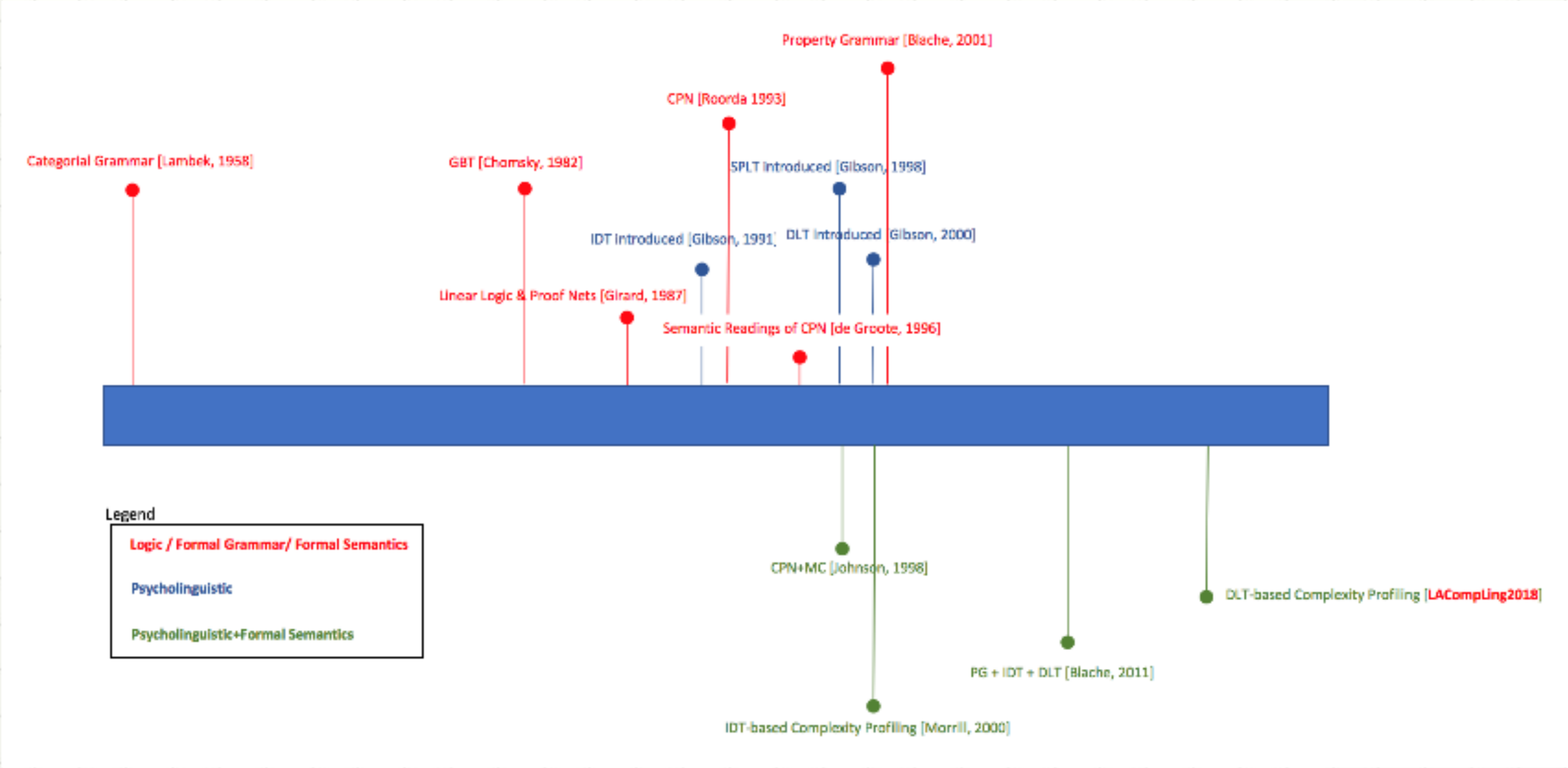




Passive Paraphrases



Big Picture:



Fair Warning: This is just a limited part of the historical line that one could work. There are definitely many interesting research that needs to be explored. We are aware of some of them and they should be even more than what we have noticed.

Limitations:

- DLT-based Complexity Profiling cannot correctly predict ranking the quantifier scoping problem.
- In fact, both IDT-based and DLT-based Complexity Profiling have this problem. [Catta, Mirzapour, 2017]
- DLT-based motivated approaches are not applicable cross-linguistically for human parsing processes. [Vasishth, 2005]
- It does not support all linguistic preference phenomenon such as *Heavy Noun Phrase Shift* while IDT-based Complexity Profiling does.

On-going Work for Overcoming the Limitations:

- Quantifier Scoping Problem. [Mirzapour, PhD, Chapter 3]
- Cross-linguistically Applicability [?, No Idea]
- Scale-up Problem [Mirzapour, PhD, Chapter 7]

Conclusion:

- DLT-based Complexity Profiling can successfully predict some linguistic phenomena such as structures with embedded pronouns, garden paths, unacceptability of center embedding, preference for lower attachment, and passive paraphrases acceptability.

- It is a kind of psycholinguistics motivated preference modeling along with the formal/lexical constructions of meaning.

Reference 1/2:

Blache, P.: A computational model for linguistic complexity. In: Proceedings of the first International Conference on Linguistics, Biology and Computer Science (2011)

Blache, P.: Evaluating language complexity in context: New parameters for a constraint- based model. In: CSLP-11, Workshop on Constraint Solving and Language Processing (2011)

Catta, D., Mirzapour, M.: Quantifier scoping and semantic preferences. In: Proceedings of the Computing Natural Language Inference Workshop (2017)

Chatzikyriakidis,S.,Pasquali,F.,Retoré,C.:Fromlogicalandlinguisticgenericstohilbert's tau and epsilon quantifiers. IfCoLog Journal of Logics and their Applications 4(2), 231–255 (2017)

Gibson,E.,Ko,K.:Anintegration-based theory of computational resources in sentence comprehension. In: Fourth Architectures and Mechanisms in Language Processing Conference, University of Freiburg, Germany (1998)

Gibson, E.: Linguistic complexity: Locality of syntactic dependencies. Cognition 68(1), 1– 76 (1998)

Gibson, E.: The dependency locality theory: A distance-based theory of linguistic complex-ity. Image, language, brain pp. 95–126 (2000)

Gibson,E.A.F.:Acomputationaltheoryofhumanlinguisticprocessing:Memorylimitations and processing breakdown. Ph.D. thesis, Carnegie Mellon University Pittsburgh, PA (1991)

Reference 2/2:

Girard,J.Y.:Linearlogic.TheoreticalCcomputerScience50,1–102(1987)

Johnson,M.E.:Proofnets and the complexity of processing center-embedded constructions.

In: Retoré, C. (ed.) Special Issue on Recent Advances in Logical and Algebraic Approaches to Grammar. Journal of Logic Language and Information, vol. 7(4), pp. 433–447. Kluwer (1998)

Lambek, J.: The mathematics of sentence structure. The American Mathematical Monthly 65(3), 154–170 (1958)

Mirzapour,M.:Findingmissingcategoriesinincompleteutterances.In:24eConférencesur le Traitement Automatique des Langues Naturelles (TALN). p. 149

Moot, R., Retoré, C.: The logic of categorial grammars: a deductive account of natural lan- guage syntax and semantics, vol. 6850. Springer (2012)

Moot, R., Retoré, C.: The logic of categorial grammars: a deductive account of natural language syntax and semantics, LNCS, vol. 6850. Springer (2012), http://www.springer.com/computer/theoretical+computer+science/book/978-3-642-31554-1

Morrill,G.:Incremental processing and acceptability.Computationallinguistics26(3),319– 338 (2000)

Retoré, C.: Calcul de Lambek et logique linéaire. Traitement Automatique des Langues 37(2), 39–70 (1996)

Roorda,D.:ProofnetsforLambekcalculus.LogicandComputation2(2),211–233(1992)

Shravan Vasishth et al. “Quantifying Processing Difficulty in Human Language Processing”. In: *In Rama Kant Agnihotri and Tista Bagchi* (2005).

Thanks For Your Attention