Evaluation and comparison of the RASP system and the Stanford unlexicalized PCFG parser

Alodie Boissonnet
Queens' College, University of Cambridge
avmb2@cam.ac.uk

Abstract

Natural language parsers aim to determine the grammatical structure of sentences. This is very useful in Natural Language Processing as it helps to extract information out of a sentence. In this work, we explore two popular parsers: the RASP system and the Stanford unlexicalized Probabilistic Context-Free Grammar (PCFG) system. While the RASP system establishes grammatical relations (GRs) between words, the Stanford tool focuses on phrase structure and constructs dependency relations. We compare these two parsing methods and analyse their errors. We find that the unlexicalized PCFG parser is more robust to common sources of errors, such as ambiguous Named-Entities or verbal forms used as non-clausal modifiers.

1 Introduction

Sentence parsing is a common task in Natural Language Processing (NLP) today to determine the structure of sentences. The latter is helpful to extract information out of sentences. For instance, we can extract noun phrases out of sentences to understand their topic. Parsing algorithms are divided into two main categories, which are the constituency parsers and the dependency parsers.

Constituency parsers attempt to assign a syntactic structure to sentences. They rely on a context-free grammar (Chomsky, 1956; Backus, 1959) that models how consecutive words can act as a constituent. To achieve this, a context-free grammar gathers different production rules to combine symbols together. These symbols are called terminal when they match words of the sentence, and non-terminals when they correspond to groups of terminal symbols. Constituency parsers then use parsing algorithms to output constituent trees, that highlight the hierarchical structure of sentences. Many parsing algorithms for CFG have already been developed. Among them, we can find the Cocke-Kasami-Younger (CKY) algorithm (Kasami, 1965; Younger, 1967) with a dynamic programming approach, or the Earley algorithm (Earley, 1970).

Constituency parsers suffer from coordination and attachment ambiguity. To solve this problem, probabilistic constituency parsers have been introduced. They use a Probabilistic Context-Free Grammar (PCFG) (Booth, 1969), that associates a probability to each production rule. Then, they compute the probability of different interpretations of a sentence and select the most probable one. Parsing algorithms used for PCFG are usually extensions of CFG parsing algorithms, such as the probabilistic CKY (Essen and Ney, 1991) or the probabilistic Earley algorithm (Stolcke, 1995).

The previous parsing methods based on a grammar require heavy rules when it comes to long-distance dependencies, sub-categorization of verbs or constituents agreement on case, gender or even number. A solution is to make better use of the lexicon. This approach is called *lexicalized* parsing, by opposition with *unlexicalized* parsing. A popular *lexicalized* grammar is the Combinatory Categorial Grammar (CCG) (Steedman, 1997, 2004) that assigns categories to words and combines them with a combinatory logic. The Charniak and Collins parsers (Charniak, 1997; Collins, 1999) are common parsing algorithms for CCG, built as extensions of the probabilistic CKY algorithm.

The second common category of parsing algorithms is the dependency one. A dependency grammar encodes the syntactic structure of sentences with directed binary grammatical relations. Grammatical relation labels are chosen from a finite set of grammatical relations. Phrase structure obtained with this kind of grammar is called a typed depen-

dency structure. Dependency parsing presents the advantage of being able to deal with more complex languages with free word order. Additionally, semantic relationships between words can be inferred from grammatical relations. Such semantic relationships can be of great importance for different NLP tasks, such as coreference resolution, information extraction or question answering. Relations are also more easily understandable by people without linguistic expertise.

Grammatical relations establish a binary relation between a "head" word and a "dependent" word which modifies the head. The head of a constituent is the word that determines the syntactic category of the constituent and around which the constituent is organized. The other words in the constituent are direct or indirect dependents of the head. Therefore, grammatical relations identify these relations between a head and its dependent words, and assign a grammatical function to each relation that makes explicit the role of the dependent over the head. A dependency analysis of a sentence produces a dependency graph or a dependency tree that can be described with different schemes, such as the CoNLL (Buchholz and Marsi, 2006), Stanford Dependencies (de Marneffe and Manning, 2008), Universal Dependencies (Nivre et al., 2016) or RASP (Briscoe et al., 2006) schemes.

On one hand, a transition-based parsing algorithm creates a dependency tree thanks to a greedy stack-based approach called shift-reduce. This bottom-up approach relies on a context-free grammar, a stack and a sentence to be parsed. It looks for sequences of words that correspond to the right-hand side of a grammar production rule and replaces them with the left-hand side until the whole sentence is reduced to a final state. To achieve this, two transition operations are used: shifting, to pushed the current symbol to the stack, and reducing, to replace symbols thanks to the grammar rules. On the other hand, graph-based dependency parsers factor the possible dependency trees of sentences into directed graphs. Sentence parsing is then transformed into a searching problem in a graph. The Maximum Spanning Tree (MST) algorithm is commonly used to find the optimal solution.

In this report, we explore two parsers, the RASP system and the unlexicalized PCFG parser released by the Stanford NLP group. In the first section, we study their underlying parsing algorithm. Then, we specify the methodology of this work and the data used. The next section presents the methods used to evaluate the output of the parsers. Finally, we examine the errors made by each parser and compare their performance.

2 Overview of methods

In the following section, we present two parsing algorithms: the RASP system and the unlexicalized PCFG parser released by the Stanford NLP group. We examine their structure and their analysis mechanism.

2.1 The RASP system

The Robust Accurate Statistical Parsing (RASP) system has been introduced by Briscoe and Carroll in 2002 as a domain-independent approach to statistical parsing. This parsing algorithm can be described as a series of different modules, gathered in a pipeline (Briscoe et al., 2006). Therefore, the system architecture is comprised of a tokeniser, a part-of-speech (PoS) tagger, a lemmatiser, a grammatical parser and a statistical disambiguator, as illustrated in Figure 1.

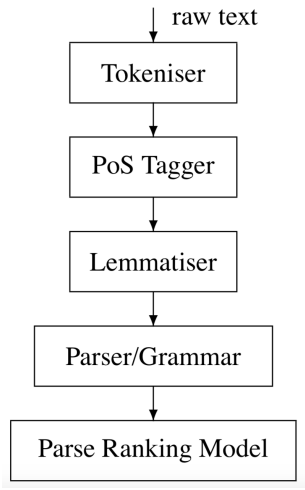

Figure 1: Pipeline architecture of the RASP system (Briscoe et al., 2006).

To begin with, the tokeniser has as its objective to transform raw and unannotated text into a sequence of tokens in which punctuation is separated from adjacent words and sentence boundaries are marked. It relies on a set of deterministic finite-state rules, which makes it easy to fix problems by adding new rules.

The tokenised text is then given as input to a PoS tagger. The latter assigns CLAWS-2 part-

```
Kim and Sandy are best friends.
(|Kim:0_NP1| |and:1_CC| |Sandy:2_NP1| |be+:3_VBR| |best:4_JJT| |
friend+s:5_NN2| |.:6_.|) 1 ; (-10.492)
gr-list: 1
(|ncsubj| |be+:3_VBR| |and:1_CC| _)
(|xcomp| _ |be+:3_VBR| |friend+s:5_NN2|)
(|ncmod| _ |friend+s:5_NN2| |best:4_JJT|)
(|conj| |and:1_CC| |Kim:0_NP1|)
(|conj| |and:1_CC| |Sandy:2_NP1|)
weighted: 1
 1.0            (|conj| |and:1_CC| |Sandy:2_NP1|)
 1.0            (|conj| |and:1_CC| |Kim:0_NP1|)
 1.0            (|ncmod| _ |friend+s:5_NN2| |best:4_JJT|)
 1.0            (|xcomp| _ |be+:3_VBR| |friend+s:5_NN2|)
 1.0            (|ncsubj| |be+:3_VBR| |and:1_CC| _)
```

Figure 2: Two possible outputs of the RASP system.

of-speech and punctuation tags (Sampson, 1995) to tokens, thanks to a first-order Hidden Markov Model (HMM) (Elworthy, 1994). A probabilistic unknown word model tags rare words with a special token. The PoS tagger is also able to return a set of more probable tags per token, in order to be able to trade-off precision against recall. To achieve this, it uses a Forward-Backward algorithm (Baum, 1972) in complement of the Viterbi algorithm (Viterbi, 1967) when predicting tags.

A morphological analysis is then conducted on tokens in order to extract lemmas and inflectional affixes from words. The lemmatiser uses about 1400 finite-state rules as part of a deterministic finite-state transducer. This task is necessary for the next modules of the pipeline to make them use the lexical information associated with lemma forms and affixes. However, lemmatisation suffers from unknown words and incorrect tokenisation.

The next step consists of parsing token sequences. A probabilistic generalized LR parser analyses tokens with their PoS to generate a parse forest containing all possible subanalyses with an associated probability. It determines the probability of some sequences of token, such as phrasal verb combinations, and encodes the conditional probability of high to mid frequency verbs. To achieve this, it uses 400 unification-based phrase structure rules in order to enumerate possible valencies for predicates.

Parsing trees can finally be output in numerous different formats, among them the n-best syntactic trees and weighted grammatical relations (GRs). For the n-best parse tree output, tokens are labeled with rules names or category names, and PoS tags are the leaves of trees. Grammatical relations display a named relation with its head and its dependent, and possibly some parameters specifying the relation. Regarding weighted GRs, GRs does not define a complete and consistent directed graph of relations, but it includes alternative weighted GRs corresponding to competing subanalyses. The weights are computed by examining the proportion of subanalyses including a given GR and their probability. Figure 2 shows these two different output formats obtained with the RASP system. In this work, we only study the grammatical relations of the best parsing tree.

2.2 The Stanford unlexicalized PCFG parser

The Stanford unlexicalized PCFG parser has been released by the Stanford NLP group in 2003 (Klein and Manning, 2003). This parser has been introduced to demonstrate that an unlexicalized PCFG parser can perform as well as lexicalized parsers, while having many advantages. It is easier to interpret and more widely understood. Thanks to its much more compact grammar, it also has a lower asymptotic complexity and is much simpler to build and optimize.

To achieve this, Klein and Manning (2003) started from a generalized CKY parser optimized by maximum-likelihood estimation, that they improved little by little. First, an unlexicalized parsing analysis relies on a grammar built on training trees. This grammar suffers from insufficiently detailed category symbols, which prevents it from making good use of the context. Indeed, subject-NP expansions should be processed differently than object-NP expansions. A solution to this issue is to use parent annotation that helps to capture some external context. Another problem with this basic grammar is that many production rules have been seen very rarely and that test sentences include production rules that have not been seen before. Collins (1999) suggests markovizing rules from the head child to capture the linguistic idea that phrases are organized around a head. Figure 3 provides an example of this new annotation strategy. These two methods to improve node annotations lead to consider a longer vertical history and a shorter horizontal history. A trade-off has to be found between a compact grammar and an increased history.

In the same vein as parent annotations for non-terminal symbols, annotating the parent of PoS tags is very helpful. Indeed, even though most tags have a canonical category, their analysis can be misled when they appear in a non-canonical position. For example, adverbs distribution varies whether adverbs are under an adverbial phrase (ADVP), a verbal phrase (VP) or a noun phrase (NP). Likewise,

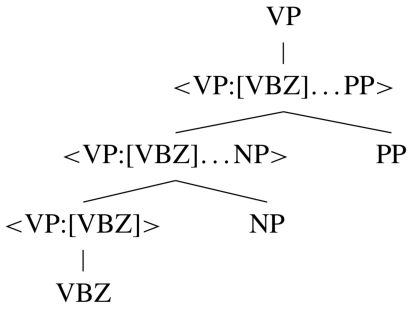

Figure 3: Example of the new annotation strategy for the parsing of VP → VBZ NP PP (Klein and Manning, 2003).

the Peen Treebank tag set (Marcus et al., 1993) does not make the difference between subordinating conjunctions (*while, as, if*), complementisers (*that, for*) and prepositions (*of, in, from*). For these reasons, annotating tags' parent is a valuable improvement of the basic CKY parsing algorithm.

Other annotation rules have also been added to the parser in order to capture the different behavior of constituents' head. Among these rules, we can find head annotation rules for possessive NPs, or head annotation rules for finite and non-finite VPs.

To finish, this PCFG parser is said to be unlexicalized as it doesn't make use of lexical class words to calculate monolexical or bilexical probabilities. It only relies on subcategorized PoS categories, that allow capturing words' behaviour in various configurations.

The output of this parser is in the CoNLL format (Buchholz and Marsi, 2006), which specifies, for each token, its part-of-speech tag, the token it depends on (if it exists) and the dependency relation between these two tokens.

3 Data and evaluation methods

3.1 Data

This work explores parsing algorithm outputs for different sentences, coming from two main sources. We first study in depth ten sentences suggested in the project handout. These sentences are very challenging for parsers as they gather many complex phrasal structures. In the next section, we will study what are the difficulties faced by parsers when analysing these sentences.

The other sentences suggested in the handout are also analysed in order to create gold standards to evaluate the parsers' performance quantitatively (see Section 3.2).

In addition, we will examine some sentences designed to assess parsers' accuracy on a precise aspect. In particular, I have chosen sentences that deal with NP composed of multiple modifiers, ambiguous Named-Entities, subject-verb inversion and verbal forms used as non-clausal modifiers. All these sentences are difficult to parse due to different kinds of ambiguities and are usually poorly analysed.

3.2 Gold standards

To evaluate the performance of parsers, it is necessary to compare parsers' outputs with gold standards. The latter has been obtained manually. For each sentence examined in this work, I have established gold standard analyses in the form of the RASP grammatical relations (Briscoe et al., 2006) and in the form of CoNLL dependencies (Buchholz and Marsi, 2006), to match the output formats of the two parsers under study. For the grammatical relations, gold standards have been established for only 10 sentences of the handout, while the 25 sentences of the handout have been parsed manually to create CoNLL gold standards. To limit human annotation errors, I have compared the gold standard I have written with the ones obtained by one of my classmates, Zébulon Goriely. Although these gold standards may contain some errors, it is still possible to identify mistakes made by parsers with good accuracy.

3.3 Evaluation metrics

Parsers' outputs can be evaluated with both quantitative and qualitative methods. In the case of the unlexicalized PCFG parser, outputs in the CoNLL format are compared to gold standards thanks to the MaltEval tool (Nilsson and Nivre, 2008). The latter breaks down a parser evaluation by dependency label, and computes, for each of them, the parser's precision, recall and F_1-score.

As for the RASP system, GRs can not be transformed into the CoNLL format easily, and thus, evaluate with the MaltEval tool. Therefore, I computed manually quantitative metrics to evaluate RASP outputs. I consider that a GR is correct if the head and all dependent slots are equal. If a relation is incorrect at a given level in the hierarchy, it can be correct higher in the hierarchy, for a subsuming relation. For example, if a `xcomp` relation is tagged as `dobj`, this relation is correct for all relations that subsume both `xcomp` and `dobj`, such as `ccomp` or `arg`. This definition of correctness

allows us to calculate a parser's precision, recall and F_1-score for each type of relation.

In addition to these three scores, we compute the micro-average and macro-average of each metric. While the macro-average scores are the mean of the scores of each relation, the micro-average ones calculate the average of the scores, weighted by the count of each label.

However, in this work, quantitative methods are not very reliable as they are based on very little data. For this reason, results will be mostly analysed with a qualitative point of view in the next section.

4 Results and discussion

Parsers' outputs for the sentences extracted from the project handout can be found in the Appendix A, with their gold standards. In the present section, we examine the errors made by each parser and make some generalisations about them.

4.1 Quantitative evaluation

Tables 1 and 2 present the results obtained for both parsers with the qualitative methods described in Section 3.3. Because the two parsers do not output results in the same format and do not use the same type of relations between words, it is difficult to compare the performance of both parsers. Some general comments can still be made. Notably, we can see that the unlexicalized PCFG parser is better at identifying the subject of sentences. While the RASP system has obtained a F_1-score of 80.04 for the `nbsubj` tag, the unlexicalized PCFG parser reaches a F_1-score of 93.5 for the `nsubj` tag. This difference is significant and suggests that the unlexicalized PCFG parser is better at determining the global structure of a sentence. However, when it comes to identifying verbs' objects, the RASP system is more efficient. It obtains a F_1-score of 79.31 for the `dobj` tag, while the unlexicalized PCFG parser reaches a F_1-score of 65.6 for the same category.

Another interesting figure is the F_1-score of 83.6 obtained by the unlexicalized PCFG parser for the `root` tag. It suggests that the parser is not able to find the root of about 20% of sentences, and therefore provides a quite poor parsing analysis. This figure is relatively high, but can be explained by the complexity of the sentences under study.

Finally, when comparing micro-average and macro-average precision, recall and F_1-scores, the unlexicalized PCFG parser achieves slightly better results. This quantitative analysis leads to the conclusion that both parsers have strengths and weaknesses, but it is not sufficient to evaluate and compare parsers output. The following qualitative analysis is therefore necessary and relevant.

4.2 Difficult tokenisation

One first difficulty faced by parsers is about tokenisation of sentences. It is possible to observe some wrong token identification, which obviously leads to an incorrect parsing. We support our claims with the two following sentences. Parsers' outputs for sentence 1 can be found in Appendix A, while the Figure 4 provides them for sentence 2.

1. The slightly simplified version of the Viterbi algorithm that we present takes as input a single HMM and a sequence of observed words $O = (o_1, o_2, ...o_T)$ **and returns** the most probable state/tag sequence $Q = (q_1, q_2, q_T)$ together with its probability.

2. Molly Andrews **Ph.D.** retired.

```
Mooly Andrews Ph.D. retired.

• RASP output:
(|Mooly:0_NP1| |Andrews:1_NP1| |Ph:2_NP1| |.:3_.|) 0 ; ()
(|ncmod| _ |Ph:2_NP1| |Mooly:0_NP1|)
(|ncmod| _ |Ph:2_NP1| |Andrews:1_NP1|)

(|d.:0_NNSA1| |retire+ed:1_VVD| |.:2_.|) 0 ; ()

• Unlexicalized PCFG parser output:
1       Molly     _       NNP     NNP     _       3       compound _      _
2       Andrews   _       NNP     NNP     _       3       compound _      _
3       Ph.D.     _       NNP     NNP     _       4       nsubj    _      _
4       retired   _       VBD     VBD     _       0       root     _      _
```

Figure 4: Example of parsers' output for sentence 2.

To begin with, these two sentences generate some errors about sentence boundaries detection. In the first sentence, the unlexicalized PCFG parser identifies two distinct fragments, with the second fragment starting from $O = (...)$. The RASP system also outputs two fragments for this sentence, but the second fragment starts from *and returns*. This error on sentence boundaries is probably linked with the difficulty of these algorithms to analyse sequences of symbols such as $O = (o_1, o_2, ...o_T)$, which contribute to an inconsistent parsing.

The tokenisation of the symbol sequences in sentence 1 can also be questioned. Both the parsers and the gold standard have assigned a unique token to each symbol, but it can also be argued that such symbol sequences form only one token. This tokenisation of symbol sequences would help a lot to process the first sentence as it would make it much simpler.

Relation	Count	Precision	Recall	F_1
dep	238	76.85	60.75	73.13
aux	6	100	83.33	90.91
conj	30	80.0	53.33	64.0
ta	16	80.0	50.0	61.54
det	20	84.21	80.0	82.05
arg_mod	164	74.53	73.17	73.85
mod	75	76.19	64.0	69.57
ncmod	67	78.95	67.16	72.58
xmod	0	-	-	-
cmod	6	60.0	50.0	54.55
pmod	2	-	-	-
arg	85	72.83	78.82	75.71
subj_or_dobj	66	72.83	78.82	79.37
subj	23	81.82	78.26	80.0
ncsubj	22	80.95	77.27	79.07
xsubj	1	100	100	100
csubj	0	-	-	-
comp	62	70.0	79.03	74.24
obj	53	77.78	79.25	78.50
dobj	43	84.21	74.42	79.01
obj2	1	50.0	100	66.67
iobj	8	66.67	100	80.0
clausal	6	35.71	83.33	50.0
xcomp	5	50.0	80.0	61.54
ccomp	1	16.67	100	28.57
pcomp	1	-	-	-
macro-average	-	71.85	77.13	71.58
micro-average	-	76.37	72.13	73.80

Table 1: Results obtained for the RASP parser, broken down by grammatical relation label.

Relation	Count	Precision	Recall	F_1
advcl	10	10	100	18.2
advmod	37	83.8	91.2	87.3
amod	45	95.6	87.8	91.5
appos	5	20	7.7	11.1
aux	15	100	71.4	83.3
aux:pass	7	85.7	100	92.3
cc	29	89.7	96.3	92.9
ccomp	10	50	83.3	62.5
compound	28	75	80.8	77.8
compound:prt	3	100	75	85.7
conj	29	93.1	67.5	78.3
cop	10	70	77.8	73.7
csubj	2	50	100	66.7
dep	31	0	-	-
det	52	98.1	100	99
det:predet	1	100	100	100
discourse	2	0	-	-
expl	2	100	100	100
fixed	5	20	20	20
iobj	3	66.7	100	80
mark	20	30	85.7	44.4
nmod:poss	9	100	69.2	81.8
nsubj	54	92.6	94.3	93.5
nsubj:pass	7	85.7	100	92.3
nummod	9	100	100	100
obj	31	67.7	63.6	65.6
obl:tmod	1	100	50	66.7
parataxis	3	66.7	28.6	40
pobj	30	90	43.5	58.7
prep	61	93.4	89.1	91.2
punct	6	100	75	85.7
rcmod	9	100	60	75
root	28	82.1	85.2	83.6
xcomp	6	66.7	66.7	66.7
macro-average	-	73.02	77.18	73.92
micro-average	-	79.00	77.89	76.43

Table 2: Results obtained for the Stanford unlexicalized PCFG parser, broken down by dependency label.

For sentence 2, the RASP algorithm wrongly considers the first dot of the *Ph.D.* abbreviation as a full stop. This error is more surprising as title abbreviations are quite common in text. But this shows the limitation of parsers relying on hand-written rules: some particular cases may have not been captured by the existing rules.

4.3 Incorrect PoS tagging

Even with correct tokenisation of sentences, the pre-processing steps of data can mislead parsing algorithms. Notably, part-of-speech tags provide useful information for parsers, but these tags are sometimes incorrect. Over the 10 sentences of the handout examined in depth in this work, it appears that, on average, the RASP system and the unlexicalized PCFG parser accurately identify 94% and 84% of PoS tags respectively. These figures show that incorrect PoS tagging is a real issue for parsing algorithms.

In the sentence,

> ***It won't rain*** *but there might be snow on high ground if the temperature stays about the same over the next 24 hours.*

the unlexicalized PCFG parser assigned the tag `NN` to *rain*, while *rain* is actually the base form of a verb. Evidently, the parser was unable to analyse properly the sentence with these PoS tags (see Appendix A).

Likewise, in the sentence,

> *The slightly simplified version of the Viterbi algorithm that we present takes as input a* ***single HMM*** *and a sequence of observed words* $O = (o_1, o_2, ...o_T)$ *and returns the most probable state/tag sequence* $Q = (q_1, q_2, q_T)$ *together with its probability.*

the RASP system tags *single* as a base form of a verb and *HMM* as an interjection. It considers that *HMM* is a complement of the verb *single*, whereas the gold standard identifies *single* as an adjective modifying the Named-Entity *HMM*.

These two examples illustrate the fact that parsers suffer from an incorrect part-of-speech tagging during data pre-processing steps, which can mislead them in their sentence analysis.

4.4 Parentheses

Another difficulty faced by parsers is the use of parentheses or parenthetical hyphens. There is no grammatical rule that explains how parentheses relate to other words. In addition, such a phrasal construction can have several roles. For example, in the following sentence, it provides some details about a previous word (sentence 1), gives an example of a previous concept (sentence 2) or makes reference (sentence 3).

1. Natural disasters **– storms, flooding, hurricanes –** occur infrequently but cause devastation that strains resources to breaking point.

2. English also has many words of more or less unique function, including interjections **(oh, ah)**, negatives **(no, not)**, politeness markers **(please, thank you)**, and the existential 'there' **(there are horses but not unicorns)** among others.

3. Making these decisions requires sophisticated knowledge of syntax; tagging manuals **(Santorini, 1990)** give various heuristics that can help human coders make these decisions and that can also provide useful features for automatic taggers.

After parsing, it appears that these parenthetical constructions effectively cause both parsers difficulty (see Appendix A). In the first sentence, the RASP system considers that *storms* is a non-clausal modifier of *hurricanes*, *flooding* is linked to *storms* with a `ta` relation and only *hurricanes* is a non-clausal modifier of *disasters*. The gold standard rather considers that *storms* is a non-clausal modifier of *disasters*, and that *flooding* is linked to *storms* and *hurricanes* is linked to *flooding*, both with a punctuation. Likewise, the unlexicalized PCFG algorithm performs poorly and links *storms*, *flooding* and *hurricanes* with a similar relation to *disasters*.

In sentence 2, both parsers fail to accurately analyse the structures with parentheses. Indeed, most grammatical relations referring to parenthetical constructions are missing in the RASP output (see Appendix A). Thus, the latter does not specify any relation for *oh* or *please*. As for the unlexicalized PCFG parser, the three first parentheses are correctly analysed but it does not consider that the last one *(there are horses and not unicorns)* relates to the previous *there*, but rather to *many words* at the beginning of the sentence. This sentence highlights that parentheses contain a wide variety of

phrase structures, that makes it difficult to understand the relation between parentheses and their head words.

Finally, sentence 3 is very interesting as both parsers conducted a correct analysis of the parentheses. This kind of parenthetical structure is more common as it corresponds to a standard way to make reference to a research paper. This example indicates that the RASP system may have a special grammatical rule to handle such a situation.

Overall, the RASP system struggles with parenthetical constructions, even though some grammatical rules try to capture them. On the other hand, the unlexicalized PCFG parser is reasonably accurate but can face difficulties with uncommon parenthetical structures.

4.5 Noun phrases with multiple modifiers and complements

Some phrasal constructions appear to be very complex when they include numerous modifiers or complements. In this subsection, we focus on long noun phrases with multiple modifiers. We use the following sentences to support our claims. Parsers' outputs and gold standards for the first example can be found in Appendix A, while Figure 5 presents parsers' outputs for the second example.

1. **The slightly simplified version of the Viterbi algorithm that we present** takes as input a single HMM and a sequence of observed words $O = (o_1, o_2, ...o_T)$ and returns the most probable state/tag sequence $Q = (q_1, q_2, q_T)$ together with its probability.

2. He hits **the girl with the umbrella in the park**.

```
He hits the girl with the umbrella in the park.

• RASP output:
(|ncsubj| |hit+s:1_VVZ| |he:0_PPHS1| _)
(|iobj| |hit+s:1_VVZ| |with:4_IW|)
(|dobj| |hit+s:1_VVZ| |girl:3_NN1|)
(|dobj| |with:4_IW| |umbrella:6_NN1|)
(|det| |umbrella:6_NN1| |the:5_AT|)
(|ncmod| _ |umbrella:6_NN1| |in:7_II|)
(|dobj| |in:7_II| |park:9_NNL1|)
(|det| |park:9_NNL1| |the:8_AT|)
(|det| |girl:3_NN1| |the:2_AT|)

• Unlexicalized PCFG parser output:
1       He      _       PRP     PRP     _       2       nsubj   _       _
2       hits    _       VBZ     VBZ     _       0       root    _       _
3       the     _       DT      DT      _       4       det     _       _
4       girl    _       NN      NN      _       2       obj     _       _
5       with    _       IN      IN      _       7       case    _       _
6       the     _       DT      DT      _       7       det     _       _
7       umbrella _      NN      NN      _       2       obl     _       _
8       in      _       IN      IN      _       10      case    _       _
9       the     _       DT      DT      _       10      det     _       _
10      park    _       NN      NN      _       7       nmod    _       _
```

Figure 5: Example of parsers' output for a noun phrase with multiple modifiers (sentence 2).

In sentence 1, both parsers consider that the relative clause *that we present* relates to the previous word *algorithm*. In reality, this relative clause refers rather to the word *version*. Parsers have been misled by the prepositional phrase *of the Viterbi algorithm* that complements the term *version* and that is located between *version* and its relative clause. Parsing algorithms do have a tendency to relate modifiers or complements to the immediately preceding word.

This is exactly the same error that we encounter in the 2nd sentence. Both parsers relate *in the park* to *umbrella*, while it is rather the *girl* who is *in the park*. Once again, parsers create a dependency relation between a modifier and its immediately previous words, which leads to poor analyses when a noun has multiple modifiers and complements.

4.6 Ambiguous Named-Entities

Sentences from the project handout do not contain any ambiguous Named-Entities, but the latter can pose a problem to parsers. Two types of ambiguity can be identified for Named-Entities: Named-Entities including common words such as *Apple*, *American Airlines* or *France Gall* (see Figure 6), and person title such as *Professor*, *Prof.* or *Director* (see Figure 7). We examine the following sentences:

1. **France Gall** is dead.

2. **Apple** is bankrupt.

3. **Vice-Chief Andrews** retired.

4. **Vice Chief Andrews** retired.

```
France Gall is dead.

• RASP output:
(|ncsubj| |be+s:2_VBZ| |gall:1_NN1| _)
(|xcomp| _ |be+s:2_VBZ| |dead:3_JJ|)
(|ncmod| _ |gall:1_NN1| |France:0_NP1|)

• Unlexicalized PCFG parser output:
1       France  _       NNP     NNP     _       2       compound _      _
2       Gall    _       NNP     NNP     _       4       nsubj   _       _
3       is      _       VBZ     VBZ     _       4       cop     _       _
4       dead    _       JJ      JJ      _       0       root    _       _


Apple is bankrupt.

• RASP output:
(|ncsubj| |be+s:1_VBZ| |apple:0_NN1| _)
(|xcomp| _ |be+s:1_VBZ| |bankrupt:2_JJ|)

• Unlexicalized PCFG parser output:
1       Apple   _       NNP     NNP     _       3       nsubj   _       _
2       is      _       VBZ     VBZ     _       3       cop     _       _
3       bankrupt _      JJ      JJ      _       0       root    _       _
```

Figure 6: Example of parsers' output for Named-Entities containing common words (sentences 1 and 2).

First, Figure 6 shows two examples of Named-Entities that contain common words (sentences 1

and 2). We notice that the RASP system fails at recognizing these Named-Entities and considers them as nouns. On the contrary, the unlexicalized PCFG parser from Stanford tools is able to identify these Named-Entities and treats them correctly as proper nouns.

```
Vice-Chief Andrews retired.

• RASP output:
(|ncsubj| |retire+ed:2_VVD| |Andrews:1_NP1| _)
(|ncmod| _ |Andrews:1_NP1| |vice-chief:0_NNSB1|)

• Unlexicalized PCFG parser output:
1       Vice-Chief _      NNP      NNP      _       2       compound _       _
2       Andrews    _      NNP      NNP      _       3       nsubj    _       _
3       retired    _      VBD      VBD      _       0       root     _       _


Vice Chief Andrews retired.

• RASP output:
(|ncsubj| |retire+ed:3_VVD| |Andrews:2_NP1| _)
(|ncmod| _ |Andrews:2_NP1| |chief:1_NNS1|)
(|ncmod| _ |chief:1_NNS1| |vice:0_NN1|)

• Unlexicalized PCFG parser output:
1       Vice     _        NNP      NNP      _       2       compound _       _
2       Chief    _        NNP      NNP      _       3       compound _       _
3       Andrews  _        NNP      NNP      _       4       nsubj    _       _
4       retired  _        VBD      VBD      _       0       root     _       _
```

Figure 7: Example of parsers' output for Named-Entities with person title (sentences 3 and 4).

Figure 7 gathers two examples to study parsers' outputs for Named-Entities containing person title (sentences 3 and 4). Once again, the unlexicalized PCFG parser performs better than the RASP system. Even though the RASP system recognizes the Named-Entity *Vice-Chief Andrews*, it does not for *Vice Chief Andrews*, which leads to a wrong sentence analysis. On the contrary, the unlexicalized PCFG succeeds in both cases.

These two errors on Named-Entities have been observed with many similar Named-Entities. They refer to the data processing step carried on by both parsers. While the unlexicalized PCFG parser is able to identify accurately Named-Entities, the Named-Entity Recognition task performed by the RASP system can be improved.

4.7 Subject-verb inversions

The sentences suggested in the project handout do not include any subject-verb inversions. Such phrase structures are not common in English, even though they are grammatically correct. For this reason, we have a closer look at parsers' analyses of sentences with subject-verb inversions. We consider the following examples and their parsing analyses in Figure 8 to illustrate our results.

1. **The dog came** down the stairs.
2. Down the stairs **came the dog**.

We see that both parsers fail to analyse the sentence with a subject-verb inversion (2nd sentence),

```
The dog came down the stairs.

• RASP output:
(|ncsubj| |come+ed:2_VVD| |dog:1_NN1| _)
(|dobj| |come+ed:2_VVD| |stair+s:5_NN2|)
(|det| |stair+s:5_NN2| |the:4_AT|)
(|ncmod| |prt| |come+ed:2_VVD| |down:3_RP|)
(|det| |dog:1_NN1| |the:0_AT|)

• Unlexicalized PCFG parser output:
1       The      _        DT       DT       _       2       det          _  _
2       dog      _        NN       NN       _       3       nsubj        _  _
3       came     _        VBD      VBD      _       0       root         _  _
4       down     _        RP       RP       _       3       compound:prt _  _
5       the      _        DT       DT       _       6       det          _  _
6       stairs   _        NNS      NNS      _       3       obj          _  _


Down the stairs came the dog.

• RASP output:
(|ncmod| _ |come+ed:3_VVD| |down:0_RP|)
(|ncsubj| |come+ed:3_VVD| |stair+s:2_NN2| _)
(|dobj| |come+ed:3_VVD| |dog:5_NN1|)
(|det| |dog:5_NN1| |the:4_AT|)
(|det| |stair+s:2_NN2| |the:1_AT|)

• Unlexicalized PCFG parser output:
1       Down     _        RB       RB       _       4       advmod   _       _
2       the      _        DT       DT       _       3       det      _       _
3       stairs   _        NNS      NNS      _       4       nsubj    _       _
4       came     _        VBD      VBD      _       0       root     _       _
5       the      _        DT       DT       _       6       det      _       _
6       dog      _        NN       NN       _       4       obj      _       _
```

Figure 8: Example of parsers' output for a subject-verb inversion phrase structure (sentences 1 and 2).

while they perform well when the subject is before the verb (1st sentence). Indeed, when the subject is placed after the verb, both parsers analyse the subject as an object and produce a wrong analysis. This error appears with many other sentences containing subject-verb inversions.

However, this is not a surprising error as the RASP reads input text from left to right without backing up, and the Stanford PCFG parser relies on production rules with a left-to-right order. We can therefore assume that both parsers have integrated the traditional left-to-right order of English sentences, and are not able to deal correctly with a right-to-left order.

4.8 Verbal forms used as non-clausal modifiers

We detail here a last difficulty faced by parsers, that is dealing with verbal forms used as non-clausal modifiers. Past participle and gerund verbs are usually employed as adjectives, and entail a non-clausal modifier grammatical relation with the noun it is related to (`ncmod` for RASP relations or `amod` for the Stanford dependencies). However, parsers struggle to identify these verbal forms as adjectives and make errors when analysing them. This problem can also refers to an incorrect part-of-speech tagging (see Section 4.3). We illustrate this assertion with the two following sentences, and their parsing analyses in Figure 9.

1. I love these **recorded** videos.
2. An application needs **computing** units to run.

```
I love these recorded videos.

• RASP output:
(|ncsubj| |love:1_VV0| |I:0_PPIS1| _)
(|ccomp| _ |love:1_VV0| |record+ed:3_VVN|)
(|ncsubj| |record+ed:3_VVN| |these:2_DD2| _)
(|obj| |record+ed:3_VVN| |video+s:4_NN2|)

• Unlexicalized PCFG parser output:
1       I         _       PRP     PRP     _       2       nsubj   _       _
2       love      _       VBP     VBP     _       0       root    _       _
3       these     _       DT      DT      _       5       det     _       _
4       recorded  _       VBN     VBN     _       5       amod    _       _
5       videos    _       NNS     NNS     _       2       obj     _       _


An application needs computing units to run.

• RASP output:
(|ncsubj| |need+s:2_VVZ| |application:1_NN1| _)
(|xcomp| _ |need+s:2_VVZ| |compute+ing:3_VVG|)
(|dobj| |compute+ing:3_VVG| |unit+s:4_NN2|)
(|xmod| |to| |unit+s:4_NN2| |run:6_VV0|)
(|det| |application:1_NN1| |an:0_AT1|)

• Unlexicalized PCFG parser output:
1       An          _       DT      DT      _       2       det     _       _
2       application _       NN      NN      _       3       nsubj   _       _
3       needs       _       VBZ     VBZ     _       0       root    _       _
4       computing   _       NN      NN      _       5       compound _      _
5       units       _       NNS     NNS     _       3       obj     _       _
6       to          _       TO      TO      _       7       mark    _       _
7       run         _       VB      VB      _       3       advcl   _       _
```

Figure 9: Example of parsers' output for verbal forms used as non-clausal modifiers (sentences 1 and 2).

In sentence 1, we observe that *recorded* is tagged as a past participle verb by both parsers. While the unlexicalized PCFG parser succeeds in analysing this word as an adjectival modifier of *videos*, the RASP system does not. It treats *recorded* as a verb and considers that *videos* is an object of the verb.

Sentence 2 leads to the same observations: both parsers tag *computing* as a gerund verb, and the unlexicalized PCFG parser does consider it as an adjectival modifier of *units* while the RASP system analyses it as a verb with *units* as a direct object.

These results lead to the conclusion that the unlexicalized PCFG parser is able to accurately identify verbal forms used as non-clausal modifiers, while the RASP system usually produces poor analyses of these verbal forms.

4.9 Speed

Beyond accuracy, parsers' performance can also be compared in terms of speed. We examine the running time of both parsers when parsing the 25 sentences suggested in the project handout, containing a total of 725 words. The RASP system needs 3.087s to parse sentences, while the unlexicalized PCFG parser takes 46.726s for the same task. These figures have been obtained on my personal computer that uses a CPU Intel Core i5 of 2.7GHz. The RASP system is therefore about 15 times faster than the Stanford PCFG parser. This difference of running times is significant, and can argue in favor of the RASP system when parsing a very long text.

This result can be explained by the fact that the RASP system uses an LR parser, which reads input text only once from left to right, without backing up. As for the unlexicalized PCFG parser, it relies on a CKY algorithm which requires more computations.

5 Conclusion

In conclusion, we examine two parsing algorithms, the RASP system and the unlexicalized PCFG parser from the Stanford NLP tools. In addition to a quantitative analysis, a qualitative evaluation of parsers' outputs highlights the strengths and weaknesses of both parsers. Notably, it seems that the unlexicalized PCFG parser is more robust to common sources of errors, such as ambiguous Named-Entities or verbal forms used as non-clausal modifiers.

This work tries to identify several different criteria to evaluate parsers in order to obtain a reliable comparison of their performance. However, the previous results and conclusions must be qualified, as they rely on very little data.

Word count: 4,831 words

References

J. Backus. 1959. The syntax and semantics of the proposed international algebraic language of the zurich acm-gamm conference. In *IFIP Congress*.

Leonard E. Baum. 1972. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. In *Inequalities III: Proceedings of the Third Symposium on Inequalities*, pages 1–8, University of California, Los Angeles. Academic Press.

Taylor L. Booth. 1969. Probabilistic representation of formal languages. In *Proceedings of the 10th Annual Symposium on Switching and Automata Theory (Swat 1969)*, SWAT '69, page 74–81, USA. IEEE Computer Society.

Ted Briscoe and John Carroll. 2002. Robust accurate statistical annotation of general text. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Canary Islands - Spain. European Language Resources Association (ELRA).

Ted Briscoe, John Carroll, and Rebecca Watson. 2006. The second release of the RASP system. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, pages 77–80, Sydney, Australia. Association for Computational Linguistics.

Sabine Buchholz and Erwin Marsi. 2006. CoNLL-X shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*,

pages 149–164, New York City. Association for Computational Linguistics.

Eugene Charniak. 1997. Statistical parsing with a context-free grammar and word statistics. In *AAAI/IAAI*.

N. Chomsky. 1956. Three models for the description of language. *IRE Transactions on Information Theory*, 2(3):113–124.

Michael Collins. 1999. Book reviews: Beyond grammar: An experience-based theory of language. *Computational Linguistics*, 25(3).

Jay Earley. 1970. An efficient context-free parsing algorithm. *Commun. ACM*, 13(2):94–102.

David Elworthy. 1994. Does baum-welch re-estimation help taggers? *CoRR*, abs/cmp-lg/9410012.

U. Essen and H. Ney. 1991. On smoothing techniques for bigram-based natural language modelling. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, pages 825–828, Los Alamitos, CA, USA. IEEE Computer Society.

T. Kasami. 1965. An efficient recognition and syntax-analysis algorithm for context-free languages.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 423–430, Sapporo, Japan. Association for Computational Linguistics.

Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: The penn treebank. *Comput. Linguist.*, 19(2):313–330.

Marie-Catherine de Marneffe and Christopher D. Manning. 2008. The Stanford typed dependencies representation. In *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pages 1–8, Manchester, UK. Coling 2008 Organizing Committee.

Jens Nilsson and Joakim Nivre. 2008. Malteval: An evaluation and visualization tool for dependency parsing. In *In Proceedings of the Sixth International Language Resources and Evaluation. LREC*.

Joakim Nivre, Marie-Catherine de Marneffe, F. Ginter, Y. Goldberg, Jan Hajic, Christopher D. Manning, R. McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and D. Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In *LREC*.

G. Sampson. 1995. English for the computer.

Mark Steedman. 1997. Surface structure and interpretation. In *Linguistic inquiry*.

Mark Steedman. 2004. The syntactic process. In *Language, speech, and communication*.

Andreas Stolcke. 1995. An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, 21(2):165–201.

A. Viterbi. 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theor.*, 13(2):260–269.

Daniel H. Younger. 1967. Recognition and parsing of context-free languages in time n^3. *Inf. Control.*, $10 : 189 - -208$.

A Parsers' outputs

A.1 Sentences of the handout

Sentences in bold are the 10 sentences studied in depth in this work.

1. The old car broke down in the car park.

2. At least two men broke in and stole my TV.

3. The horses were broken in and ridden in two weeks.

4. Kim and Sandy both broke up with their partners.

5. The horse which Kim sometimes rides is more bad tempered than mine.

6. The horse as well as the rabbits which we wanted to eat have escaped.

7. It was my aunt's car which we sold at auction last year in February.

8. The only rabbit that I ever liked was eaten by my parents one summer.

9. The veterans who I thought that we would meet at the reunion were dead.

10. Natural disasters – storms, flooding, hurricanes – occur infrequently but cause devastation that strains resources to breaking point.

11. Letters delivered on time by old-fashioned means are increasingly rare, so it is as well that that is not the only option available.

12. It won't rain but there might be snow on high ground if the temperature stays about the same over the next 24 hours.

13. The long and lonely road to redemption begins with self-reflection: the need to delve inwards to deconstruct layers of psychological obfuscation.

14. My wildest dream is to build a POS tagger which processes 10K words per second and uses only 1MB of RAM, but it may prove too hard.

15. English also has many words of more or less unique function, including interjections (oh, ah), negatives (no, not), politeness markers (please, thank you), and the existential 'there' (there are horses but not unicorns) among others.

16. Making these decisions requires sophisticated knowledge of syntax; tagging manuals (Santorini, 1990) give various heuristics that can help human coders make these decisions and that can also provide useful features for automatic taggers.

17. The Penn Treebank tagset was culled from the original 87-tag tagset for the Brown Corpus. For example the original Brown and C5 tagsets include a separate tag for each of the different forms of the verbs *do* (e.g. C5 tag VDD for *did* and VDG tag for *doing*), *be* and *have*.

18. The slightly simplified version of the Viterbi algorithm that we present takes as input a single HMM and a sequence of observed words $O = (o_1, o_2, ...o_T)$ and returns the most probable state/tag sequence $Q = (q_1, q_2, q_T)$ together with its probability.

19. Thus the EM-trained "pure HMM" tagger is probably best suited to cases where no training data is available, for example, when tagging languages for which no data was previously hand-tagged.

20. Coming home from very lonely places, all of us go a little mad: whether from great personal success, or just an all-night drive, we are the sole survivors of a world no one else has ever seen.

21. Skill without imagination is craftsmanship and gives us many useful objects such as wicker-work picnic baskets. Imagination without skill gives us modern art.

22. But far fewer people fully understand how the Media Lab operates, fits into MIT, and encourages such a creative environment; about half of the anniversary celebration's program focused on simply defining what the Media Lab is.

23. An MoD spokesman said: "Surveys of Astute have now been completed and she will proceed to Faslane under her own power. She is being escorted by tugs and HMS Shoreham."

24. Instead of constantly worrying about funding, the faculty and students can focus on their project, with the exception of sponsors' weeks, when they have to convince companies to start or continue their support.

25. The doctors are warning that the NHS cannot make the £20bn of savings by 2014 that ministers expect, while simultaneously undertaking a huge reorganisation that will see England's 152 primary care trusts (PCTs) abolished and consortiums of GPs assume responsibility for the commissioning of services for patients.

A.2 RASP outputs and gold standards

RASP output for the 1st sentence of the handout:

```
(|ncsubj| |break+ed:3_VVD| |car:2_NN1| _)
(|iobj| |break+ed:3_VVD| |in:5_II|)
(|dobj| |in:5_II| |park:8_NNL1|)
(|det| |park:8_NNL1| |the:6_AT|)
(|ncmod| _ |park:8_NNL1| |car:7_NN1|)
(|ncmod| |prt| |break+ed:3_VVD| |down:4_RP|)
(|det| |car:2_NN1| |the:0_AT|)
(|ncmod| _ |car:2_NN1| |old:1_JJ|)
```

Gold standards in the RASP format for the 1st sentence of the handout:

```
(|ncsubj| |broke:3| |car:2| _)
(|iobj| |broke:3| |in:5|)
(|dobj| |in:5| |park:8|)
(|det| |park:8| |the:6|)
(|ncmod| _ |park:8| |car:7|)
(|ncmod| prt |broke:3| |down:4|)
(|det| |car:2| |the:0|)
(|ncmod| |car:2| |old:1|)
```

RASP output for the 2nd sentence of the handout:

```
(|ncmod| _ |and:6_CC| |at:0_RR|)
(|ncmod| _ |and:6_CC| |least:1_RR|)
(|ncsubj| |and:6_CC| |man+s:3_NN2| _)
```

(|conj| |and:6_CC| |break+ed:4_VVD|)
(|conj| |and:6_CC| |steal+ed:7_VVD|)
(|dobj| |steal+ed:7_VVD| |tv:9_NN1|)
(|det| |tv:9_NN1| |my:8_APP$|)
(|ncmod| |prt| |break+ed:4_VVD| |in:5_RP|)
(|ncmod| |num| |man+s:3_NN2| |two:2_MC|)

Gold standards in the RASP format for the 2nd sentence of the handout:
(|dobj| |at:0| |least:1|)
(|ncmod| _ |two:2| |at:0|)
(|ncsubj| |and:6| |men:3| _)
(|conj| |and:6| |broke:5|)
(|conj| |and:6| |stole:7|)
(|dobj| |stole:7| |TV:9|)
(|det| |TV:9| |my:8|)
(|ncmod| prt |broke:4| |in:5|)
(|ncmod| num |men:3| |two:2|)

RASP output for the 7th sentence of the handout:
(|ncsubj| |be+ed:1_VBDZ| |it:0_PPH1| _)
(|xcomp| _ |be+ed:1_VBDZ| |car:5_NN1|)
(|ncmod| |poss| |car:5_NN1| |aunt:3_NN1|)
(|cmod| _ |car:5_NN1| |sell+ed:8_VVD|)
(|obj| |sell+ed:8_VVD| |which:6_DDQ|)
(|ncsubj| |sell+ed:8_VVD| |we:7_PPIS2| _)
(|ncmod| _ |sell+ed:8_VVD| |year:12_NNT1|)
(|ncmod| _ |year:12_NNT1| |in:13_II|)
(|dobj| |in:13_II| |February:14_NPM1|)
(|ncmod| |num| |year:12_NNT1| |last:11_MD|)
(|iobj| |sell+ed:8_VVD| |at:9_II|)
(|dobj| |at:9_II| |auction:10_NN1|)
(|det| |aunt:3_NN1| |my:2_APP$|)

Gold standards in the RASP format for the 7th sentence of the handout:
(|ncsubj| |was:1| |it:0| _)
(|xcomp| |was:1| |car:5|)
(|ncmod| poss |car:5| |aunt:3|)
(|cmod| _ |car:5| |sold:8|)
(|dobj| |sold:8| |which:6|)
(|ncsubj| |sold:8| |we:7| _)
(|ncmod| _ |sold:8| |year:12|)
(|ncmod| _ |sold:8| |in:13|)
(|dobj| |in:12| |February:14|)
(|ncmod| _ |year:12| |last:11|)
(|ncmod| _ |sold:8| |at:9|)

```
(|dobj| |at:9| |auction:10|)
(|det| |aunt:3| |my:2|)
```

RASP output for the 10th sentence of the handout:

```
(|conj| |but:11_CCB| |devastation:13_NN1|)
(|ncmod| _ |devastation:13_NN1| |cause:12_NN1|)
(|ncsubj| |strain+s:15_VVZ| |devastation:13_NN1| _)
(|cmod| |that| |devastation:13_NN1| |strain+s:15_VVZ|)
(|dobj| |strain+s:15_VVZ| |resource+s:16_NN2|)
(|xmod| _ |resource+s:16_NN2| |to:17_II|)
(|xcomp| _ |to:17_II| |break+ing:18_VVG|)
(|dobj| |break+ing:18_VVG| |point:19_NN1|)
(|ccomp| _ |natural:0_JJ| |occur:9_VV0|)
(|ncsubj| |occur:9_VV0| |disaster+s:1_NN2| _)
(|ncmod| _ |occur:9_VV0| |infrequently:10_RR|)
(|ta| |bal| |disaster+s:1_NN2| |hurricane+s:7_NN2|)
(|ncmod| _ |hurricane+s:7_NN2| |storm+s:3_NN2|)
(|ta| |bal| |storm+s:3_NN2| |flood+ing:5_VVG|)
```

Gold standards in the RASP format for the 10th sentence of the handout:

```
(|conj| |but:11| |occur:9|)
(|conj| |but:11| |cause:12|)
(|dobj| |devastation:13| |cause:12|)
(|nbsubj| |strains:15| |devastation:13| _)
(|cmod| that |devastation:13| |strains:15|)
(|dobj| |strains:15| |resources:16|)
(|pmod| |strains:15| |to:17|)
(|dobj| |to:17| |point:19|)
(|ncmod| _ |point:19| |breaking:18|)
(|ncmod| _ |disasters:1| |natural:0|)
(|ncsubj| |but:11| |disasters:1| _)
(|ncmod| _ |occur:9| |infrequently:10|)
(|ta| bal |disasters:1| |storms:3|)
(|ta| bal |disasters:1| |flooding:5|)
(|ta| bal |disasters:1| |hurricanes:7|)
```

RASP output for the 12th sentence of the handout:

```
(|conj| |but:4_CCB| |rain:3_VV0|)
(|conj| |but:4_CCB| |be:7_VB0|)
(|ncsubj| |be:7_VB0| |there:5_EX| _)
(|cmod| _ |be:7_VB0| |if:12_CS|)
(|ccomp| _ |if:12_CS| |stay+s:15_VVZ|)
(|ncsubj| |stay+s:15_VVZ| |temperature:14_NN1| _)
(|iobj| |stay+s:15_VVZ| |about:16_II|)
(|dobj| |about:16_II| |same:18_DA|)
```

(|det| |same:18_DA| |the:17_AT|)
(|ncmod| _ |same:18_DA| |over:19_II|)
(|dobj| |over:19_II| |hour+s:23_NNT2|)
(|ncmod| _ |hour+s:23_NNT2| |next:21_MD|)
(|ncmod| |num| |hour+s:23_NNT2| |24:22_MC|)
(|det| |next:21_MD| |the:20_AT|)
(|det| |temperature:14_NN1| |the:13_AT|)
(|ncmod| _ |be:7_VB0| |on:9_II|)
(|dobj| |on:9_II| |ground:11_NN1|)
(|ncmod| _ |ground:11_NN1| |high:10_JJ|)
(|aux| |be:7_VB0| |might:6_VM|)
(|xcomp| _ |be:7_VB0| |snow:8_NN1|)
(|ncsubj| |rain:3_VV0| |it:0_PPH1| _)
(|aux| |rain:3_VV0| |will+:1_VM|)
(|ncmod| _ |rain:3_VV0| |not+:2_XX|)

Gold standards in the RASP format for the 12th sentence of the handout:

(|conj| |but:4| |rain:3|)
(|conj| |but:4| |be:7|)
(|ncsubj| |be:7| |there:5| _)
(|cmod| _ |be:7| |if:12|)
(|ccomp| |if:12| |stays:15|)
(|ncsubj| |stays:15| |temperature:14| _)
(|iobj| _ |stays:15| |about:16|)
(|dojb| _ |about:16| |same:18|)
(|det| |same:18| |the:17|)
(|ncmod| _ |stays:15| |over:19|)
(|dobj| |over:19| |hours:23|)
(|ncmod| _ |hours:23| |next:21|)
(|ncmod| num |hours:23| |24:22|)
(|det| |hours:23| |the:20|)
(|det| |temperature:14| |the:13|)
(|ncmod| _ |be:7| |on:9|)
(|dobj| |on:9| |ground:11|)
(|ncmod| _ |ground:11| |high:10|)
(|aux| |be:7| |might:6|)
(|xcomp| |be:7| |snow:8|)
(|ncsubj| |rain:3| |it:0| _)
(|aux| |rain:3| |will:1|)
(|ncmod| _ |rain:3| |not:2|)

RASP output for the 15th sentence of the handout:

(|dobj| |among:50_II| |other+s:51_NN2|)
(|ta| |bal| |there:40_EX| |be+:44_VBR|)
(|ncsubj| |be+:44_VBR| |there:43_EX| _)
(|xcomp| _ |be+:44_VBR| |but:46_CCB|)

```
(|conj| |but:46_CCB| |horse+s:45_NN2|)
(|ncmod| _ |unicorn+s:48_NN2| |not+:47_XX|)
(|conj| |but:46_CCB| |unicorn+s:48_NN2|)
(|ta| |quote| |ellip| |there:40_EX|)
(|conj| |and:36_CC| |existential:38_JJ|)
(|det| |existential:38_JJ| |the:37_AT|)
(|ta| |bal| |marker+s:28_NN2| |thank:32_VV0|)
(|dobj| |thank:32_VV0| |you:33_PPY|)
(|ncmod| _ |marker+s:28_NN2| |politeness:27_NN1|)
(|ncmod| _ |have+s:2_VHZ| |also:1_RR|)
(|dobj| |have+s:2_VHZ| |or:7_CC|)
(|conj| |or:7_CC| |word+s:4_NN2|)
(|conj| |or:7_CC| |negative+s:20_NN2|)
(|ncmod| _ |negative+s:20_NN2| |less:8_DAR|)
(|ncmod| _ |negative+s:20_NN2| |unique:9_JJ|)
(|ncmod| _ |negative+s:20_NN2| |function:10_NN1|)
(|ta| |bal| |function:10_NN1| |include+ing:12_VVG|)
(|dobj| |include+ing:12_VVG| |interjection+s:13_NN2|)
(|ta| |bal| |interjection+s:13_NN2| |ah:17_UH|)
(|ncmod| _ |word+s:4_NN2| |many:3_DA2|)
(|iobj| |word+s:4_NN2| |of:5_IO|)
(|dobj| |of:5_IO| |more:6_DAR|)
```

Gold standards in the RASP format for the 15th sentence of the handout:

```
(|ncmod| _ |and:36| |among:50|)
(|dobj| |among:50| |others:51|)
(|ta| brack |there:40| |are:44|)
(|ncsubj| |are:44| |there:43| _)
(|dobj| |are:44| |but:46|)
(|conj| |but:46| |horses:45|)
(|ncmod| _ |unicorns:48| |not:47|)
(|conj| |but:46| |unicorns:48|)
(|ta| quote |ellip| | there:40|)
(|conj| |and:36| |there:40|)
(|det| |there:40| |the:24|)
(|ncmod| |there:40| |existential:38|)
(|ta| brack |markers:28| |thank:32|)
(|dobj| |thank:32| |you:33|)
(|ta| brack |markers:28| |please:20|)
(|ncmod| _ |markers:28| |politeness:27|)
(|ncmod | _ |has:2| |also:1|)
(|dobj| |has:2| |words:4|)
(|conj| |or:7| |more:6|)
(|conj| |or:7| |less:8|)
(|ncmod | _ |unique:9| |or:7|)
(|iobj| |words:4| |of:5|)
```

```
(|dobj| |of:5| |function:10|)
(|ncmod | _ |function:10| |unique:9|)
(|xcomp| _ |function:10| |including:12|)
(|ncsubj| |has:2| |English:0| _)
(|ncmod| _ |words:4| |many:3|)
(|dobj| |including:12| |and:36|)
(|conj| |and:36| |interjections:13|)
(|conj| |and:36| |negatives:20|)
(|conj| |and:36| |markers:28|)
(|ta| brack |interjections:13| |oh:15|)
(|ta| brack |interjections:13| |ah:17|)
(|ta| brack |negatives:20| |no:15|)
(|ta| brack |negatives:20| |not:17|)
```

RASP output for the 16th sentence of the handout:

```
(|ncmod| _ |provide:31_VV0| |also:30_RR|)
(|iobj| |provide:31_VV0| |for:34_IF|)
(|dobj| |provide:31_VV0| |feature+s:33_NN2|)
(|dobj| |for:34_IF| |tagger+s:36_NN2|)
(|ncmod| _ |tagger+s:36_NN2| |automatic:35_JJ|)
(|ncmod| _ |feature+s:33_NN2| |useful:32_JJ|)
(|conj| ; |give:16_VV0|)
(|conj| ; |require+s:3_VVZ|)
(|ncsubj| |give:16_VV0| |manual+s:10_NN2| _)
(|ccomp| _ |give:16_VV0| |make:24_VV0|)
(|ncsubj| |make:24_VV0| |heuristic+s:18_NN1| _)
(|dobj| |make:24_VV0| |decision+s:26_NN2|)
(|det| |decision+s:26_NN2| |these:25_DD2|)
(|ncmod| _ |heuristic+s:18_NN1| |various:17_JJ|)
(|ncsubj| |help:21_VV0| |heuristic+s:18_NN1| _)
(|cmod| |that| |heuristic+s:18_NN1| |help:21_VV0|)
(|aux| |help:21_VV0| |can:20_VM|)
(|dobj| |help:21_VV0| |coder+s:23_NN2|)
(|ncmod| _ |coder+s:23_NN2| |human:22_JJ|)
(|ta| |bal| |manual+s:10_NN2| |Santorini:12_NP1|)
(|ncmod| _ |Santorini:12_NP1| 1990:14_MC)
(|ncsubj| |tag+ing:9_VVG| |manual+s:10_NN2| _)
(|ncmod| _ |manual+s:10_NN2| |tag+ing:9_VVG|)
(|xsubj| |require+s:3_VVZ| |make+ing:0_VVG| _)
(|dobj| |require+s:3_VVZ| |knowledge:5_NN1|)
(|ncmod| _ |knowledge:5_NN1| |sophisticated:4_JJ|)
(|iobj| |knowledge:5_NN1| |of:6_IO|)
(|dobj| |of:6_IO| |syntax:7_NN1|)
(|dobj| |make+ing:0_VVG| |decision+s:2_NN2|)
(|det| |decision+s:2_NN2| |these:1_DD2|)
```

Gold standards in the RASP format for the 16th sentence of the handout:

```
(|ncmod| _ |provide:31| |also:30|)
(|iobj| |provide:31| |for:34|)
(|dobj| |provide:31| |features:33|)
(|dobj| |for:34| |taggers:36|)
(|ncmod| _ |taggers:36| |automatic:35|)
(|ncmod| _ |features:33| |useful:32|)
(|conj| ; |give:16|)
(|conj| ; |requires:3|)
(|conj| |and:27| |that:19|)
(|conj| |and:27| |that:28|)
(|ncsubj| |provide:31| |that:28| _)
(|aux| | provide:31| |can:29|)
(|ncsubj| |give:16| |manuals:10| _)
(|dobj| |give:16| |heuristics:18|)
(|ncsubj| |make:24| |coders:23| _)
(|dobj| |make:24| |decisions:26|)
(|det| |decisions:26| |these:25_DD2|)
(|ncmod| _ |heuristics:18| |various:17|)
(|ncsubj| |help:21| |that:19| _)
(|cmod| |that| |heuristics:18| |and:27|)
(|aux| |help:21| |can:20|)
(|dobj| |help:21| |coders:23|)
(|ncmod| _ |coders:23| |human:22|)
(|ta| |brack| |manuals:10| |Santorini:12|)
(|ncmod| _ |Santorini:12| |1990:14|)
(|ncmod| _ |manuals:10| |tagging:9|)
(|xsubj| |requires:3| |making:0| _)
(|dobj| |requires:3| |knowledge:5|)
(|ncmod| _ |knowledge:5| |sophisticated:4|)
(|ncmod| _ |knowledge:5| |of:6|)
(|dobj| |of:6| |syntax:7|)
(|dobj| |making:0| |decisions:2|)
(|det| |decisions:2| |these:1|)
```

RASP output for the 18th sentence of the handout:

```
(|ta| |bal| O1:26_NP1 O2:28_NP1)
(|conj| |and:17_CC| |sequence:19_NN1|)
(|ncmod| _ |sequence:19_NN1| |=:24_FO|)
(|ncmod| _ |=:24_FO| |o:23_ZZ1|)
(|det| |sequence:19_NN1| |a:18_AT1|)
(|iobj| |sequence:19_NN1| |of:20_IO|)
(|dobj| |of:20_IO| |word+s:22_NN2|)
(|passive| |observe+ed:21_VVN|)
(|ncsubj| |observe+ed:21_VVN| |word+s:22_NN2| |obj|)
(|ncmod| _ |word+s:22_NN2| |observe+ed:21_VVN|)
```

```
(|ccomp| _ |single:15_VV0| |hmm:16_UH|)
(|det| |version:3_NN1| |the:0_AT|)
(|ncmod| _ |version:3_NN1| |simplify+ed:2_VVN|)
(|iobj| |version:3_NN1| |of:4_IO|)
(|dobj| |of:4_IO| |algorithm:7_NN1|)
(|det| |algorithm:7_NN1| |the:5_AT|)
(|ncmod| _ |algorithm:7_NN1| |Viterbi:6_NP1|)
(|ccomp| |that:8_CST| |algorithm:7_NN1| |present:10_VV0|)
(|ncsubj| |present:10_VV0| |we:9_PPIS2| _)
(|iobj| |present:10_VV0| |as:12_CSA|)
(|dobj| |present:10_VV0| |take+s:11_NN2|)
(|dobj| |as:12_CSA| |input:13_NN1|)
(|ncmod| _ |simplify+ed:2_VVN| |slightly:1_RR|)

(|dobj| |with:19_IW| |probability:21_NN1|)
(|det| |probability:21_NN1| |its:20_APP$|)
(|ta| |bal| |q1:12_NNU| Q2:14_NP1)
(|conj| |and:2_CC| |return+s:3_VVZ|)
(|obj2| |return+s:3_VVZ| |=:10_FO|)
(|dobj| |return+s:3_VVZ| |sequence:8_NN1|)
(|ncmod| _ |=:10_FO| |q:9_ZZ1|)
(|det| |sequence:8_NN1| |the:4_AT|)
(|ncmod| _ |sequence:8_NN1| |most:5_DAT|)
(|ncmod| _ |sequence:8_NN1| |probable:6_JJ|)
(|ncmod| _ |sequence:8_NN1| |state/tag:7_NN1|)
```

Gold standards in the RASP format for the 18th sentence of the handout:

```
(|ta| bal |o1:26| |o2:26|)
(|ta| bal |o2:28| |oT:30|)
(|conj| |and:17| |HMM:16|)
(|conj| |and:17| |sequence:19|)
(|ncmod| _ |sequence:19| |O:23|)
(|ncmod| _ |O:23| |=:24|)
(|ncmod| _ |=:24| |(:25|)
(|det| |sequence:19| |a:18|)
(|iobj| |sequence:19| |of:20|)
(|dobj| |of:20| |words:22|)
(|passive| | observed:21|)
(|ncsubj| |observed:21| |words:22| obj)
(|ncmod| _ |words:22| |observed:21|)
(|ncmod| _ |HMM:16| |single:15|)
(|det| |version:3| |the:0|)
(|ncmod| _ |version:3| |simplified:2|)
(|ncmod| _ |simplified:2| |slightly:1|)
(|ncmod| _ |version:3| |of:4|)
(|dobj| |of:4| |algorithm:7|)
```

```
(|det| |algorithm:7| |the:5|)
(|ncmod| _ |algorithm:7| |Viterbi:6|)
(|cmod| that |algorithm:7| |present:10|)
(|dobj| |present:10| |that:8|)
(|ncsubj| |present:10| |we:9| _)
(|iobj| |takes:11| |as:12|)
(|dobj| |as:12| |input:13|)
(|dobj| |takes:11| |and:32|)
(|det| |HMM:16| |a:14|)
(|ncsubj| |and:24| |version:3| _)
(|conj| |and:32| |takes:11|)
(|conj| |and:32| |returns:33|)
(|dobj| |with:51| |probability:53|)
(|det| |probability:53| |its:52|)
(|ta| bal |q1:43| |q2:45|)
(|ta| bal |q2:45| |qT:47|)
(|ncmod| _ |sequence:40| |Q:41|)
(|ncmod| _ |Q:41| |=:42|)
(|ncmod| _ |=:42| |(:43|)
(|dobj| |returns:33| |sequence:40|)
(|det| |sequence:40| |the:34|)
(|ncmod| _ |sequence:40| |/:38|)
(|conj| |/ :38| |state :37|)
(|conj| |/ :38| |tag :39|)
(|ncmod| _ |probable:36| |most:35|)
(|ncmod| _ |sequence:40| |probable:36|)
(|ncmod| _ |sequence:40| |with:51|)
(|pcomp| |together:50| |with:51|)
```

RASP output for the 21th sentence of the handout:

```
(|ncmod| _ |such:11_DA| |as:12_CSA|)
(|dobj| |as:12_CSA| |basket+s:15_NN2|)
(|ncmod| _ |basket+s:15_NN2| |wickerwork:13_NN1|)
(|ncmod| _ |basket+s:15_NN2| |picnic:14_NN1|)
(|ccomp| _ |skill:0_NN1| |and:5_CC|)
(|ncsubj| |and:5_CC| |without:1_IW| |inv|)
(|conj| |and:5_CC| |be+s:3_VBZ|)
(|conj| |and:5_CC| |give+s:6_VVZ|)
(|obj2| |give+s:6_VVZ| |object+s:10_NN2|)
(|dobj| |give+s:6_VVZ| |we+:7_PPIO2|)
(|ncmod| _ |object+s:10_NN2| |many:8_DA2|)
(|ncmod| _ |object+s:10_NN2| |useful:9_JJ|)
(|xcomp| _ |be+s:3_VBZ| |craftsmanship:4_NN1|)
(|dobj| |without:1_IW| |imagination:2_NN1|)

(|ncmod| _ |art:6_NN1| |modern:5_JJ|)
```

(|ncsubj| |give+s:3_VVZ| |imagination:0_NN1| _)
(|dobj| |give+s:3_VVZ| |we+:4_PPIO2|)
(|ncmod| _ |imagination:0_NN1| |without:1_IW|)
(|dobj| |without:1_IW| |skill:2_NN1|)

Gold standards in the RASP format for the 21th sentence of the handout:
(|ncmod| _ |objects:10| |such:11|)
(|ncmod| _ |such:11| |as:12|)
(|dobj| |as:12| |baskets:15|)
(|ncmod| _ |baskets:15| |wickerwork:13|)
(|ncmod| _ |baskets:15| |picnic:14|)
(|ncsubj| |and:5| | skill:2|)
(|conj| |and:5| |is:3|)
(|conj| |and:5| |gives:6|)
(|obj2| |gives:6| |objects:10|)
(|dobj| |gives:6| |us:7|)
(|ncmod| _ |objects:10| |many:8|)
(|ncmod| _ |objects:10| |useful:9|)
(|dojb| |is:3| |craftsmanship:4|)
(|dobj| |without:1| |imagination:2|)
(|ncmod| _ |skill:0| |without:1|)

(|ncmod| _ |art:6| |modern:5|)
(|iobj2| |gives:3| |art:6|)
(|ncsubj| |gives:3| |imagination:0| _)
(|dobj| |gives:3| |us:4|)
(|dobj| |without:1| |skill:2|)
(|ncmod| _ |imagination:0| |without:1|)

RASP output for the 24th sentence of the handout:
(|ncmod| _ |focus:12_VV0| |instead:0_RR|)
(|ncmod| _ |focus:12_VV0| |with:17_IW|)
(|dobj| |with:17_IW| |exception:19_NN1|)
(|det| |exception:19_NN1| |the:18_AT|)
(|iobj| |exception:19_NN1| |of:20_IO|)
(|dobj| |of:20_IO| |sponsor+s:21_NN2|)
(|cmod| _ |sponsor+s:21_NN2| |convince:29_VV0|)
(|ncmod| _ |convince:29_VV0| |week+s:23_NNT2|)
(|arg_mod| _ |convince:29_VV0| |when:25_RRQ|)
(|ncsubj| |convince:29_VV0| |they:26_PPHS2| _)
(|ncmod| |prt| |have:27_VH0| |to|)
(|aux| |convince:29_VV0| |have:27_VH0|)
(|ncsubj| |or:33_CC| |company+s:30_NNJ2| _)
(|xcomp| |to| |convince:29_VV0| |or:33_CC|)
(|dobj| |convince:29_VV0| |company+s:30_NNJ2|)
(|conj| |or:33_CC| |start:32_VV0|)

(|conj| |or:33_CC| |continue:34_VV0|)
(|dobj| |continue:34_VV0| |support:36_NN1|)
(|det| |support:36_NN1| |their:35_APP$|)
(|ncsubj| |focus:12_VV0| |and:9_CC| _)
(|aux| |focus:12_VV0| |can:11_VM|)
(|iobj| |focus:12_VV0| |on:13_II|)
(|dobj| |on:13_II| |project:15_NN1|)
(|det| |project:15_NN1| |their:14_APP$|)
(|det| |and:9_CC| |the:7_AT|)
(|conj| |and:9_CC| |faculty:8_NN1|)
(|conj| |and:9_CC| |student+s:10_NN2|)
(|xcomp| _ |instead:0_RR| |of:1_IO|)
(|xcomp| _ |of:1_IO| |worry+ing:3_VVG|)
(|ncmod| _ |worry+ing:3_VVG| |constantly:2_RR|)
(|iobj| |worry+ing:3_VVG| |about:4_II|)
(|dobj| |about:4_II| |funding:5_NN1|)

Gold standards in the RASP format for the 24th sentence of the handout:

(|ncmod| _ |focus:12| |instead:0|)
(|ncmod| _ |focus:12| |with:17|)
(|dobj| |with:17| |exception:19|)
(|det| |exception:19| |the:18|)
(|ncmod| _ |exception:19| |of:20|)
(|dobj| |of:20| |weeks:23|)
(|ncmod| |poss| | weeks:23| | sponsors:21|)
(|cmod| _ |weeks:23| |when:25|)
(|ncsubj| |convince:29| |they:26| _)
(|ncmod| |prt| |have:27| |to|)
(|aux| |convince:29| |have:27|)
(|dobj| |convince:29| |companies:30|)
(|conj| |or:33| |start:32|)
(|conj| |or:33| |continue:34|)
(|xcomp| |to| |convince:29| |or:33|)
(|ncsubj| |or:33| |companies:30| _)
(|dobj| |or:33| |support:36|)
(|det| |support:36| |their:35|)
(|ncsubj| |focus:12| |and:9| _)
(|aux| |focus:12| |can:11|)
(|iobj| |focus:12| |on:13|)
(|dobj| |on:13| |project:15|)
(|det| |project:15| |their:14|)
(|det| | faculty:8| |the:7|)
(|conj| |and:9| |faculty:8|)
(|conj| |and:9| |students:10|)
(|pmod| |instead:0| |of:1|)
(|xcomp| _ |of:1| |worry+ing:3|)

```
(|ncmod| _ |worrying:3| |constantly:2|)
(|iobj| |worrying:3| |about:4|)
(|dobj| |about:4| |funding:5|)
```

A.3 Stanford unlexicalized PCFG parser's outputs and CoNLL gold standards

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	3	det	_	_
2	old	_	JJ	JJ	_	3	amod	_	_
3	car	_	NN	NN	_	4	nsubj	_	_
4	broke	_	VBD	VBD	_	0	root	_	_
5	down	_	RP	RP	_	4	compound:prt	_	_
6	in	_	IN	IN	_	9	prep	_	_
7	the	_	DT	DT	_	9	det	_	_
8	car	_	NN	NN	_	9	compound	_	_
9	park	_	NN	NN	_	4	pobj	_	_

Table 1: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 1st sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	3	det	_	_
2	old	_	JJ	JJ	_	3	amod	_	_
3	car	_	NN	NN	_	4	nsubj	_	_
4	broke	_	VBD	VBD	_	0	root	_	_
5	down	_	RP	RP	_	4	compound:prt	_	_
6	in	_	IN	IN	_	4	prep	_	_
7	the	_	DT	DT	_	9	det	_	_
8	car	_	NN	NN	_	9	compound	_	_
9	park	_	NN	NN	_	6	pobj	_	_

Table 2: Gold standards in the CoNLL format for the 1st sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	At	_	IN	IN	_	2	prep	_	_
2	least	_	JJS	JJS	_	3	obl:npmod	_	_
3	two	_	CD	CD	_	4	nummod	_	_
4	men	_	NNS	NNS	_	5	nsubj	_	_
5	broke	_	VBD	VBD	_	0	root	_	_
6	in	_	RP	RP	_	5	compound:prt	_	_
7	and	_	CC	CC	_	8	cc	_	_
8	stole	_	VBD	VBD	_	5	conj	_	_
9	my	_	PRP$	PRP$	_	10	nmod:poss	_	_
10	TV	_	NN	NN	_	8	obj	_	_

Table 3: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 2nd sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	At	_	IN	IN	_	3	prep	_	_
2	least	_	RBS	RBS	_	1	pobj	_	_
3	two	_	CD	CD	_	4	nummod	_	_
4	men	_	NNS	NNS	_	5	nsubj	_	_
5	broke	_	VBD	VBD	_	0	root	_	_
6	in	_	RP	RP	_	5	compound:prt	_	_
7	and	_	CC	CC	_	5	cc	_	_
8	stole	_	VBD	VBD	_	5	conj	_	_
9	my	_	PRP$	PRP$	_	10	nmod:poss	_	_
10	TV	_	NN	NN	_	8	obj	_	_

Table 4: Gold standards in the CoNLL format for the 2nd sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	2	det	_	_
2	horses	_	NNS	NNS	_	4	nsubj:pass	_	_
3	were	_	VBD	VBD	_	4	aux:pass	_	_
4	broken	_	VBN	VBN	_	0	root	_	_
5	in	_	IN	IN	_	4	pobj	_	_
6	and	_	CC	CC	_	7	cc	_	_
7	ridden	_	VBN	VBN	_	4	conj	_	_
8	in	_	IN	IN	_	10	prep	_	_
9	two	_	CD	CD	_	10	nummod	_	_
10	weeks	_	NNS	NNS	_	7	pobj	_	_

Table 5: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 3rd sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	2	det	_	_
2	horses	_	NNS	NNS	_	4	nsubj:pass	_	_
3	were	_	VBD	VBD	_	4	aux:pass	_	_
4	broken	_	VBN	VBN	_	0	root	_	_
5	in	_	IN	IN	_	4	compound:prt	_	_
6	and	_	CC	CC	_	4	cc	_	_
7	ridden	_	VBN	VBN	_	4	conj	_	_
8	in	_	IN	IN	_	4	prep	_	_
9	two	_	CD	CD	_	10	nummod	_	_
10	weeks	_	NNS	NNS	_	8	pobj	_	_

Table 6: Gold standards in the CoNLL format for the 3rd sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Kim	_	NNP	NNP	_	5	nsubj	_	_
2	and	_	CC	CC	_	3	cc	_	_
3	Sandy	_	NNP	NNP	_	1	conj	_	_
4	both	_	DT	DT	_	5	dep	_	_
5	broke	_	VBD	VBD	_	0	root	_	_
6	up	_	RP	RP	_	5	compound:prt	_	_
7	with	_	IN	IN	_	9	prep	_	_
8	their	_	PRP$	PRP$	_	9	nmod:poss	_	_
9	partners	_	NNS	NNS	_	5	pobj	_	_

Table 7: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 4th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Kim	_	NNP	NNP	_	5	nsubj	_	_
2	and	_	CC	CC	_	1	cc	_	_
3	Sandy	_	NNP	NNP	_	1	conj	_	_
4	both	_	PDT	PDT	_	1	appos	_	_
5	broke	_	VBD	VBD	_	0	root	_	_
6	up	_	RP	RP	_	5	compound:prt	_	_
7	with	_	IN	IN	_	5	prep	_	_
8	their	_	PRP$	PRP$	_	9	nmod:poss	_	_
9	partners	_	NNS	NNS	_	7	pobj	_	_

Table 8: Gold standards in the CoNLL format for the 4th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	2	det	_	_
2	horse	_	NN	NN	_	10	nsubj:pass	_	_
3	which	_	WDT	WDT	_	6	obj	_	_
4	Kim	_	NNP	NNP	_	6	nsubj	_	_
5	sometimes	_	RB	RB	_	6	advmod	_	_
6	rides	_	VBZ	VBZ	_	2	rcmod	_	_
7	is	_	VBZ	VBZ	_	10	aux:pass	_	_
8	more	_	RBR	RBR	_	9	advmod	_	_
9	bad	_	RB	RB	_	10	advmod	_	_
10	tempered	_	VBN	VBN	_	0	root	_	_
11	than	_	IN	IN	_	12	prep	_	_
12	mine	_	NN	NN	_	10	pobj	_	_

Table 9: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 5th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	2	det	_	_
2	horse	_	NN	NN	_	10	nsubj	_	_
3	which	_	WP	WP	_	2	ref	_	_
4	Kim	_	NNP	NNP	_	6	nsubj	_	_
5	sometimes	_	RB	RB	_	6	advmod	_	_
6	rides	_	VBZ	VBZ	_	2	rcmod	_	_
7	is	_	VBZ	VBZ	_	10	cop	_	_
8	more	_	RBR	RBR	_	10	advmod	_	_
9	bad	_	RB	RB	_	10	advmod	_	_
10	tempered	_	JJ	JJ	_	0	root	_	_
11	than	_	IN	IN	_	10	prep	_	_
12	mine	_	NN	NN	_	11	pobj	_	_

Table 10: Gold standards in the CoNLL format for the 5th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	2	det	_	_
2	horse	_	NN	NN	_	14	nsubj	_	_
3	as	_	RB	RB	_	7	cc	_	_
4	well	_	RB	RB	_	3	fixed	_	_
5	as	_	IN	IN	_	3	fixed	_	_
6	the	_	DT	DT	_	7	det	_	_
7	rabbits	_	NNS	NNS	_	2	conj	_	_
8	which	_	WDT	WDT	_	10	obj	_	_
9	we	_	PRP	PRP	_	10	nsubj	_	_
10	wanted	_	VBD	VBD	_	7	rcmod	_	_
11	to	_	TO	TO	_	12	mark	_	_
12	eat	_	VB	VB	_	10	xcomp	_	_
13	have	_	VBP	VBP	_	14	aux	_	_
14	escaped	_	VBN	VBN	_	0	root	_	_

Table 11: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 6th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	2	det	_	_
2	horse	_	NN	NN	_	14	nsubj	_	_
3	as	_	IN	IN	_	4	fixed	_	_
4	well	_	RB	RB	_	2	cc	_	_
5	as	_	IN	IN	_	4	fixed	_	_
6	the	_	DT	DT	_	7	det	_	_
7	rabbits	_	NNS	NNS	_	2	conj	_	_
8	which	_	WP	WP	_	2	ref	_	_
9	we	_	PRP	PRP	_	10	nsubj	_	_
10	wanted	_	VBD	VBD	_	2	rcmod	_	_
11	to	_	TO	TO	_	12	aux	_	_
12	eat	_	VB	VB	_	10	xcomp	_	_
13	have	_	VBP	VBP	_	14	aux	_	_
14	escaped	_	VBN	VBN	_	0	root	_	_

Table 12: Gold standards in the CoNLL format for the 6th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	It	_	PRP	PRP	_	6	nsubj	_	_
2	was	_	VBD	VBD	_	6	cop	_	_
3	my	_	PRP$	PRP$	_	6	nmod:poss	_	_
4	aunt	_	NN	NN	_	6	compound	_	_
5	's	_	NN	NN	_	6	compound	_	_
6	car	_	NN	NN	_	0	root	_	_
7	which	_	WDT	WDT	_	9	obj	_	_
8	we	_	PRP	PRP	_	9	nsubj	_	_
9	sold	_	VBD	VBD	_	6	dep	_	_
10	at	_	IN	IN	_	11	prep	_	_
11	auction	_	NN	NN	_	9	pobj	_	_
12	last	_	JJ	JJ	_	13	amod	_	_
13	year	_	NN	NN	_	9	obl:tmod	_	_
14	in	_	IN	IN	_	15	prep	_	_
15	February	_	NNP	NNP	_	9	pobj	_	_

Table 13: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 7th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	It	_	PRP	PRP	_	2	nsubj	_	_
2	was	_	VBD	VBD	_	0	root	_	_
3	my	_	PRP$	PRP$	_	4	nmod:poss	_	_
4	aunt	_	NN	NN	_	6	nmod:poss	_	_
5	's	_	POS	POS	_	4	possessive	_	_
6	car	_	NN	NN	_	2	obj	_	_
7	which	_	WP	WP	_	6	ref	_	_
8	we	_	PRP	PRP	_	9	nsubj	_	_
9	sold	_	VBD	VBD	_	6	rcmod	_	_
10	at	_	IN	IN	_	9	prep	_	_
11	auction	_	NN	NN	_	10	pobj	_	_
12	last	_	JJ	JJ	_	13	amod	_	_
13	year	_	NN	NN	_	9	obl:tmod	_	_
14	in	_	IN	IN	_	9	prep	_	_
15	February	_	NNP	NNP	_	14	pobj	_	_

Table 14: Gold standards in the CoNLL format for the 7th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	3	det	_	_
2	only	_	JJ	JJ	_	3	amod	_	_
3	rabbit	_	NN	NN	_	9	nsubj:pass	_	_
4	that	_	IN	IN	_	7	mark	_	_
5	I	_	PRP	PRP	_	7	nsubj	_	_
6	ever	_	RB	RB	_	7	advmod	_	_
7	liked	_	VBD	VBD	_	3	dep	_	_
8	was	_	VBD	VBD	_	9	aux:pass	_	_
9	eaten	_	VBN	VBN	_	0	root	_	_
10	by	_	IN	IN	_	12	prep	_	_
11	my	_	PRP$	PRP$	_	12	nmod:poss	_	_
12	parents	_	NNS	NNS	_	9	pobj	_	_
13	one	_	CD	CD	_	14	nummod	_	_
14	summer	_	NN	NN	_	12	dep	_	_

Table 15: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 8th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	3	det	_	_
2	only	_	RB	RB	_	3	amod	_	_
3	rabbit	_	NN	NN	_	9	nsubj:pass	_	_
4	that	_	WP	WP	_	3	ref	_	_
5	I	_	PRP	PRP	_	7	nsubj	_	_
6	ever	_	RB	RB	_	7	advmod	_	_
7	liked	_	VBD	VBD	_	3	rcmod	_	_
8	was	_	VBD	VBD	_	9	aux:pass	_	_
9	eaten	_	VBN	VBN	_	0	root	_	_
10	by	_	IN	IN	_	9	prep	_	_
11	my	_	PRP$	PRP$	_	12	nmod:poss	_	_
12	parents	_	NNS	NNS	_	9	pobj	_	_
13	one	_	CD	CD	_	14	nummod	_	_
14	summer	_	NN	NN	_	9	obl:tmod	_	_

Table 16: Gold standards in the CoNLL format for the 8th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	2	det	_	_
2	veterans	_	NNS	NNS	_	14	nsubj	_	_
3	who	_	WP	WP	_	5	obj	_	_
4	I	_	PRP	PRP	_	5	nsubj	_	_
5	thought	_	VBD	VBD	_	2	rcmod	_	_
6	that	_	IN	IN	_	9	mark	_	_
7	we	_	PRP	PRP	_	9	nsubj	_	_
8	would	_	MD	MD	_	9	aux	_	_
9	meet	_	VB	VB	_	5	ccomp	_	_
10	at	_	IN	IN	_	12	prep	_	_
11	the	_	DT	DT	_	12	det	_	_
12	reunion	_	NN	NN	_	9	pobj	_	_
13	were	_	VBD	VBD	_	14	cop	_	_
14	dead	_	JJ	JJ	_	0	root	_	_

Table 17: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 9th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	2	det	_	_
2	veterans	_	NNS	NNS	_	14	nsubj	_	_
3	who	_	WP	WP	_	2	ref	_	_
4	I	_	PRP	PRP	_	5	nsubj	_	_
5	thought	_	VBD	VBD	_	2	rcmod	_	_
6	that	_	IN	IN	_	9	mark	_	_
7	we	_	PRP	PRP	_	9	nsubj	_	_
8	would	_	MD	MD	_	9	aux	_	_
9	meet	_	VB	VB	_	5	ccomp	_	_
10	at	_	IN	IN	_	9	prep	_	_
11	the	_	DT	DT	_	12	det	_	_
12	reunion	_	NN	NN	_	10	pobj	_	_
13	were	_	VBD	VBD	_	14	cop	_	_
14	dead	_	JJ	JJ	_	0	root	_	_

Table 18: Gold standards in the CoNLL format for the 9th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Natural	_	JJ	JJ	_	2	amod	_	_
2	disasters	_	NNS	NNS	_	0	root	_	_
3	–	_	SYM	SYM	_	4	dep	_	_
4	storms	_	NNS	NNS	_	2	nmod	_	_
6	flooding	_	NN	NN	_	2	appos	_	_
8	hurricanes	_	NNS	NNS	_	2	appos	_	_
10	occur	_	VB	VB	_	2	dep	_	_
11	infrequently	_	RB	RB	_	10	advmod	_	_
12	but	_	CC	CC	_	13	cc	_	_
13	cause	_	VB	VB	_	10	conj	_	_
14	devastation	_	NN	NN	_	13	obj	_	_
15	that	_	WDT	WDT	_	16	nsubj	_	_
16	strains	_	VBZ	VBZ	_	14	rcmod	_	_
17	resources	_	NNS	NNS	_	16	obj	_	_
18	to	_	IN	IN	_	19	mark	_	_
19	breaking	_	VBG	VBG	_	16	advcl	_	_
20	point	_	NN	NN	_	19	obj	_	_

Table 19: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 10^{th} sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Natural	_	JJ	JJ	_	2	amod	_	_
2	disasters	_	NNS	NNS	_	10	nsubj	_	_
3	–	_	:	:	_	4	punct	_	_
4	storms	_	NNS	NNS	_	2	appos	_	_
6	flooding	_	NN	NN	_	4	conj	_	_
8	hurricanes	_	NNS	NNS	_	4	conj	_	_
10	occur	_	VBP	VBP	_	0	root	_	_
11	infrequently	_	RB	RB	_	10	advmod	_	_
12	but	_	CC	CC	_	10	cc	_	_
13	cause	_	VBP	VBP	_	10	conj	_	_
14	devastation	_	NN	NN	_	13	obj	_	_
15	that	_	WP	WP	_	16	nsubj	_	_
16	strains	_	VBZ	VBZ	_	14	rcmod	_	_
17	resources	_	NNS	NNS	_	16	obj	_	_
18	to	_	TO	TO	_	16	prep	_	_
19	breaking	_	JJ	JJ	_	16	amod	_	_
20	point	_	NN	NN	_	18	pobj	_	_

Table 20: Gold standards in the CoNLL format for the 10^{th} sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Letters	_	NNPS	NNPS	_	12	nsubj	_	_
2	delivered	_	VBN	VBN	_	1	acl	_	_
3	on	_	IN	IN	_	4	prep	_	_
4	time	_	NN	NN	_	2	pobj	_	_
5	by	_	IN	IN	_	9	prep	_	_
6	old	_	JJ	JJ	_	8	amod	_	_
7	-	_	HYPH	HYPH	_	8	punct	_	_
8	fashioned	_	JJ	JJ	_	9	amod	_	_
9	means	_	NNS	NNS	_	2	pobj	_	_
10	are	_	VBP	VBP	_	12	cop	_	_
11	increasingly	_	RB	RB	_	12	advmod	_	_
12	rare	_	JJ	JJ	_	0	root	_	_
14	so	_	IN	IN	_	16	mark	_	_
15	it	_	PRP	PRP	_	16	nsubj	_	_
16	is	_	VBZ	VBZ	_	12	advcl	_	_
17	as	_	RB	RB	_	16	advmod	_	_
18	well	_	RB	RB	_	17	fixed	_	_
19	that	_	IN	IN	_	26	mark	_	_
20	that	_	DT	DT	_	26	nsubj	_	_
21	is	_	VBZ	VBZ	_	26	cop	_	_
22	not	_	RB	RB	_	26	advmod	_	_
23	the	_	DT	DT	_	25	det	_	_
24	only	_	JJ	JJ	_	25	amod	_	_
25	option	_	NN	NN	_	26	obl:npmod	_	_
26	available	_	JJ	JJ	_	17	advcl	_	_

Table 21: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 11^{th} sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Letters	_	NNS	NNS	_	12	nsubj	_	_
2	delivered	_	VBN	VBN	_	1	vmod	_	_
3	on	_	IN	IN	_	2	prep	_	_
4	time	_	NN	NN	_	3	pobj	_	_
5	by	_	IN	IN	_	2	prep	_	_
6	old	_	JJ	JJ	_	8	amod	_	_
7	-	_	:	:	_	8	punct	_	_
8	fashioned	_	JJ	JJ	_	9	amod	_	_
9	means	_	NNS	NNS	_	5	pobj	_	_
10	are	_	VBP	VBP	_	12	cop	_	_
11	increasingly	_	RB	RB	_	12	advmod	_	_
12	rare	_	JJ	JJ	_	0	root	_	_
14	so	_	RB	RB	_	18	mark	_	_
15	it	_	PRP	PRP	_	18	nsubj	_	_
16	is	_	VBZ	VBZ	_	18	cop	_	_
17	as	_	RB	RB	_	18	fixed	_	_
18	well	_	RB	RB	_	12	parataxis	_	_
19	that	_	IN	IN	_	25	mark	_	_
20	that	_	WDT	WDT	_	25	nsubj	_	_
21	is	_	VBZ	VBZ	_	25	cop	_	_
22	not	_	RB	RB	_	24	advmod	_	_
23	the	_	DT	DT	_	25	det	_	_
24	only	_	JJ	JJ	_	25	amod	_	_
25	option	_	NN	NN	_	18	ccomp	_	_
26	available	_	JJ	JJ	_	25	amod	_	_

Table 22: Gold standards in the CoNLL format for the 11^{th} sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	It	_	PRP	PRP	_	3	nsubj	_	_
2	wo	_	MD	MD	_	3	aux	_	_
3	n't	_	VB	VB	_	0	root	_	_
4	rain	_	NN	NN	_	3	obj	_	_
5	but	_	CC	CC	_	9	cc	_	_
6	there	_	EX	EX	_	9	expl	_	_
7	might	_	MD	MD	_	9	aux	_	_
8	be	_	VB	VB	_	9	cop	_	_
9	snow	_	NN	NN	_	3	conj	_	_
10	on	_	IN	IN	_	12	prep	_	_
11	high	_	JJ	JJ	_	12	amod	_	_
12	ground	_	NN	NN	_	9	nmod	_	_
13	if	_	IN	IN	_	16	mark	_	_
14	the	_	DT	DT	_	15	det	_	_
15	temperature	_	NN	NN	_	16	nsubj	_	_
16	stays	_	VBZ	VBZ	_	9	advcl	_	_
17	about	_	IN	IN	_	19	prep	_	_
18	the	_	DT	DT	_	19	det	_	_
19	same	_	JJ	JJ	_	16	pobj	_	_
20	over	_	IN	IN	_	24	prep	_	_
21	the	_	DT	DT	_	24	det	_	_
22	next	_	JJ	JJ	_	24	amod	_	_
23	24	_	CD	CD	_	24	nummod	_	_
24	hours	_	NNS	NNS	_	19	nmod	_	_

Table 23: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 12^{th} sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	It	_	PRP	PRP	_	4	nsubj	_	_
2	wo	_	MD	MD	_	4	aux	_	_
3	n't	_	RB	RB	_	4	neg	_	_
4	rain	_	VB	VB	_	0	root	_	_
5	but	_	CC	CC	_	4	cc	_	_
6	there	_	EX	EX	_	8	expl	_	_
7	might	_	MD	MD	_	8	aux	_	_
8	be	_	VB	VB	_	4	conj	_	_
9	snow	_	NN	NN	_	8	nsubj	_	_
10	on	_	IN	IN	_	8	prep	_	_
11	high	_	JJ	JJ	_	12	amod	_	_
12	ground	_	NN	NN	_	10	pobj	_	_
13	if	_	IN	IN	_	16	mark	_	_
14	the	_	DT	DT	_	15	det	_	_
15	temperature	_	NN	NN	_	16	nsubj	_	_
16	stays	_	VBZ	VBZ	_	8	advcl	_	_
17	about	_	RB	RB	_	19	advmod	_	_
18	the	_	DT	DT	_	19	det	_	_
19	same	_	JJ	JJ	_	16	acomp	_	_
20	over	_	IN	IN	_	16	prep	_	_
21	the	_	DT	DT	_	24	det	_	_
22	next	_	JJ	JJ	_	24	amod	_	_
23	24	_	CD	CD	_	24	nummod	_	_
24	hours	_	NNS	NNS	_	20	pobj	_	_

Table 24: Gold standards in the CoNLL format for the 12th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	5	det	_	_
2	long	_	JJ	JJ	_	5	amod	_	_
3	and	_	CC	CC	_	4	cc	_	_
4	lonely	_	JJ	JJ	_	2	conj	_	_
5	road	_	NN	NN	_	8	nsubj	_	_
6	to	_	IN	IN	_	7	prep	_	_
7	redemption	_	NN	NN	_	5	nmod	_	_
8	begins	_	VBZ	VBZ	_	0	root	_	_
9	with	_	IN	IN	_	12	prep	_	_
10	self	_	NN	NN	_	12	compound	_	_
11	-	_	HYPH	HYPH	_	12	punct	_	_
12	reflection	_	NN	NN	_	8	pobj	_	_
14	the	_	DT	DT	_	15	det	_	_
15	need	_	NN	NN	_	8	obj	_	_
16	to	_	TO	TO	_	17	mark	_	_
17	delve	_	VB	VB	_	8	parataxis	_	_
18	inwards	_	RB	RB	_	17	advmod	_	_
19	to	_	IN	IN	_	21	prep	_	_
20	deconstruct	_	JJ	JJ	_	21	amod	_	_
21	layers	_	NNS	NNS	_	17	pobj	_	_
22	of	_	IN	IN	_	24	prep	_	_
23	psychological	_	JJ	JJ	_	24	amod	_	_
24	obfuscation	_	NN	NN	_	21	nmod	_	_

Table 25: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 13th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	5	det	_	_
2	long	_	JJ	JJ	_	5	amod	_	_
3	and	_	CC	CC	_	2	cc	_	_
4	lonely	_	JJ	JJ	_	2	conj	_	_
5	road	_	NN	NN	_	8	nsubj	_	_
6	to	_	IN	IN	_	5	prep	_	_
7	redemption	_	NN	NN	_	6	pobj	_	_
8	begins	_	VBZ	VBZ	_	0	root	_	_
9	with	_	IN	IN	_	8	prep	_	_
10	self	_	NN	NN	_	12	compound	_	_
11	-	_	:	:	_	12	punct	_	_
12	reflection	_	NN	NN	_	9	pobj	_	_
14	the	_	DT	DT	_	15	det	_	_
15	need	_	NN	NN	_	12	appos	_	_
16	to	_	TO	TO	_	17	aux	_	_
17	delve	_	VB	VB	_	15	vmod	_	_
18	inwards	_	RB	RB	_	17	advmod	_	_
19	to	_	IN	IN	_	20	aux	_	_
20	deconstruct	_	VB	VB	_	17	vmod	_	_
21	layers	_	NNS	NNS	_	20	obj	_	_
22	of	_	IN	IN	_	21	prep	_	_
23	psychological	_	JJ	JJ	_	24	amod	_	_
24	obfuscation	_	NN	NN	_	22	pobj	_	_

Table 26: Gold standards in the CoNLL format for the 13th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	My	_	PRP$	PRP$	_	3	nmod:poss	_	_
2	wildest	_	JJ	JJ	_	3	amod	_	_
3	dream	_	NN	NN	_	4	nsubj	_	_
4	is	_	VBZ	VBZ	_	0	root	_	_
5	to	_	TO	TO	_	6	mark	_	_
6	build	_	VB	VB	_	4	xcomp	_	_
7	a	_	DT	DT	_	9	det	_	_
8	POS	_	NN	NN	_	9	dep	_	_
9	tagger	_	NN	NN	_	6	obj	_	_
10	which	_	WDT	WDT	_	11	nsubj	_	_
11	processes	_	VBZ	VBZ	_	9	rcmod	_	_
12	10K	_	CD	CD	_	13	nummod	_	_
13	words	_	NNS	NNS	_	11	obj	_	_
14	per	_	IN	IN	_	15	prep	_	_
15	second	_	NN	NN	_	13	nmod	_	_
16	and	_	CC	CC	_	17	cc	_	_
17	uses	_	VBZ	VBZ	_	11	conj	_	_
18	only	_	RB	RB	_	19	advmod	_	_
19	1	_	CD	CD	_	20	nummod	_	_
20	MB	_	NNS	NNS	_	17	obj	_	_
21	of	_	IN	IN	_	22	prep	_	_
22	RAM	_	NNP	NNP	_	20	nmod	_	_
24	but	_	CC	CC	_	27	cc	_	_
25	it	_	PRP	PRP	_	27	nsubj	_	_
26	may	_	MD	MD	_	27	aux	_	_
27	prove	_	VB	VB	_	4	conj	_	_
28	too	_	RB	RB	_	29	advmod	_	_
29	hard	_	JJ	JJ	_	27	xcomp	_	_

Table 27: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 14^{th} sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	My	_	PRP$	PRP$	_	3	nmod:poss	_	_
2	wildest	_	JJS	JJS	_	3	amod	_	_
3	dream	_	NN	NN	_	4	nsubj	_	_
4	is	_	VBZ	VBZ	_	0	root	_	_
5	to	_	TO	TO	_	6	aux	_	_
6	build	_	VB	VB	_	4	xcomp	_	_
7	a	_	DT	DT	_	9	det	_	_
8	POS	_	NN	NN	_	9	compound	_	_
9	tagger	_	NN	NN	_	6	obj	_	_
10	which	_	WP	WP	_	11	nsubj	_	_
11	processes	_	VBZ	VBZ	_	9	rcmod	_	_
12	10K	_	CD	CD	_	13	nummod	_	_
13	words	_	NNS	NNS	_	11	obj	_	_
14	per	_	IN	IN	_	11	prep	_	_
15	second	_	NN	NN	_	14	pobj	_	_
16	and	_	CC	CC	_	11	cc	_	_
17	uses	_	VBZ	VBZ	_	11	conj	_	_
18	only	_	RB	RB	_	19	advmod	_	_
19	1	_	CD	CD	_	20	nummod	_	_
20	MB	_	NNP	NNP	_	17	obj	_	_
21	of	_	IN	IN	_	20	prep	_	_
22	RAM	_	NNP	NNP	_	21	pobj	_	_
24	but	_	CC	CC	_	4	cc	_	_
25	it	_	PRP	PRP	_	27	nsubj	_	_
26	may	_	MD	MD	_	27	aux	_	_
27	prove	_	VB	VB	_	4	conj	_	_
28	too	_	RB	RB	_	29	advmod	_	_
29	hard	_	JJ	JJ	_	27	acomp	_	_

Table 28: Gold standards in the CoNLL format for the 14th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	English	_	NNP	NNP	_	3	nsubj	_	_
2	also	_	RB	RB	_	3	advmod	_	_
3	has	_	VBZ	VBZ	_	0	root	_	_
4	many	_	JJ	JJ	_	5	amod	_	_
5	words	_	NNS	NNS	_	3	obj	_	_
6	of	_	IN	IN	_	11	prep	_	_
7	more	_	RBR	RBR	_	10	advmod	_	_
8	or	_	CC	CC	_	9	cc	_	_
9	less	_	RBR	RBR	_	7	conj	_	_
10	unique	_	JJ	JJ	_	11	amod	_	_
11	function	_	NN	NN	_	5	nmod	_	_
13	including	_	VBG	VBG	_	14	prep	_	_
14	interjections	_	NNS	NNS	_	11	nmod	_	_
16	oh	_	CD	CD	_	14	dep	_	_
18	ah	_	JJ	JJ	_	16	dep	_	_
21	negatives	_	NNS	NNS	_	14	conj	_	_
23	no	_	UH	UH	_	25	discourse	_	_
25	not	_	RB	RB	_	21	dep	_	_
28	politeness	_	JJ	JJ	_	29	amod	_	_
29	markers	_	NNS	NNS	_	14	conj	_	_
31	please	_	UH	UH	_	34	discourse	_	_
33	thank	_	UH	UH	_	31	dep	_	_
34	you	_	PRP	PRP	_	29	dep	_	_
37	and	_	CC	CC	_	41	cc	_	_
38	the	_	DT	DT	_	41	det	_	_
39	existential	_	JJ	JJ	_	41	amod	_	_
41	there	_	RB	RB	_	14	conj	_	_
44	there	_	EX	EX	_	45	expl	_	_
45	are	_	VBP	VBP	_	5	dep	_	_
46	horses	_	NNS	NNS	_	45	nsubj	_	_
47	but	_	CC	CC	_	48	cc	_	_
48	not	_	RB	RB	_	49	cc	_	_
49	unicorns	_	NNS	NNS	_	46	conj	_	_
51	among	_	IN	IN	_	52	prep	_	_
52	others	_	NNS	NNS	_	5	nmod	_	_

Table 29: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 15th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	English	_	NNP	NNP	_	3	nsubj	_	_
2	also	_	RB	RB	_	3	advmod	_	_
3	has	_	VBZ	VBZ	_	0	root	_	_
4	many	_	PDT	PDT	_	5	amod	_	_
5	words	_	NNS	NNS	_	3	obj	_	_
6	of	_	IN	IN	_	5	prep	_	_
7	more	_	RBR	RBR	_	11	advmod	_	_
8	or	_	CC	CC	_	7	cc	_	_
9	less	_	RBR	RBR	_	7	conj	_	_
10	unique	_	JJ	JJ	_	11	amod	_	_
11	function	_	NN	NN	_	6	pobj	_	_
13	including	_	VBG	VBG	_	5	vmod	_	_
14	interjections	_	NNS	NNS	_	13	obj	_	_
16	oh	_	UH	UH	_	14	appos	_	_
18	ah	_	UH	UH	_	16	conj	_	_
21	negatives	_	NNS	NNS	_	14	conj	_	_
23	no	_	DT	DT	_	21	appos	_	_
25	not	_	RB	RB	_	23	conj	_	_
28	politeness	_	NN	NN	_	29	compound	_	_
29	markers	_	NNS	NNS	_	14	conj	_	_
31	please	_	UH	UH	_	29	appos	_	_
33	thank	_	VBP	VBP	_	31	conj	_	_
34	you	_	PRP	PRP	_	33	obj	_	_
37	and	_	CC	CC	_	14	cc	_	_
38	the	_	DT	DT	_	41	det	_	_
39	existential	_	JJ	JJ	_	41	amod	_	_
41	there	_	EX	EX	_	14	conj	_	_
44	there	_	EX	EX	_	45	expl	_	_
45	are	_	VBP	VBP	_	41	parataxis	_	_
46	horses	_	NNS	NNS	_	45	nsubj	_	_
47	but	_	CC	CC	_	46	cc	_	_
48	not	_	RB	RB	_	49	neg	_	_
49	unicorns	_	NNS	NNS	_	46	conj	_	_
51	among	_	IN	IN	_	14	prep	_	_
52	others	_	NNS	NNS	_	51	pobj	_	_

Table 30: Gold standards in the CoNLL format for the 15th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Making	_	VBG	VBG	_	4	csubj	_	_
2	these	_	DT	DT	_	3	det	_	_
3	decisions	_	NNS	NNS	_	1	obj	_	_
4	requires	_	VBZ	VBZ	_	0	root	_	_
5	sophisticated	_	JJ	JJ	_	6	amod	_	_
6	knowledge	_	NN	NN	_	4	obj	_	_
7	of	_	IN	IN	_	8	prep	_	_
8	syntax	_	NN	NN	_	6	nmod	_	_
10	tagging	_	NN	NN	_	11	compound	_	_
11	manuals	_	NNS	NNS	_	17	nsubj	_	_
13	Santorini	_	NNP	NNP	_	11	appos	_	_
15	1990	_	CD	CD	_	13	dep	_	_
17	give	_	VBP	VBP	_	4	parataxis	_	_
18	various	_	JJ	JJ	_	19	amod	_	_
19	heuristics	_	NNS	NNS	_	17	iobj	_	_
20	that	_	WDT	WDT	_	22	nsubj	_	_
21	can	_	MD	MD	_	22	aux	_	_
22	help	_	VB	VB	_	17	obj	_	_
23	human	_	JJ	JJ	_	24	amod	_	_
24	coders	_	NNS	NNS	_	25	nsubj	_	_
25	make	_	VB	VB	_	22	ccomp	_	_
26	these	_	DT	DT	_	27	det	_	_
27	decisions	_	NNS	NNS	_	25	obj	_	_
28	and	_	CC	CC	_	32	cc	_	_
29	that	_	WDT	WDT	_	32	nsubj	_	_
30	can	_	MD	MD	_	32	aux	_	_
31	also	_	RB	RB	_	32	advmod	_	_
32	provide	_	VB	VB	_	22	conj	_	_
33	useful	_	JJ	JJ	_	34	amod	_	_
34	features	_	NNS	NNS	_	32	obj	_	_
35	for	_	IN	IN	_	37	prep	_	_
36	automatic	_	JJ	JJ	_	37	amod	_	_
37	taggers	_	NNS	NNS	_	32	pobj	_	_

Table 31: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 16th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Making	_	VBG	VBG	_	4	csubj	_	_
2	these	_	DT	DT	_	3	det	_	_
3	decisions	_	NNS	NNS	_	1	obj	_	_
4	requires	_	VBZ	VBZ	_	0	root	_	_
5	sophisticated	_	JJ	JJ	_	6	amod	_	_
6	knowledge	_	NN	NN	_	4	obj	_	_
7	of	_	IN	IN	_	6	prep	_	_
8	syntax	_	NN	NN	_	7	pobj	_	_
10	tagging	_	JJ	JJ	_	11	amod	_	_
11	manuals	_	NNS	NNS	_	17	nsubj	_	_
13	Santorini	_	NNP	NNP	_	11	appos	_	_
15	1990	_	CD	CD	_	13	npadvmod	_	_
17	give	_	VBP	VBP	_	4	parataxis	_	_
18	various	_	JJ	JJ	_	19	amod	_	_
19	heuristics	_	NNS	NNS	_	17	obj	_	_
20	that	_	WP	WP	_	22	nsubj	_	_
21	can	_	MD	MD	_	22	aux	_	_
22	help	_	VB	VB	_	19	rcmod	_	_
23	human	_	JJ	JJ	_	24	amod	_	_
24	coders	_	NNS	NNS	_	22	obj	_	_
25	make	_	VB	VB	_	22	ccomp	_	_
26	these	_	DT	DT	_	27	det	_	_
27	decisions	_	NNS	NNS	_	25	obj	_	_
28	and	_	CC	CC	_	22	cc	_	_
29	that	_	WP	WP	_	32	nsubj	_	_
30	can	_	MD	MD	_	32	aux	_	_
31	also	_	RB	RB	_	32	advmod	_	_
32	provide	_	VB	VB	_	22	conj	_	_
33	useful	_	JJ	JJ	_	34	amod	_	_
34	features	_	NNS	NNS	_	32	obj	_	_
35	for	_	IN	IN	_	34	prep	_	_
36	automatic	_	JJ	JJ	_	37	amod	_	_
37	taggers	_	NNS	NNS	_	35	pobj	_	_

Table 32: Gold standards in the CoNLL format for the 16th sentence of the handout.

height=0.77

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	4	det	_	_
2	Penn	_	NNP	NNP	_	4	compound	_	_
3	Treebank	_	NNP	NNP	_	4	compound	_	_
4	tagset	_	NNP	NNP	_	6	nsubj:pass	_	_
5	was	_	VBD	VBD	_	6	aux:pass	_	_
6	culled	_	VBN	VBN	_	0	root	_	_
7	from	_	IN	IN	_	13	prep	_	_
8	the	_	DT	DT	_	13	det	_	_
9	original	_	JJ	JJ	_	13	amod	_	_
10	87	_	CD	CD	_	12	nummod	_	_
11	-	_	HYPH	HYPH	_	12	punct	_	_
12	tag	_	NN	NN	_	13	compound	_	_
13	tagset	_	NN	NN	_	6	pobj	_	_
14	for	_	IN	IN	_	17	prep	_	_
15	the	_	DT	DT	_	17	det	_	_
16	Brown	_	NNP	NNP	_	17	compound	_	_
17	Corpus	_	NNP	NNP	_	13	nmod	_	_
1	For	_	IN	IN	_	2	prep	_	_
2	example	_	NN	NN	_	9	pobj	_	_
3	the	_	DT	DT	_	8	det	_	_
4	original	_	JJ	JJ	_	8	amod	_	_
5	Brown	_	NNP	NNP	_	8	compound	_	_
6	and	_	CC	CC	_	7	cc	_	_
7	C5	_	NNP	NNP	_	5	conj	_	_
8	tagsets	_	NNS	NNS	_	9	nsubj	_	_
9	include	_	VBP	VBP	_	0	root	_	_
10	a	_	DT	DT	_	12	det	_	_
11	separate	_	JJ	JJ	_	12	amod	_	_
12	tag	_	NN	NN	_	22	nsubj	_	_
13	for	_	IN	IN	_	14	prep	_	_
14	each	_	DT	DT	_	12	nmod	_	_
15	of	_	IN	IN	_	18	prep	_	_
16	the	_	DT	DT	_	18	det	_	_
17	different	_	JJ	JJ	_	18	amod	_	_
18	forms	_	NNS	NNS	_	14	nmod	_	_
19	of	_	IN	IN	_	21	prep	_	_
20	the	_	DT	DT	_	21	det	_	_
21	verbs	_	NNS	NNS	_	18	nmod	_	_
22	do	_	VB	VB	_	9	ccomp	_	_
24	e.g.	_	FW	FW	_	27	dep	_	_
25	C5	_	NNP	NNP	_	27	compound	_	_
26	tag	_	NNP	NNP	_	27	compound	_	_
27	VDD	_	NNP	NNP	_	22	dep	_	_
28	for	_	IN	IN	_	29	mark	_	_
29	did	_	VBD	VBD	_	27	dep	_	_
30	and	_	CC	CC	_	31	cc	_	_
31	VDG	_	VBD	VBD	_	29	conj	_	_
32	tag	_	NN	NN	_	29	obj	_	_
33	for	_	IN	IN	_	34	mark	_	_
34	doing	_	VBG	VBG	_	29	advcl	_	_
37	be	_	VB	VB	_	22	conj	_	_

height=0.77

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	4	det	_	_
2	Penn	_	NNP	NNP	_	3	compound	_	_
3	Treebank	_	NNP	NNP	_	4	compound	_	_
4	tagset	_	NN	NN	_	6	nsubj:pass	_	_
5	was	_	VBD	VBD	_	6	aux:pass	_	_
6	culled	_	VBN	VBN	_	0	root	_	_
7	from	_	IN	IN	_	6	prep	_	_
8	the	_	DT	DT	_	13	det	_	_
9	original	_	JJ	JJ	_	13	amod	_	_
10	87	_	CD	CD	_	12	nummod	_	_
11	-	_	:	:	_	12	punct	_	_
12	tag	_	NN	NN	_	13	compound	_	_
13	tagset	_	NN	NN	_	7	pobj	_	_
14	for	_	IN	IN	_	13	prep	_	_
15	the	_	DT	DT	_	17	det	_	_
16	Brown	_	NNP	NNP	_	17	compound	_	_
17	Corpus	_	NNP	NNP	_	14	pobj	_	_
1	For	_	IN	IN	_	9	prep	_	_
2	example	_	NN	NN	_	1	pobj	_	_
3	the	_	DT	DT	_	8	det	_	_
4	original	_	JJ	JJ	_	8	amod	_	_
5	Brown	_	NNP	NNP	_	8	compound	_	_
6	and	_	CC	CC	_	5	cc	_	_
7	C5	_	NNP	NNP	_	5	conj	_	_
8	tagsets	_	NNS	NNS	_	9	nsubj	_	_
9	include	_	VBP	VBP	_	0	root	_	_
10	a	_	DT	DT	_	12	det	_	_
11	separate	_	JJ	JJ	_	12	amod	_	_
12	tag	_	NN	NN	_	9	obj	_	_
13	for	_	IN	IN	_	12	prep	_	_
14	each	_	PDT	PDT	_	13	pobj	_	_
15	of	_	IN	IN	_	14	prep	_	_
16	the	_	DT	DT	_	18	det	_	_
17	different	_	JJ	JJ	_	18	amod	_	_
18	forms	_	NNS	NNS	_	15	pobj	_	_
19	of	_	IN	IN	_	18	prep	_	_
20	the	_	DT	DT	_	21	det	_	_
21	verbs	_	NNS	NNS	_	22	pobj	_	_
22	do	_	VB	VB	_	19	npadvmod	_	_
24	e.g.	_	FW	FW	_	22	prep	_	_
25	C5	_	NNP	NNP	_	27	compound	_	_
26	tag	_	NN	NN	_	27	compound	_	_
27	VDD	_	NNP	NNP	_	24	pobj	_	_
28	for	_	IN	IN	_	27	prep	_	_
29	did	_	VBD	VBD	_	28	pobj	_	_
30	and	_	CC	CC	_	27	cc	_	_
31	VDG	_	NNP	NNP	_	32	compound	_	_
32	tag	_	NN	NN	_	27	conj	_	_
33	for	_	IN	IN	_	32	prep	_	_
34	doing	_	VBG	VBG	_	33	pobj	_	_
37	be	_	VB	VB	_	22	conj	_	_

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	4	det	_	_
2	slightly	_	RB	RB	_	4	advmod	_	_
3	simplified	_	VBN	VBN	_	4	amod	_	_
4	version	_	NN	NN	_	12	nsubj	_	_
5	of	_	IN	IN	_	8	prep	_	_
6	the	_	DT	DT	_	8	det	_	_
7	Viterbi	_	NNP	NNP	_	8	compound	_	_
8	algorithm	_	NN	NN	_	4	nmod	_	_
9	that	_	IN	IN	_	11	mark	_	_
10	we	_	PRP	PRP	_	11	nsubj	_	_
11	present	_	VBP	VBP	_	8	dep	_	_
12	takes	_	VBZ	VBZ	_	24	dep	_	_
13	as	_	IN	IN	_	14	prep	_	_
14	input	_	NN	NN	_	12	pobj	_	_
15	a	_	DT	DT	_	17	det	_	_
16	single	_	JJ	JJ	_	17	amod	_	_
17	HMM	_	NN	NN	_	14	dep	_	_
18	and	_	CC	CC	_	20	cc	_	_
19	a	_	DT	DT	_	20	det	_	_
20	sequence	_	NN	NN	_	14	conj	_	_
21	of	_	IN	IN	_	23	prep	_	_
22	observed	_	VBN	VBN	_	23	amod	_	_
23	words	_	NNS	NNS	_	20	nmod	_	_
24	O	_	NN	NN	_	0	root	_	_
27	o1	_	CD	CD	_	24	dep	_	_
29	o2	_	NNP	NNP	_	27	dep	_	_
32	oT	_	NN	NN	_	29	dep	_	_
34	and	_	CC	CC	_	24	cc	_	_
35	returns	_	NNS	NNS	_	24	dep	_	_
36	the	_	DT	DT	_	39	det	_	_
37	most	_	RBS	RBS	_	38	advmod	_	_
38	probable	_	JJ	JJ	_	39	amod	_	_
39	state	_	NN	NN	_	35	dep	_	_
41	tag	_	NN	NN	_	43	compound	_	_
42	sequence	_	NN	NN	_	43	compound	_	_
43	Q	_	NN	NN	_	46	nsubj	_	_
44	=	_	SYM	SYM	_	46	dep	_	_
46	q1	_	NNS	NNS	_	35	dep	_	_
48	q2	_	NNS	NNS	_	46	appos	_	_
50	qT	_	NN	NN	_	46	appos	_	_
52	together	_	RB	RB	_	55	advmod	_	_
53	with	_	IN	IN	_	55	prep	_	_
54	its	_	PRP$	PRP$	_	55	nmod:poss	_	_
55	probability	_	NN	NN	_	46	nmod	_	_

Table 35: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 18th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	4	det	_	_
2	slightly	_	RB	RB	_	3	advmod	_	_
3	simplified	_	JJ	JJ	_	4	amod	_	_
4	version	_	NN	NN	_	12	nsubj	_	_
5	of	_	IN	IN	_	4	prep	_	_
6	the	_	DT	DT	_	8	det	_	_
7	Viterbi	_	NNP	NNP	_	8	compound	_	_
8	algorithm	_	NN	NN	_	5	pobj	_	_
9	that	_	WP	WP	_	4	ref	_	_
10	we	_	PRP	PRP	_	11	nsubj	_	_
11	present	_	VBP	VBP	_	4	rcmod	_	_
12	takes	_	VBZ	VBZ	_	0	root	_	_
13	as	_	IN	IN	_	12	prep	_	_
14	input	_	NN	NN	_	13	pobj	_	_
15	a	_	DT	DT	_	17	det	_	_
16	single	_	JJ	JJ	_	17	amod	_	_
17	HMM	_	NNP	NNP	_	12	obj	_	_
18	and	_	CC	CC	_	17	cc	_	_
19	a	_	DT	DT	_	20	det	_	_
20	sequence	_	NN	NN	_	17	conj	_	_
21	of	_	IN	IN	_	20	prep	_	_
22	observed	_	JJ	JJ	_	23	amod	_	_
23	words	_	NNS	NNS	_	21	pobj	_	_
24	O	_	SYM	SYM	_	20	appos	_	_
27	o1	_	SYM	SYM	_	24	appos	_	_
29	o2	_	SYM	SYM	_	27	conj	_	_
32	oT	_	SYM	SYM	_	29	conj	_	_
34	and	_	CC	CC	_	12	cc	_	_
35	returns	_	VBZ	VBZ	_	12	conj	_	_
36	the	_	DT	DT	_	42	det	_	_
37	most	_	RBS	RBS	_	38	advmod	_	_
38	probable	_	JJ	JJ	_	42	amod	_	_
39	state	_	NN	NN	_	42	compound	_	_
41	tag	_	NN	NN	_	39	conj	_	_
42	sequence	_	NN	NN	_	35	obj	_	_
43	Q	_	SYM	SYM	_	42	appos	_	_
44	=	_	SYM	SYM	_	43	punct	_	_
46	q1	_	SYM	SYM	_	43	appos	_	_
48	q2	_	SYM	SYM	_	46	conj	_	_
50	qT	_	SYM	SYM	_	48	conj	_	_
52	together	_	RB	RB	_	53	advmod	_	_
53	with	_	IN	IN	_	42	prep	_	_
54	its	_	PRP$	PRP$	_	55	nmod:poss	_	_
55	probability	_	NN	NN	_	53	pobj	_	_

Table 36: Gold standards in the CoNLL format for the 18th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Thus	_	RB	RB	_	14	advmod	_	_
2	the	_	DT	DT	_	10	det	_	_
3	EM	_	JJ	JJ	_	5	dep	_	_
4	-	_	HYPH	HYPH	_	5	punct	_	_
5	trained	_	VBN	VBN	_	10	amod	_	_
7	pure	_	JJ	JJ	_	8	amod	_	_
8	HMM	_	NN	NN	_	10	compound	_	_
10	tagger	_	NN	NN	_	14	nsubj	_	_
11	is	_	VBZ	VBZ	_	14	cop	_	_
12	probably	_	RB	RB	_	14	advmod	_	_
13	best	_	RBS	RBS	_	14	advmod	_	_
14	suited	_	JJ	JJ	_	0	root	_	_
15	to	_	IN	IN	_	16	prep	_	_
16	cases	_	NNS	NNS	_	14	pobj	_	_
17	where	_	WRB	WRB	_	22	advmod	_	_
18	no	_	DT	DT	_	20	det	_	_
19	training	_	NN	NN	_	20	compound	_	_
20	data	_	NNS	NNS	_	22	nsubj	_	_
21	is	_	VBZ	VBZ	_	22	cop	_	_
22	available	_	JJ	JJ	_	14	advcl	_	_
24	for	_	IN	IN	_	25	prep	_	_
25	example	_	NN	NN	_	22	pobj	_	_
27	when	_	WRB	WRB	_	29	advmod	_	_
28	tagging	_	JJ	JJ	_	29	amod	_	_
29	languages	_	NNS	NNS	_	38	obj	_	_
30	for	_	IN	IN	_	31	prep	_	_
31	which	_	WDT	WDT	_	29	nmod	_	_
32	no	_	DT	DT	_	33	det	_	_
33	data	_	NNS	NNS	_	38	nsubj:pass	_	_
34	was	_	VBD	VBD	_	38	aux:pass	_	_
35	previously	_	RB	RB	_	38	advmod	_	_
36	hand	_	NN	NN	_	38	dep	_	_
37	-	_	HYPH	HYPH	_	38	punct	_	_
38	tagged	_	VBN	VBN	_	25	rcmod	_	_

Table 37: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 19th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Thus	_	RB	RB	_	14	advmod	_	_
2	the	_	DT	DT	_	10	det	_	_
3	EM	_	NNP	NNP	_	5	amod	_	_
4	-	_	:	:	_	5	punct	_	_
5	trained	_	JJ	JJ	_	10	amod	_	_
7	pure	_	JJ	JJ	_	8	amod	_	_
8	HMM	_	NNP	NNP	_	10	compound	_	_
10	tagger	_	NN	NN	_	14	nsubj	_	_
11	is	_	VBZ	VBZ	_	14	cop	_	_
12	probably	_	RB	RB	_	14	advmod	_	_
13	best	_	RBS	RBS	_	14	advmod	_	_
14	suited	_	JJ	JJ	_	0	root	_	_
15	to	_	IN	IN	_	14	prep	_	_
16	cases	_	NNS	NNS	_	15	pobj	_	_
17	where	_	WP	WP	_	16	ref	_	_
18	no	_	DT	DT	_	20	det	_	_
19	training	_	NN	NN	_	20	compound	_	
20	data	_	NN	NN	_	22	nsubj	_	_
21	is	_	VBZ	VBZ	_	22	cop	_	_
22	available	_	JJ	JJ	_	14	rcmod	_	_
24	for	_	IN	IN	_	16	prep	_	_
25	example	_	NN	NN	_	24	pobj	_	_
27	when	_	WRB	WRB	_	28	advmod	_	_
28	tagging	_	JJ	JJ	_	14	amod	_	_
29	languages	_	NNS	NNS	_	28	obj	_	_
30	for	_	IN	IN	_	38	prep	_	_
31	which	_	WDT	WDT	_	30	pobj	_	_
32	no	_	DT	DT	_	33	det	_	_
33	data	_	NN	NN	_	38	nsubj:pass	_	_
34	was	_	VBD	VBD	_	38	aux:pass	_	_
35	previously	_	RB	RB	_	38	advmod	_	_
36	hand	_	NN	NN	_	38	compound	_	_
37	-	_	:	:	_	38	punct	_	_
38	tagged	_	JJ	JJ	_	29	rcmod	_	_

Table 38: Gold standards in the CoNLL format for the 19th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Coming	_	VBG	VBG	_	11	advcl	_	_
2	home	_	RB	RB	_	1	advmod	_	_
3	from	_	IN	IN	_	6	prep	_	_
4	very	_	RB	RB	_	5	advmod	_	_
5	lonely	_	JJ	JJ	_	6	amod	_	_
6	places	_	NNS	NNS	_	1	pobj	_	_
8	all	_	DT	DT	_	11	nsubj	_	_
9	of	_	IN	IN	_	10	prep	_	_
10	us	_	PRP	PRP	_	8	nmod	_	_
11	go	_	VBP	VBP	_	0	root	_	_
12	a	_	DT	DT	_	13	det	_	_
13	little	_	JJ	JJ	_	14	obl:npmod	_	_
14	mad	_	JJ	JJ	_	11	xcomp	_	_
16	whether	_	IN	IN	_	34	mark	_	_
17	from	_	IN	IN	_	20	prep	_	_
18	great	_	JJ	JJ	_	20	amod	_	_
19	personal	_	JJ	JJ	_	20	amod	_	_
20	success	_	NN	NN	_	34	pobj	_	_
22	or	_	CC	CC	_	23	cc	_	_
23	just	_	RB	RB	_	28	cc	_	_
24	an	_	DT	DT	_	28	det	_	_
25	all	_	DT	DT	_	27	det	_	_
26	-	_	HYPH	HYPH	_	27	punct	_	_
27	night	_	NN	NN	_	28	compound	_	_
28	drive	_	NN	NN	_	20	conj	_	_
30	we	_	PRP	PRP	_	34	nsubj	_	_
31	are	_	VBP	VBP	_	34	cop	_	_
32	the	_	DT	DT	_	34	det	_	_
33	sole	_	JJ	JJ	_	34	amod	_	_
34	survivors	_	NNS	NNS	_	11	ccomp	_	_
35	of	_	IN	IN	_	37	prep	_	_
36	a	_	DT	DT	_	37	det	_	_
37	world	_	NN	NN	_	34	nmod	_	_
38	no	_	DT	DT	_	39	det	_	_
39	one	_	NN	NN	_	43	nsubj	_	_
40	else	_	RB	RB	_	39	advmod	_	_
41	has	_	VBZ	VBZ	_	43	aux	_	_
42	ever	_	RB	RB	_	43	advmod	_	_
43	seen	_	VBN	VBN	_	37	rcmod	_	_

Table 39: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 20th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Coming	_	VBG	VBG	_	11	appos	_	_
2	home	_	NN	NN	_	1	obj	_	_
3	from	_	IN	IN	_	1	prep	_	_
4	very	_	RB	RB	_	6	advmod	_	_
5	lonely	_	JJ	JJ	_	6	amod	_	_
6	places	_	NNS	NNS	_	3	pobj	_	_
8	all	_	PDT	PDT	_	11	nsubj	_	_
9	of	_	IN	IN	_	8	prep	_	_
10	us	_	PRP	PRP	_	9	pobj	_	_
11	go	_	VBP	VBP	_	0	root	_	_
12	a	_	DT	DT	_	13	det	_	_
13	little	_	JJ	JJ	_	14	amod	_	_
14	mad	_	JJ	JJ	_	11	acomp	_	_
16	whether	_	WRB	WRB	_	34	advmod	_	_
17	from	_	IN	IN	_	16	prep	_	_
18	great	_	JJ	JJ	_	20	amod	_	_
19	personal	_	JJ	JJ	_	20	amod	_	_
20	success	_	NN	NN	_	17	pobj	_	_
22	or	_	CC	CC	_	20	cc	_	_
23	just	_	RB	RB	_	28	advmod	_	_
24	an	_	DT	DT	_	28	det	_	_
25	all	_	PDT	PDT	_	27	amod	_	_
26	-	_	:	:	_	27	punct	_	_
27	night	_	NN	NN	_	28	compound	_	_
28	drive	_	NN	NN	_	20	conj	_	_
30	we	_	PRP	PRP	_	34	nsubj	_	_
31	are	_	VBP	VBP	_	34	cop	_	_
32	the	_	DT	DT	_	34	det	_	_
33	sole	_	JJ	JJ	_	34	amod	_	_
34	survivors	_	NNS	NNS	_	11	parataxis	_	_
35	of	_	IN	IN	_	34	prep	_	_
36	a	_	DT	DT	_	37	det	_	_
37	world	_	NN	NN	_	35	pobj	_	_
38	no	_	DT	DT	_	39	det	_	_
39	one	_	NN	NN	_	43	nsubj	_	_
40	else	_	RB	RB	_	39	advmod	_	_
41	has	_	VBZ	VBZ	_	43	aux	_	_
42	ever	_	RB	RB	_	43	advmod	_	_
43	seen	_	VBN	VBN	_	37	rcmod	_	_

Table 40: Gold standards in the CoNLL format for the 20th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Skill	_	NNP	NNP	_	5	nsubj	_	_
2	without	_	IN	IN	_	3	prep	_	_
3	imagination	_	NN	NN	_	1	nmod	_	_
4	is	_	VBZ	VBZ	_	5	cop	_	_
5	craftsmanship	_	JJ	JJ	_	0	root	_	_
6	and	_	CC	CC	_	7	cc	_	_
7	gives	_	VBZ	VBZ	_	5	conj	_	_
8	us	_	PRP	PRP	_	7	iobj	_	_
9	many	_	JJ	JJ	_	11	amod	_	_
10	useful	_	JJ	JJ	_	11	amod	_	_
11	objects	_	NNS	NNS	_	7	obj	_	_
12	such	_	JJ	JJ	_	16	prep	_	_
13	as	_	IN	IN	_	12	fixed	_	_
14	wickerwork	_	NN	NN	_	16	compound	_	_
15	picnic	_	NN	NN	_	16	compound	_	_
16	baskets	_	NNS	NNS	_	11	nmod	_	_
1	Imagination	_	NNP	NNP	_	4	nsubj	_	_
2	without	_	IN	IN	_	3	prep	_	_
3	skill	_	NN	NN	_	1	nmod	_	_
4	gives	_	VBZ	VBZ	_	0	root	_	_
5	us	_	PRP	PRP	_	4	iobj	_	_
6	modern	_	JJ	JJ	_	7	amod	_	_
7	art	_	NN	NN	_	4	obj	_	_

Table 41: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 21th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Skill	_	NN	NN	_	5	nsubj	_	_
2	without	_	RB	RB	_	1	prep	_	_
3	imagination	_	NN	NN	_	2	pobj	_	_
4	is	_	VBZ	VBZ	_	5	cop	_	_
5	craftsmanship	_	NN	NN	_	0	root	_	_
6	and	_	CC	CC	_	5	cc	_	_
7	gives	_	VBZ	VBZ	_	5	conj	_	_
8	us	_	PRP	PRP	_	7	iobj	_	_
9	many	_	PDT	PDT	_	11	amod	_	_
10	useful	_	JJ	JJ	_	11	amod	_	_
11	objects	_	NNS	NNS	_	7	obj	_	_
12	such	_	IN	IN	_	13	fixed	_	_
13	as	_	IN	IN	_	11	prep	_	_
14	wickerwork	_	NN	NN	_	16	compound	_	_
15	picnic	_	NN	NN	_	16	compound	_	_
16	baskets	_	NNS	NNS	_	13	pobj	_	_
1	Imagination	_	NN	NN	_	4	nsubj	_	_
2	without	_	RB	RB	_	1	prep	_	_
3	skill	_	NN	NN	_	2	pobj	_	_
4	gives	_	VBZ	VBZ	_	0	root	_	_
5	us	_	PRP	PRP	_	4	iobj	_	_
6	modern	_	JJ	JJ	_	7	amod	_	_
7	art	_	NN	NN	_	4	obj	_	_

Table 42: Gold standards in the CoNLL format for the 21st sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	An	_	DT	DT	_	3	det	_	_
2	MoD	_	NNP	NNP	_	3	compound	_	_
3	spokesman	_	NN	NN	_	4	nsubj	_	_
4	said	_	VBD	VBD	_	0	root	_	_
7	Surveys	_	NNS	NNS	_	13	nsubj:pass	_	_
8	of	_	IN	IN	_	9	prep	_	_
9	Astute	_	NNP	NNP	_	7	nmod	_	_
10	have	_	VBP	VBP	_	13	aux	_	_
11	now	_	RB	RB	_	13	advmod	_	_
12	been	_	VBN	VBN	_	13	aux:pass	_	_
13	completed	_	VBN	VBN	_	4	ccomp	_	_
14	and	_	CC	CC	_	17	cc	_	_
15	she	_	PRP	PRP	_	17	nsubj	_	_
16	will	_	MD	MD	_	17	aux	_	_
17	proceed	_	VB	VB	_	13	conj	_	_
18	to	_	IN	IN	_	19	prep	_	_
19	Faslane	_	NNP	NNP	_	17	pobj	_	_
20	under	_	IN	IN	_	23	prep	_	_
21	her	_	PRP$	PRP$	_	23	nmod:poss	_	_
22	own	_	JJ	JJ	_	23	amod	_	_
23	power	_	NN	NN	_	17	pobj	_	_
1	She	_	PRP	PRP	_	4	nsubj:pass	_	_
2	is	_	VBZ	VBZ	_	4	aux	_	_
3	being	_	VBG	VBG	_	4	aux:pass	_	_
4	escorted	_	VBN	VBN	_	0	root	_	_
5	by	_	IN	IN	_	6	prep	_	_
6	tugs	_	NNS	NNS	_	4	pobj	_	_
7	and	_	CC	CC	_	9	cc	_	_
8	HMS	_	NNP	NNP	_	9	compound	_	_
9	Shoreham	_	NNP	NNP	_	6	conj	_	_

Table 43: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 22th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	An	_	DT	DT	_	3	det	_	_
2	MoD	_	NNP	NNP	_	3	compound	_	_
3	spokesman	_	NN	NN	_	4	nsubj	_	_
4	said	_	VBD	VBD	_	0	root	_	_
7	Surveys	_	NNS	NNS	_	13	nsubj:pass	_	_
8	of	_	IN	IN	_	7	prep	_	_
9	Astute	_	NNP	NNP	_	8	pobj	_	_
10	have	_	VBP	VBP	_	13	aux	_	_
11	now	_	RB	RB	_	13	advmod	_	_
12	been	_	VBN	VBN	_	13	aux:pass	_	_
13	completed	_	VBN	VBN	_	4	parataxis	_	_
14	and	_	CC	CC	_	13	cc	_	_
15	she	_	PRP	PRP	_	17	nsubj	_	_
16	will	_	MD	MD	_	17	aux	_	_
17	proceed	_	VB	VB	_	13	conj	_	_
18	to	_	IN	IN	_	17	prep	_	_
19	Faslane	_	NNP	NNP	_	18	pobj	_	_
20	under	_	IN	IN	_	17	prep	_	_
21	her	_	PRP$	PRP$	_	23	nmod:poss	_	_
22	own	_	JJ	JJ	_	23	amod	_	_
23	power	_	NN	NN	_	20	pobj	_	_
25	She	_	PRP	PRP	_	28	nsubj:pass	_	_
26	is	_	VBZ	VBZ	_	28	aux	_	_
27	being	_	VBG	VBG	_	28	aux:pass	_	_
28	escorted	_	VBN	VBN	_	13	parataxis	_	_
29	by	_	IN	IN	_	28	prep	_	_
30	tugs	_	NNS	NNS	_	29	pobj	_	_
31	and	_	CC	CC	_	30	cc	_	_
32	HMS	_	NNP	NNP	_	33	compound	_	_
33	Shoreham	_	NNP	NNP	_	30	conj	_	_

Table 44: Gold standards in the CoNLL format for the 22nd sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	But	_	CC	CC	_	6	cc	_	_
2	far	_	RB	RB	_	3	advmod	_	_
3	fewer	_	JJR	JJR	_	4	amod	_	_
4	people	_	NNS	NNS	_	6	nsubj	_	_
5	fully	_	RB	RB	_	6	advmod	_	_
6	understand	_	VBP	VBP	_	0	root	_	_
7	how	_	WRB	WRB	_	11	advmod	_	_
8	the	_	DT	DT	_	10	det	_	_
9	Media	_	NNP	NNP	_	10	compound	_	_
10	Lab	_	NNP	NNP	_	11	nsubj	_	_
11	operates	_	VBZ	VBZ	_	6	ccomp	_	_
13	fits	_	VBZ	VBZ	_	11	conj	_	_
14	into	_	IN	IN	_	15	prep	_	_
15	MIT	_	NNP	NNP	_	13	obj	_	_
17	and	_	CC	CC	_	18	cc	_	_
18	encourages	_	VBZ	VBZ	_	11	conj	_	_
19	such	_	PDT	PDT	_	22	det:predet	_	_
20	a	_	DT	DT	_	22	det	_	_
21	creative	_	JJ	JJ	_	22	amod	_	_
22	environment	_	NN	NN	_	18	obj	_	_
24	about	_	RB	RB	_	25	advmod	_	_
25	half	_	PDT	PDT	_	32	nsubj	_	_
26	of	_	IN	IN	_	31	prep	_	_
27	the	_	DT	DT	_	31	det	_	_
28	anniversary	_	NN	NN	_	31	compound	_	_
29	celebration	_	NN	NN	_	31	compound	_	_
30	’s	_	NN	NN	_	31	compound	_	_
31	program	_	NN	NN	_	25	nmod	_	_
32	focused	_	VBD	VBD	_	6	parataxis	_	_
33	on	_	IN	IN	_	35	mark	_	_
34	simply	_	RB	RB	_	35	advmod	_	_
35	defining	_	VBG	VBG	_	32	advcl	_	_
36	what	_	WP	WP	_	40	obj	_	_
37	the	_	DT	DT	_	39	det	_	_
38	Media	_	NNP	NNP	_	39	compound	_	_
39	Lab	_	NNP	NNP	_	40	nsubj	_	_
40	is	_	VBZ	VBZ	_	35	ccomp	_	_

Table 45: Stanford unlexicalized PCFG parser’s output in the CoNLL format for the 23th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	But	_	CC	CC	_	6	cc	_	_
2	far	_	RB	RB	_	3	advmod	_	_
3	fewer	_	JJR	JJR	_	4	amod	_	_
4	people	_	NNS	NNS	_	6	nsubj	_	_
5	fully	_	RB	RB	_	6	advmod	_	_
6	understand	_	VBP	VBP	_	0	root	_	_
7	how	_	WRB	WRB	_	11	advmod	_	_
8	the	_	DT	DT	_	10	det	_	_
9	Media	_	NNP	NNP	_	10	compound	_	_
10	Lab	_	NNP	NNP	_	11	nsubj	_	_
11	operates	_	VBZ	VBZ	_	6	ccomp	_	_
13	fits	_	VBZ	VBZ	_	11	conj	_	_
14	into	_	IN	IN	_	13	prep	_	_
15	MIT	_	NNP	NNP	_	14	pobj	_	_
17	and	_	CC	CC	_	11	cc	_	_
18	encourages	_	VBZ	VBZ	_	11	conj	_	_
19	such	_	PDT	PDT	_	22	det:predet	_	_
20	a	_	DT	DT	_	22	det	_	_
21	creative	_	JJ	JJ	_	22	amod	_	_
22	environment	_	NN	NN	_	18	obj	_	_
24	about	_	RB	RB	_	25	quantmod	_	_
25	half	_	PDT	PDT	_	31	nsubj	_	_
26	of	_	IN	IN	_	25	prep	_	_
27	the	_	DT	DT	_	29	det	_	_
28	anniversary	_	NN	NN	_	29	compound	_	_
29	celebration	_	NN	NN	_	31	nmod:poss	_	_
30	’s	_	POS	POS	_	29	possessive	_	_
31	program	_	NN	NN	_	25	pobj	_	_
32	focused	_	VBD	VBD	_	6	parataxis	_	_
33	on	_	IN	IN	_	32	prep	_	_
34	simply	_	RB	RB	_	35	advmod	_	_
35	defining	_	VBG	VBG	_	33	pcomp	_	_
36	what	_	WP	WP	_	40	obj	_	_
37	the	_	DT	DT	_	39	det	_	_
38	Media	_	NNP	NNP	_	39	compound	_	_
39	Lab	_	NNP	NNP	_	40	nsubj	_	_
40	is	_	VBZ	VBZ	_	35	ccomp	_	_

Table 46: Gold standards in the CoNLL format for the 23rd sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Instead	_	RB	RB	_	4	mark	_	_
2	of	_	IN	IN	_	1	fixed	_	_
3	constantly	_	RB	RB	_	4	advmod	_	_
4	worrying	_	VBG	VBG	_	13	advcl	_	_
5	about	_	IN	IN	_	6	prep	_	_
6	funding	_	NN	NN	_	4	pobj	_	_
8	the	_	DT	DT	_	9	det	_	_
9	faculty	_	NN	NN	_	13	nsubj	_	_
10	and	_	CC	CC	_	11	cc	_	_
11	students	_	NNS	NNS	_	9	conj	_	_
12	can	_	MD	MD	_	13	aux	_	_
13	focus	_	VB	VB	_	0	root	_	_
14	on	_	IN	IN	_	16	prep	_	_
15	their	_	PRP$	PRP$	_	16	nmod:poss	_	_
16	project	_	NN	NN	_	13	pobj	_	_
18	with	_	IN	IN	_	20	prep	_	_
19	the	_	DT	DT	_	20	det	_	_
20	exception	_	NN	NN	_	13	pobj	_	_
21	of	_	IN	IN	_	22	prep	_	_
22	sponsors	_	NNS	NNS	_	20	nmod	_	_
24	weeks	_	NNS	NNS	_	22	dep	_	_
26	when	_	WRB	WRB	_	28	advmod	_	_
27	they	_	PRP	PRP	_	28	nsubj	_	_
28	have	_	VBP	VBP	_	22	rcmod	_	_
29	to	_	TO	TO	_	30	mark	_	_
30	convince	_	VB	VB	_	28	xcomp	_	_
31	companies	_	NNS	NNS	_	30	obj	_	_
32	to	_	TO	TO	_	33	mark	_	_
33	start	_	VB	VB	_	30	xcomp	_	_
34	or	_	CC	CC	_	35	cc	_	_
35	continue	_	VB	VB	_	33	conj	_	_
36	their	_	PRP$	PRP$	_	37	nmod:poss	_	_
37	support	_	NN	NN	_	33	obj	_	_

Table 47: Stanford unlexicalized PCFG parser's output in the CoNLL format for the 24th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	Instead	_	RB	RB	_	2	fixed	_	_
2	of	_	IN	IN	_	1	prep	_	_
3	constantly	_	RB	RB	_	4	advmod	_	_
4	worrying	_	VBG	VBG	_	2	pcomp	_	_
5	about	_	RB	RB	_	4	prep	_	_
6	funding	_	NN	NN	_	5	pobj	_	_
8	the	_	DT	DT	_	9	det	_	_
9	faculty	_	NN	NN	_	13	nsubj	_	_
10	and	_	CC	CC	_	9	cc	_	_
11	students	_	NNS	NNS	_	9	conj	_	_
12	can	_	MD	MD	_	13	aux	_	_
13	focus	_	VB	VB	_	0	root	_	_
14	on	_	IN	IN	_	13	prep	_	_
15	their	_	PRP$	PRP$	_	16	nmod:poss	_	_
16	project	_	NN	NN	_	14	pobj	_	_
18	with	_	IN	IN	_	13	prep	_	_
19	the	_	DT	DT	_	20	det	_	_
20	exception	_	NN	NN	_	18	pobj	_	_
21	of	_	IN	IN	_	20	prep	_	_
22	sponsors	_	NNS	NNS	_	24	nmod:poss	_	_
24	weeks	_	NNS	NNS	_	21	pobj	_	_
26	when	_	WP	WP	_	28	mark	_	_
27	they	_	PRP	PRP	_	28	nsubj	_	_
28	have	_	VBP	VBP	_	24	rcmod	_	_
29	to	_	TO	TO	_	30	aux	_	_
30	convince	_	VB	VB	_	28	xcomp	_	_
31	companies	_	NNS	NNS	_	30	obj	_	_
32	to	_	TO	TO	_	33	aux	_	_
33	start	_	VB	VB	_	30	xcomp	_	_
34	or	_	CC	CC	_	33	cc	_	_
35	continue	_	VB	VB	_	33	conj	_	_
36	their	_	PRP$	PRP$	_	37	nmod:poss	_	_
37	support	_	NN	NN	_	33	obj	_	_

Table 48: Gold standards in the CoNLL format for the 24th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	2	det	_	_
2	doctors	_	NNS	NNS	_	4	nsubj	_	_
3	are	_	VBP	VBP	_	4	aux	_	_
4	warning	_	VBG	VBG	_	0	root	_	_
5	that	_	IN	IN	_	10	mark	_	_
6	the	_	DT	DT	_	7	det	_	_
7	NHS	_	NN	NN	_	10	nsubj	_	_
8	can	_	MD	MD	_	10	aux	_	_
9	not	_	RB	RB	_	10	advmod	_	_
10	make	_	VB	VB	_	4	ccomp	_	_
11	the	_	DT	DT	_	12	det	_	_
12	£	_	$	$	_	10	obj	_	_
13	20bn	_	CD	CD	_	12	nummod	_	_
14	of	_	IN	IN	_	15	prep	_	_
15	savings	_	NNS	NNS	_	12	nmod	_	_
16	by	_	IN	IN	_	17	prep	_	_
17	2014	_	CD	CD	_	10	pobj	_	_
18	that	_	IN	IN	_	20	mark	_	_
19	ministers	_	NNS	NNS	_	20	nsubj	_	_
20	expect	_	VBP	VBP	_	10	ccomp	_	_
22	while	_	IN	IN	_	40	mark	_	_
23	simultaneously	_	RB	RB	_	40	advmod	_	_
24	undertaking	_	VBG	VBG	_	40	csubj	_	_
25	a	_	DT	DT	_	27	det	_	_
26	huge	_	JJ	JJ	_	27	amod	_	_
27	reorganisation	_	NN	NN	_	24	obj	_	_
28	that	_	WDT	WDT	_	30	nsubj	_	_
29	will	_	MD	MD	_	30	aux	_	_
30	see	_	VB	VB	_	27	rcmod	_	_
31	England	_	NNP	NNP	_	36	nsubj	_	_
32	’s	_	VB	VB	_	36	cop	_	_
33	152	_	CD	CD	_	36	nummod	_	_
34	primary	_	JJ	JJ	_	36	amod	_	_
35	care	_	NN	NN	_	36	compound	_	_
36	trusts	_	NNS	NNS	_	30	ccomp	_	_
38	PCTs	_	NNP	NNP	_	24	dep	_	_
40	abolished	_	VBD	VBD	_	10	advcl	_	_
41	and	_	CC	CC	_	45	cc	_	_
42	consortiums	_	NNS	NNS	_	45	nsubj	_	_
43	of	_	IN	IN	_	44	prep	_	_
44	GPs	_	NNP	NNP	_	42	nmod	_	_
45	assume	_	VBP	VBP	_	40	conj	_	_
46	responsibility	_	NN	NN	_	45	obj	_	_
47	for	_	IN	IN	_	49	prep	_	_
48	the	_	DT	DT	_	49	det	_	_
49	commissioning	_	NN	NN	_	45	pobj	_	_
50	of	_	IN	IN	_	51	prep	_	_
51	services	_	NNS	NNS	_	49	nmod	_	_
52	for	_	IN	IN	_	53	prep	_	_
53	patients	_	NNS	NNS	_	51	nmod	_	_

Table 49: Stanford unlexicalized PCFG parser’s output in the CoNLL format for the 25th sentence of the handout.

Id	Form	Lemma	Upos	Xpos	Feats	Head	Deprel	Deps	Misc
1	The	_	DT	DT	_	2	det	_	_
2	doctors	_	NNS	NNS	_	4	nsubj	_	_
3	are	_	VBP	VBP	_	4	aux	_	_
4	warning	_	VBG	VBG	_	0	root	_	_
5	that	_	IN	IN	_	10	mark	_	_
6	the	_	DT	DT	_	7	det	_	_
7	NHS	_	NNP	NNP	_	10	nsubj	_	_
8	can	_	MD	MD	_	10	aux	_	_
9	not	_	RB	RB	_	10	neg	_	_
10	make	_	VB	VB	_	4	ccomp	_	_
11	the	_	DT	DT	_	13	det	_	_
12	£	_	$	$	_	10	obj	_	_
13	20bn	_	CD	CD	_	12	nummod	_	_
14	of	_	IN	IN	_	13	prep	_	_
15	savings	_	NNS	NNS	_	14	pobj	_	_
16	by	_	IN	IN	_	10	prep	_	_
17	2014	_	CD	CD	_	16	pobj	_	_
18	that	_	WP	WP	_	20	obj	_	_
19	ministers	_	NNS	NNS	_	20	nsubj	_	_
20	expect	_	VBP	VBP	_	12	rcmod	_	_
22	while	_	RB	RB	_	10	mark	_	_
23	simultaneously	_	RB	RB	_	24	advmod	_	_
24	undertaking	_	VBG	VBG	_	22	xcomp	_	_
25	a	_	DT	DT	_	27	det	_	_
26	huge	_	JJ	JJ	_	27	amod	_	_
27	reorganisation	_	NN	NN	_	24	obj	_	_
28	that	_	WP	WP	_	30	nsubj	_	_
29	will	_	MD	MD	_	30	aux	_	_
30	see	_	VB	VB	_	27	rcmod	_	_
31	England	_	NNP	NNP	_	36	nmod:poss	_	_
32	’s	_	POS	POS	_	31	possessive	_	_
33	152	_	CD	CD	_	36	nummod	_	_
34	primary	_	JJ	JJ	_	35	amod	_	_
35	care	_	NN	NN	_	36	compound	_	_
36	trusts	_	NNS	NNS	_	30	obj	_	_
38	PCTs	_	NNP	NNP	_	36	appos	_	_
40	abolished	_	JJ	JJ	_	30	xcomp	_	_
41	and	_	CC	CC	_	40	cc	_	_
42	consortiums	_	NNS	NNS	_	45	nsubj	_	_
43	of	_	IN	IN	_	42	prep	_	_
44	GPs	_	NNP	NNP	_	43	pobj	_	_
45	assume	_	VBP	VBP	_	40	conj	_	_
46	responsibility	_	NN	NN	_	45	obj	_	_
47	for	_	IN	IN	_	46	prep	_	_
48	the	_	DT	DT	_	49	det	_	_
49	commissioning	_	NN	NN	_	47	pobj	_	_
50	of	_	IN	IN	_	49	prep	_	_
51	services	_	NNS	NNS	_	50	pobj	_	_
52	for	_	IN	IN	_	51	prep	_	_
53	patients	_	NNS	NNS	_	52	pobj	_	_

Table 50: Gold standards in the CoNLL format for the 25th sentence of the handout.