

Qualitative analysis of Transformer-based Parsing Model

This report analyzes syntactic ambiguities in the validation and test sentences provided to the transformer-based parsing model and the decoded CKY output. A sentence may be syntactically ambiguous if it may have more than one possible parse tree. Most of these cases are not too obvious and require a more careful reading as humans are rather good at disambiguation, especially in possibilities that are semantically unreasonable.

Most of the examples where the ambiguity was the upstream tag having noun phrase as a child, and noun phrase and prepositional phrase as grandchildren vs having noun phrase and prepositional phrase as children was resolved in the former manner. This may partly be due to the former pattern being more semantically obvious in the annotated training examples and the lack of commas, which were common in examples where the latter pattern was more explicit. For coordination ambiguity, the model seems to prefer shorter phrases for the conjunction.

Figure 1: The model considers "a group of hundreds of lawyers in Chicago" as the NP, as opposed to just "group of hundreds of lawyers" as the NP and "in Chicago" as a PP at the same level, which seems to be plausible derivation.

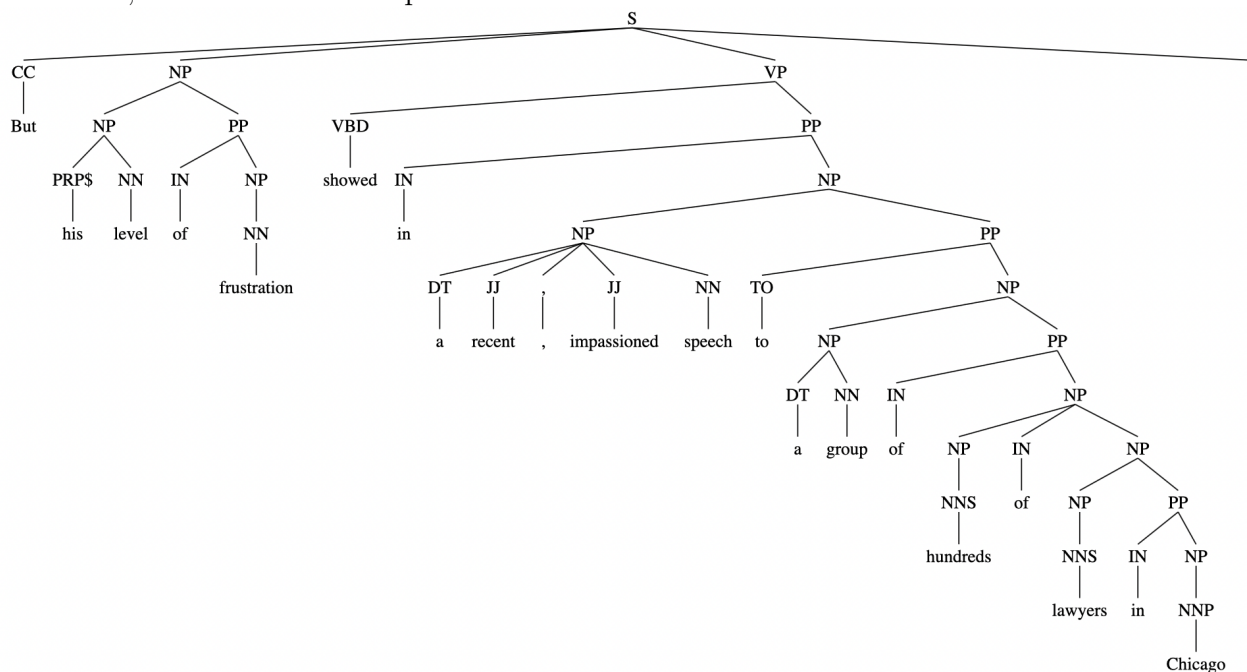


Figure 2: This is a hard to reason with semantically, but "a degree of liquidity that is not there in the market" is tagged as NP as opposed to "in the market" being tagged as a PP at the same level (if the speculators were in the 'market' and calling for liquidity). Perhaps this is partly influenced by the domain of our dataset.

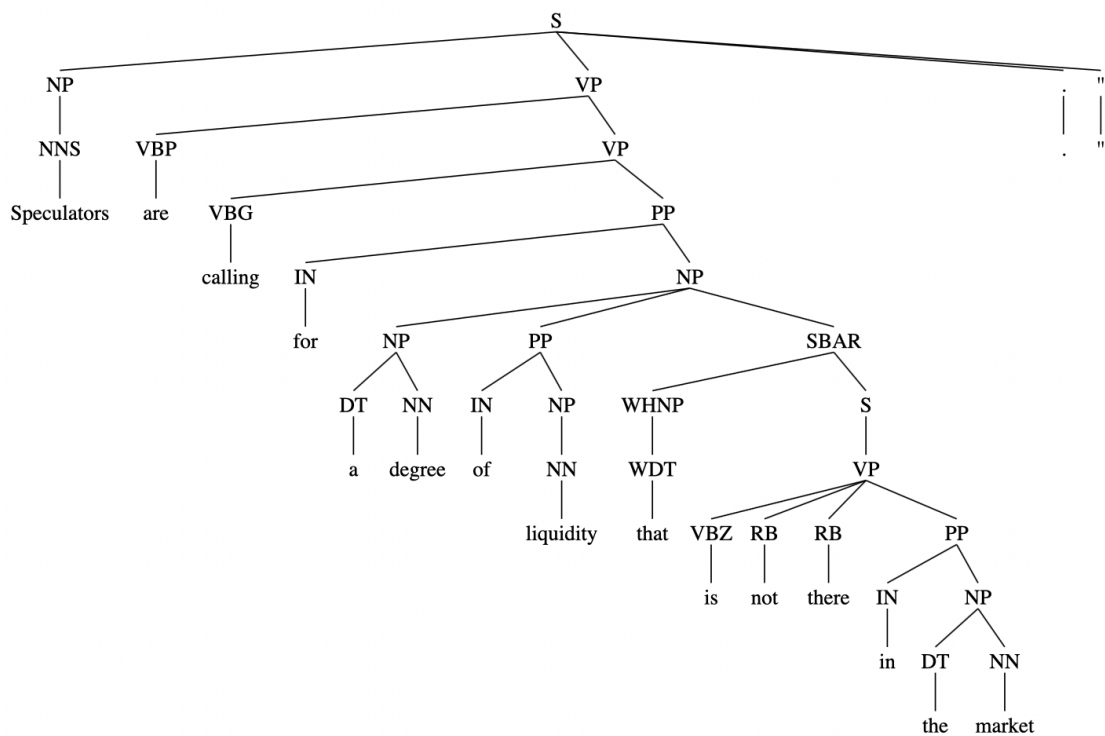


Figure 3: Model tags "IBM's outlook for the next few quarters" as NP as opposed to "IBM's outlook" as NP and rest of the phrase as a sibling PP instead of subtree. Are the analysts downbeat about IBM's outlook (for any subject) for the next few quarters or are they IBM's outlook (for the next few quarters)?

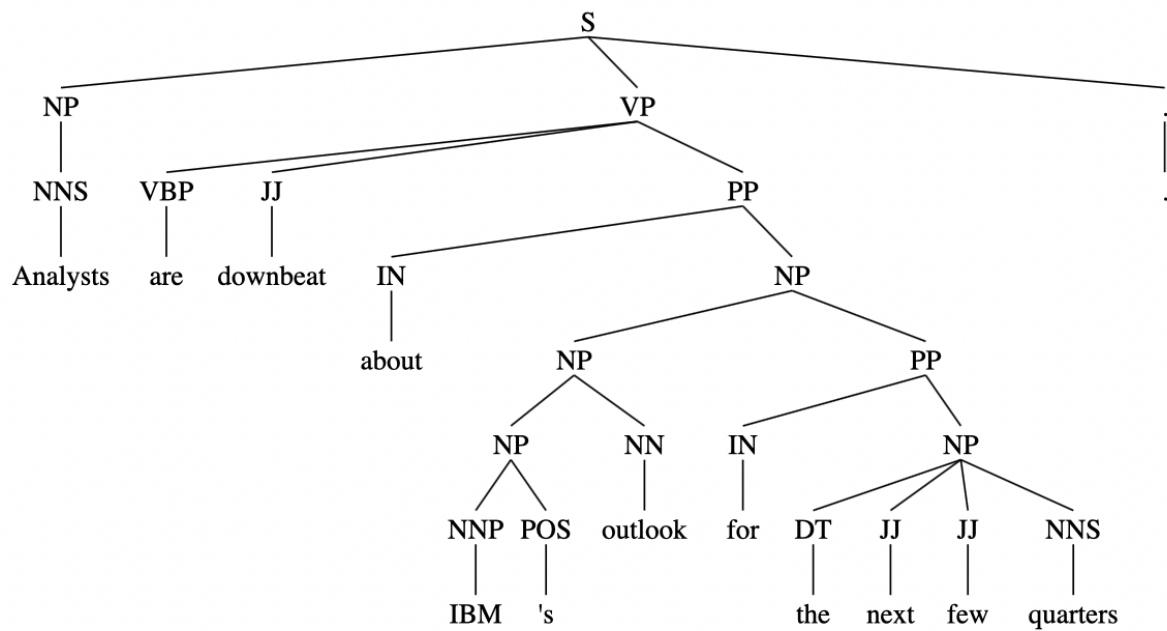
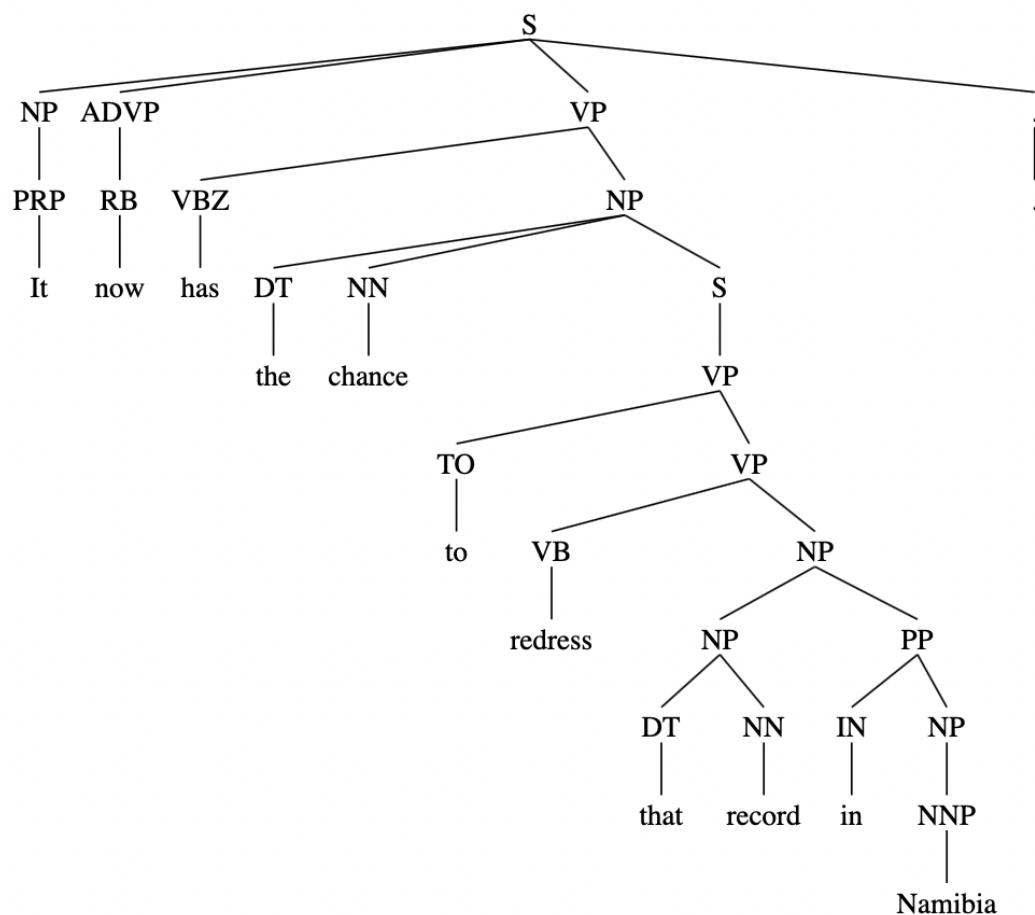


Figure 4: Yet again, our model tags "that record in Namibia" as the NP, instead of "that record" as NP and "in Namibia" as PP, both at the same level after "redress that record in Namibia" VP, which seems equally plausible semantically.



```

graph TD
    S1[S] --- NP1[NP]
    S1 --- VP1[VP]
    NP1 --- PRP[PRP]
    PRP --- They[They]
    VP1 --- VBP[VBP]
    VBP --- hope[hope]
    VP1 --- S2[S]
    S2 --- VP2[VP]
    VP2 --- TO[TO]
    TO --- to[to]
    VP2 --- VP3[VP]
    VP3 --- VB[VB]
    VB --- avoid[avoid]
    VP3 --- NP2[NP]
    NP2 --- NP3[NP]
    NP3 --- DT1[DT]
    DT1 --- the1[the]
    NP3 --- JJ1[JJ]
    JJ1 --- jammed[jammed]
    NP3 --- NN1[NN]
    NN1 --- phone[phone]
    NP3 --- NNS1[NNS]
    NNS1 --- lines[lines]
    NP2 --- CC[CC]
    CC --- and[and]
    NP2 --- NP4[NP]
    NP4 --- JJ2[JJ]
    JJ2 --- other[other]
    NP4 --- NNS2[NNS]
    NNS2 --- snags[snags]
    NP2 --- WHNP[WHNP]
    WHNP --- WDT[WDT]
    WDT --- that[that]
    NP2 --- VP4[VP]
    VP4 --- VBD[VBD]
    VBD --- infuriated[infruriated]
    NP2 --- SBAR[SBAR]
    SBAR --- S3[S]
    S3 --- NP5[NP]
    NP5 --- DT2[DT]
    DT2 --- some[some]
    NP5 --- NN2[NN]
    NN2 --- fund[fund]
    NP5 --- NNS3[NNS]
    NNS3 --- investors[investors]
    S3 --- PP[PP]
    PP --- IN[IN]
    IN --- in[in]
    PP --- NP6[NP]
    NP6 --- NNP[NNP]
    NNP --- October[October]
    NP6 --- CD[CD]
    CD --- 1987[1987]
  
```

The diagram illustrates the syntactic structure of the sentence: "Harry is now a 57-year-old writer, whose continuing flirtation with drugs is marginal types in Hollywood and New York seems quaintly out-of-synch".

The main clause structure is:

- S** (Sentence)
 - NP** (Noun Phrase): Harry
 - VP** (Verb Phrase): is
 - ADV** (Adverb): now
 - NP** (Noun Phrase): a 57-year-old writer
 - SBAR** (Subordinate Clause):
 - WHNP** (Wh-noun phrase): whose continuing flirtation
 - WHNP** (Wh-noun phrase): whose
 - VBG** (Verb-ing): continuing
 - NN** (Noun): flirtation
 - PP** (Prepositional phrase): with drugs is marginal types in Hollywood and New York
 - IN** (Preposition): with
 - NP** (Noun phrase): drugs
 - NNS** (Noun, plural): drugs
 - CC** (Coordinator): and
 - NP** (Noun phrase): marginal types
 - JJ** (Adjective): marginal
 - NNS** (Noun, plural): types
 - IN** (Preposition): in
 - PP** (Prepositional phrase): Hollywood and New York
 - NP** (Noun phrase): Hollywood
 - NNP** (Noun, proper): Hollywood
 - CC** (Coordinator): and
 - NP** (Noun phrase): New York
 - NNP** (Noun, proper): New
 - NNP** (Noun, proper): York
 - S** (Sentence): seems
 - VP** (Verb phrase):
 - VBZ** (Verb, 3rd person singular present tense): seems
 - ADJP** (Adjective phrase):
 - RB** (Adverb): quaintly
 - JJ** (Adjective): out-of-synch

Figure 7: Coordination ambiguity where it could be (Florida and Texas) vs (Florida and Texas Markets)

