Catherine DeJager

Professors Johnson and VandenBosch

English 374

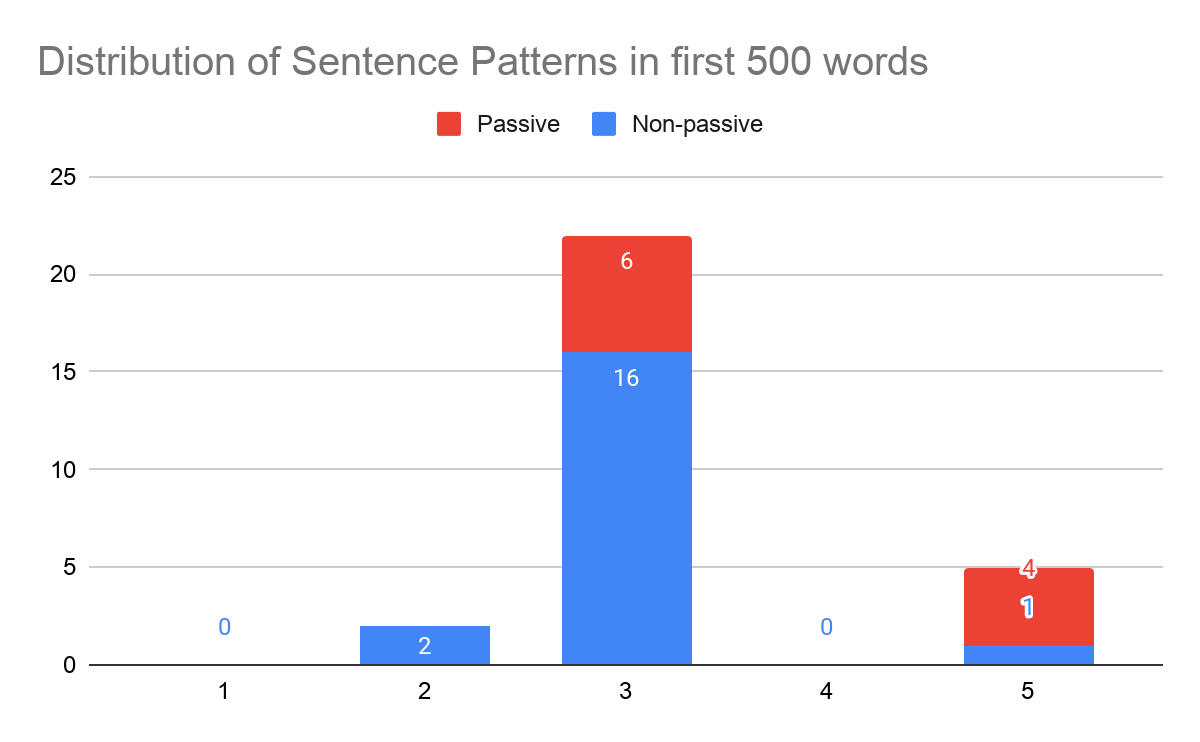
27 January 2020

Grammatical Analysis of “IMPROVING SPOKEN LANGUAGE UNDERSTANDING BY EXPLOITING ASR N-BEST HYPOTHESES”

This article discusses a problem with many spoken language understanding algorithms. The algorithm selects its best automatic speech recognition (ASR) hypothesis (i.e., the algorithm makes its best guess as to what exactly the speaker said), then continues processing based on that assumption. However, if the hypothesis was incorrect, that inaccuracy will be passed down through all the steps of the algorithm, so the result may contain egregious errors. To mitigate this issue, the article proposes the use of ASR *n*-best hypotheses; instead of just looking at the best hypothesis, take the first *n* (where *n* is a number the programmers deem appropriate) hypotheses and do additional work with them. The article describes research in great detail so that other people can apply the approaches the article suggests, so naturally the article has a lot of technical language. I had a little difficulty understanding the article at first because some of the technical details are more advanced than I can understand. In some parts of the article (especially the beginning and the end), the simplicity of the grammar made it relatively easy for me to get the gist of the article even if I didn’t understand all the minute details. However, there are a couple parts, especially section 4.3, that have long sentences with many dependent clauses and can thus be a little difficult to follow.

For the first 500 words, I looked at 2-3.1, not counting figure captions or footnotes. There were approximately 613 words.

Question 1: In the first 500 words, what is the distribution of the five sentence patterns?



Pattern 2 examples:

* “Their abilities to take advantage of multiple interpretations are actually not trained”
* “Each hidden state vector, e.g. *h*1, is the concatenation of forward *h*1f and backward *h*1b states.”

Pattern 3 examples:

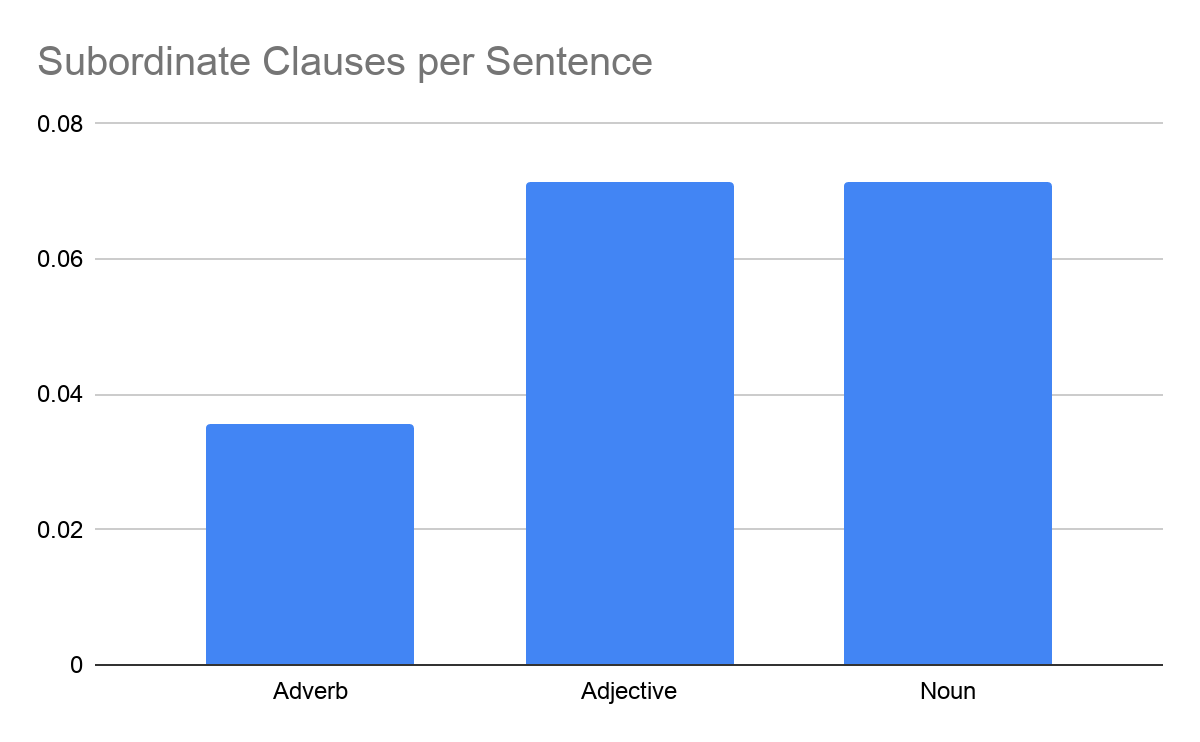
* “The preliminary architecture is shown in Fig. 1.”
* “We separate hypotheses with a special delimiter (<SEP>).”

Pattern 5 examples:

* “We name it Oracle simply because we assume that hypotheses are noisy versions of transcription”
* “The entire model can be formulated as:” (the objective complement is a set of equations)

Question 2: In the first 500 words, what is the average number of subordinate clauses per sentence?

I found 1 adverb clause, 2 adjective clauses, and 1 noun clause. There were 28 sentences total.



Adverb clause examples:

* “We name it Oracle simply **because we assume that hypotheses are noisy versions of transcription**”

Adjective clause examples:

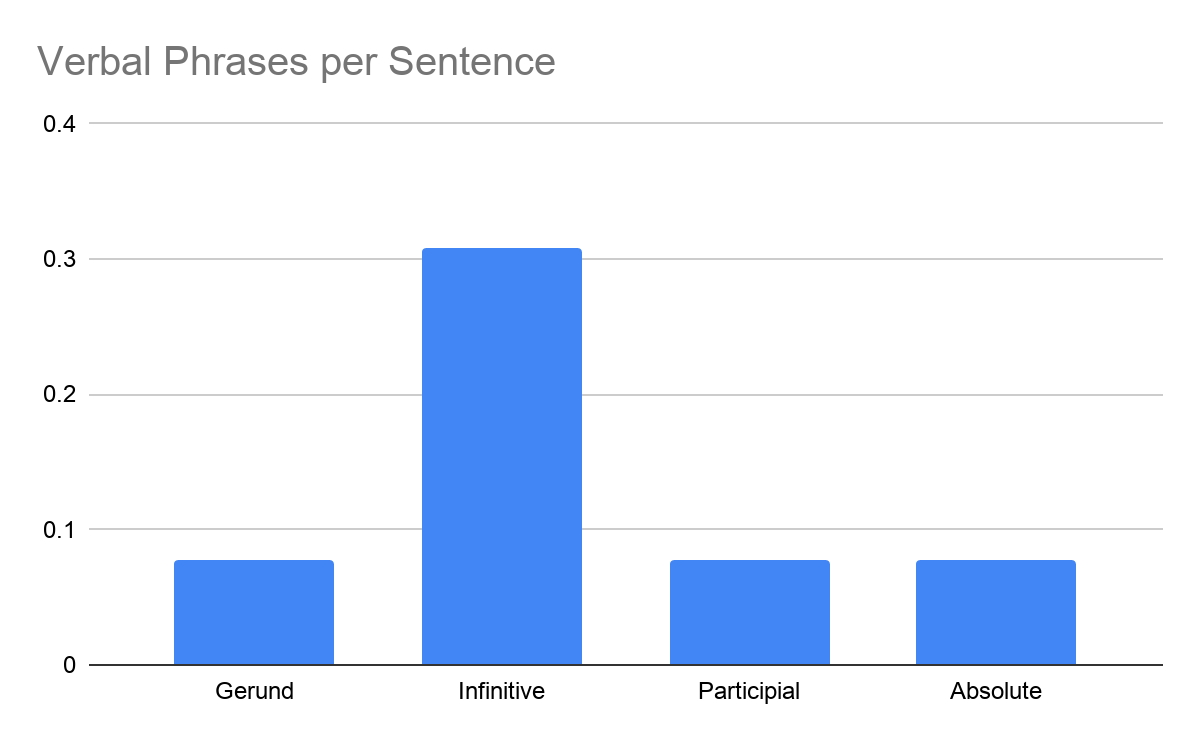
* “Besides the Baseline and Oracle, **where only ASR 1-best hypothesis is considered**, we also perform experiments to utilize ASR *n*-best hypotheses during evaluation.”
* “Since the current rerank models (e.g., [2, 4, 5]) attempt to select the hypothesis most similar to transcription, we propose the Rerank (Oracle), **which picks the hypothesis with the smallest edit distance to transcription (assume it is the *a*-th best) during evaluation and uses its corresponding prediction**.”

Noun clause examples:

* “For a given transcribed utterance, it is firstly encoded with Byte Pair Encoding (BPE) [15], **a compression algorithm splitting words to fundamental subword units (*pairs of bytes* or BPs) and reducing the embedded vocabulary size**.”
* “**The models evaluating with n-bests and a BM (pre-trained on transcription)** are called *Direct Models* (in Fig. 2):”

For the last 250 words, I looked at 4.4 to 5 (conclusions), not counting figure captions or tables. There were approximately 345 words. There were 13 sentences.

Question 3: In the final 250 words, what is the average number of verbal phrases in each sentence?



Gerund phrase examples:

* “This paper improves the SLU system robustness to ASR errors by **integrating *n*-best hypotheses in different ways**, e.g. the aggregation of predictions from hypotheses or the concatenation of hypothesis text or embedding.”

Infinitive phrase examples:

* “In the future, we aim [**to employ additional features (e.g. confidence scores for hypotheses or tokens)** [**to integrate *n*-bests more efficiently**]**, where we can train a function *f*** [**to obtain a weight for each hypothesis embedding before pooling**]]**.**”
* “Another direction is using deep learning framework **to embed the word lattice [17] or confusion network [18, 19]**, which can provide a compact representation of multiple hypotheses and more information like times, in the SLU system.”

Participial phrase examples:

* “**Changing the number of hypotheses while evaluation**, Fig. 5 shows a monotonic increase with the access to more hypotheses for the PoolingAvg and PoolingMax (Sort by Score is shown because it is the best achievable direct model while the Rerank (Oracle) is not realistic).” (I think there may be a typo here (“while evaluation” is incorrect and should be “while evaluating” or “during evaluation”), but either way “changing the number of hypotheses” is the same so this is definitely a participial phrase.)

Absolute phrase examples:

* “We also observe that **with more hypotheses utilized**, the performance can be further improved.”

I was surprised that there were so few dependent clauses and verbal phrases. It was also interesting to see how so many of the sentences were pattern 3 and some sentence patterns were entirely absent. In addition, most sentences were not in the passive voice. This simplicity helped me follow the article without getting lost in the sentences. While some of these observations may be particular to this article, I think the prevalence of pattern 3 sentences and the lack of passive voice are fairly common in computer science. The agent is usually quite important (typically a computer, an algorithm, or a person is doing something), so passive voice is not as useful. There is also usually a direct object. I learned from comparing with my classmates that most people had articles with more complex language (more dependent clauses per sentence, more verbal phrases, etc.). My classmates also tended to find a lot of infinitive phrases in their articles.