44 to 45

An analysis of structural complexity as a linguistic feature of US Presidents Obama and Trump.

Conor Evans

24th April 2020

LI-4032

Contents

1 Introduction

In this section we will introduce structural complexity. We will then outline three software tools that will allow us to conduct structural complexity analysis.

1.1 Structural Complexity

Structural complexity and processes by which it can be measured have been the subject of much research, particularly in the area of psycholinguistics. Structural complexity has been considered particularly finely with respect to Child Language Acquisition (Chomsky, 1969; Olson, 1973) and Second-Language Acquisition (Housen & Kuiken, 2009). Nonetheless, native adult speakers have also been the subject of many analyses, e.g. Kemper et al. (1989). It has proven difficult to reach a consensus on metrics by which structural complexity can be measured. The aforementioned Kemper et al. analysis was an early proponent of Mean Length of Utterance (MLU) as a measure of structural complexity in oral language. This term is closely mirrored by Mean Length of Clause (MLC) or Mean Length of Sentence (MLS) in written language analyses, e.g. Rimmer (2006). Rimmer, however, considers it “doubtful [that] there is a straight correspondence between length of sentence and complexity” and that “unit length is unlikely to be a strong indicator of complexity” (idem, p. 506). This analysis will consider unit length as a potential metric but will endeavour to take a holistic approach and to incorporate a variety of metrics.

Sentence vs. T-unit Given a text, any layperson can identify a sentence. It is demarcated on both sides by a full-stop. It may seem logical to consider that the length of a sentence is a fine measure of syntactic complexity. This is somewhat true and MLU/MLC/MLS have been used in many analyses, e.g. Kemper (1987). Sentence length is a non-decreasing measure of complexity in that a sentence can only become more complex if additional elements are added to the sentence. However, that does not mean that a sentence must be long to be complex. Consider the following two sentences:

1. I know [[why I was expecting [to see something else]].[1]

2. I like [bananas and oranges and grapes and pineapples and apples].

It does not take a linguist to determine that (1) is more complex than (2), yet both sentences are of equal length. The T-unit provides a better measure of

[1]Example taken and slightly modified from Biber et. al (2011, p. 7).

determining structural complexity as it analyzes the sentence with respect to its components rather than its length. Hunt (1970, p. 4) defines the T-unit as "one main clause plus any sub-ordinate clause or non-clausal structure that is attached to or embedded in it". Through conjunction, a sentence can have several T-units, i.e. several main clauses.

Complex T-unit Casanave (1994) proposed that a T-unit can be considered complex when it contains a dependent clause. Relative clauses are embedded subclauses with a filler-gap dependency. The subclause is missing a constituent – either its subject or an object (De Villiers et al., 1979, p. 500). The subclause may have its own embedded subclause and, in theory, this may continue ad infinitum:

3. The man_i [that hated_i the woman_j [that loved_j ...]] was_i enraged.

Both relative clauses in (3) are subject relative. This means that the relative pronoun occupies the role of subject in the embedded clause – this role is known as its *focus*. Subject relatives are centre-embedded in English (idem). Centre-embedded clauses have been considered to be more complex than right-embedded (object) relative clauses since seminal work in the area during the early sixties – (Yngve, 1960; Miller, 1962). This generality spans contexts and groups, from comprehension of relative clauses by native children (Kidd & Bavin, 2002, p. 608) to production of relative clauses by adult L2 speakers (Izumi, 2003, p. 304).

Dangling Preposition Prepositional noun phrases (PNP) are a typical object clause in English – give *to the man*, read *from the page*, etc. The preposition is referred to as 'dangling' or 'hanging' when the PNP has been split. This primarily occurs when a relative clause fronts the NP of a PNP but not its preposition, i.e. when the focus of the relative clause is the object:

4. The man_i [that/who/whom I_j was talking_j to $[]_i$] was_i enraged.

5. The bag_i [that/which I_j was referring_j to $[]_i$] was_i red.

Certain grammarians such as the American Heritage Dictionary (AHA) do not consider these constructions to be proper and judge this filler-gap dependency to be ungrammatical (Kunsmann, Gordesch, & Dretzke, 1998, p. 214–215). The 'proper' constructions would be as follow:

6. The man_i [to *that/*who/whom I_j was talking_j] was_i enraged.

7. The bag_i [to *that/which I_j was referring_j] was_i red.

The prepositional phrase is no longer disjointed. We may note that this form is more restrictive – *that* can no longer be used as the relative pronoun in either clause. Furthermore, we may note the case marking of the relative pronoun. Both nominative *who* and objective *whom* are acceptable in (4), whereas only the latter can be used in (6). These restrictions are the logic behind considering dangling prepositions with respect to structural complexity. This approach has not been explored extensively but Demirezen (2012) did find Turkish L2 university students to have difficulty understanding and producing sentences with dangling elements. Moreover, dangling constructions are "typical of oral language" (Dorgeloh & Wanner, 2010, p. 114). It will thus be interesting to examine whether the structures of speeches by Presidents Obama and Trump corroborate this generality or whether they reflect the more 'proper' and restricted AHA structure.

1.2 Structural Complexity Analysis

L2SC Analyzer The L2SC Analyzer (L2SCA) was developed to analyze structural complexity in L2 speaker data. However, it can still be used in analyzing native speaker data. Like other structural complexity analyzers (SCAs) such as the D-Level Analyzer,[2] the L2SCA preprocessing step uses the Stanford Parser for sentence segmentation, tokenization, and POS tagging. This parser in turn uses the Penn Treebank (Marcus et al., 1993) which was trained on native speaker data. The clause trees returned by the Stanford Parser are analyzed using Tregex (Levy & Andrew, 2006). Tregex allows L2SCA to count the number of occurrences of nine production units and syntactic structures: words, sentences, clauses, dependent clauses, T-units, complex T-units, coordinate phrase, complex nominals, and verb phrases. The Tregex patterns describing each of these are provided by Lu (2010) in his description of the system. L2SCA computes fourteen metrics that can be divided into five subcategories: length of production unit, sentence complexity, subordination, coordination, and particular structures (idem, p. 479). We have stated previously that some consider unit length to be an unreliable indicator of complexity and examples (1) and (2) corroborated this statement. Below are the metrics that we will consider:

[2]Another SCA that assigns texts to eight increasingly complex developmental levels (Lu, 2010).

Measure	Definition
Sentence complexity ratio (SCR)	# clauses / # sentences
Complex T-unit ratio (CTR)	# complex T-units / # T-units
Dependent clauses per T-unit (DCT)	# dependent clauses / # T-units

Table 1: Complexity measures calculated by L2SCA (Lu, 2010, p. 479)

SCR will allow us to better define the structural complexity of sentences such as (1) and (2) where length is not an excellent indicator of complexity. The latter two will allow us to consider the use of dependent clauses such as relative clauses.

Fry Graph The Fry Graph is commonly used to measure readability. It determines the 'grade-reading level' of a text with respect to the US educational system.[3] It is calculated based on two metrics: # sentences / 100 words, # syllables / 100 words. The former is much the same, albeit slightly more nuanced, than L2SCA's Mean Length of Sentence. The latter is moreso a measure of morphemic complexity than structural complexity. Nonetheless, the Fry Graph has been used to increase the structural complexity of texts (Hague & Mason, 1986). It also provides a more tangible metric than those we have introduced thus far.

Coh-Metrix Analyzer The Coh-Metrix Analyzer (CMA) (Graesser et al., 2004) provides a broad range of text analytics and is commonly used to evaluate the 'developmental level' at which L2 speakers can produce and understand English (Vyatkina, 2013; Crossley & McNamara, 2014). Speakers are expected to both produce and understand increasingly complex structures at higher developmental levels. CMA's structural analysis is based on the ApplePie Parser (Sekine & Grishman, 1996) and uses Brill's POS tagger (1995) to determine syntactic tree structure. Left embeddedness (LE) measures the mean number of words before the main verb in each sentence. This may be particularly effective in evaluating relative clause usage. The relative clauses in (3), (4), and (5) each had between six and eight words before the main clause, whereas the simple sentence in (2) had just one. Modifiers per noun phrase (MNP) described the number of elements, e.g. adjectives, that modify the "head [noun] of the phrase" (Graesser et al., 2004, p. 198).

[3] Note that the scale has an upper limit of grade 15 whereas secondary education, the last level to commonly refer to years of education as 'grades', ends at grade 12. Grades 13-15 are based on college levels (Fry, 1977).

Finally, both Adjacent Sentence Syntax Similarity (ASSS) and Parapraph Sentence Syntax Similarity (PSSS) measure the structural coherence of a text. We may wish to consider these metrics with respect to our Fry Graph evaluation. In fact, CMA was first used as a predictor of text readability (McNamara, Louwerse, & Graesser, 2002).

2 Methodology

Null Hypothesis Before we consider what statistics we may use to evaluate our data, we must consider our expectations for the data. The two subjects are assumed to be independent or *unpaired*. We also assume the following null hypothesis:
« The structural complexity of speeches by Presidents Obama and Trump will not differ significantly with respect to the metrics outlined hitherto. »

Corpus AmericanRhetoric[4] is a website that archives many speeches given by influential US figures, including Presidents Obama and Trump. The transcriptions are authenticated, but a customary partial verification was nonetheless carried out personally in order to ensure their authenticity. Chosen source materials were mirrored to best ensure the integrity of the comparison. The corpus is comprised of the first three State of the Union[5] (SOTU) addresses delivered by each President. Links to source materials can be found in the appendices. The corpus for each speaker includes over 1100 sentences:

	No. Words	No. Sentences
Obama		
2010 SOTU	7523	430
2011 SOTU	7263	409
2012 SOTU	7476	420
Trump		
2018 SOTU	5985	376
2019 SOTU	5672	366
2020 SOTU	6482	392

The analyzers have varying upper limits of words that can be processed within a single request. As such, the speeches have been analyzed in parts of roughly equal length. This is perhaps beneficial as it will give us more data

[4] `https://www.americanrhetoric.com/`

[5] `https://history.house.gov/Institution/SOTU/State-of-the-Union/`

points to consider when examining the distribution and the normality of the data. This is a prerequisite to deciding on any statistical test. The below table represents the Shapiro-Wilk p-value. A value above 0.05 implies that the distribution of data is not significantly different from normal distribution and therefore that we can assume normality.

	Obama	**Trump**
L2SCA		
SCR	.9634*	.2465*
CTR	.9727*	.02789
DCT	.8917*	.05292*
Fry Graph		
# syllables / 100 words	.9491*	.8605*
# sentences / 100 words	.1644*	.5016*
CMA		
LE	.4761*	.04694
MNP	.9886*	.3545*
ASSS	.414*	.5013*
PSSS	.9829*	.8047*

Any measure for which data normality can be assumed will be evaluated using an unpaired two-tailed t-test. Any measure for which data normality cannot be assumed will be evaluated using an unpaired two-tailed Wilcoxon test.

Dangling Preposition As this phenomenon is not accounted for by the analyzers described hitherto, it is necessary to implement a custom procedure. The initial implementation is basic and flags prepositions that are at sentence-end or at clause-end before coordination. Similarly, fronted PNPs were flagged. A manual verification was then carried out to see which cases were relevant with respect to the dangling preposition. As this was an experimental implementation that required manual verification, each President's acceptance speeches (94-161 sentences) were considered.

3 Results

Fry Graph The two speakers exhibited significant differences in # sentences / 100 words ($p < 0.01$) and # syllables / 100 words ($p < .025$), the former of which is a more typical measure of structural complexity:

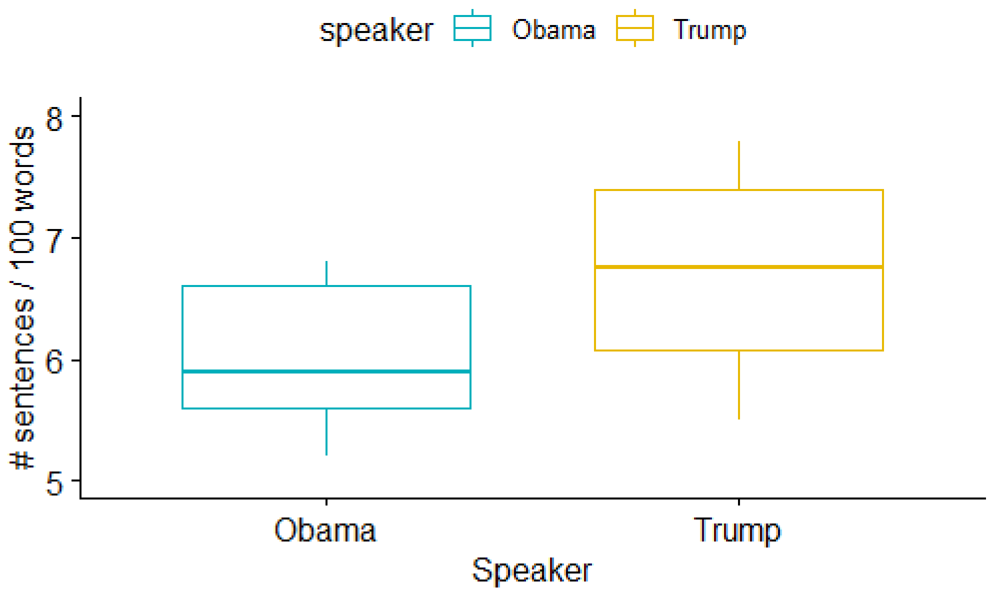

Figure 1: Box plot for the number of sentences per 100 words

Figure 1 shows that Obama uses significantly longer sentences. We used examples (1) and (2) to show that sentence length is not always a reliable indicator of structural complexity. However, we would expect such counter-examples to become less relevant over an extended corpus. Interestingly, the two speakers place relatively similarly on the Fry Graph:

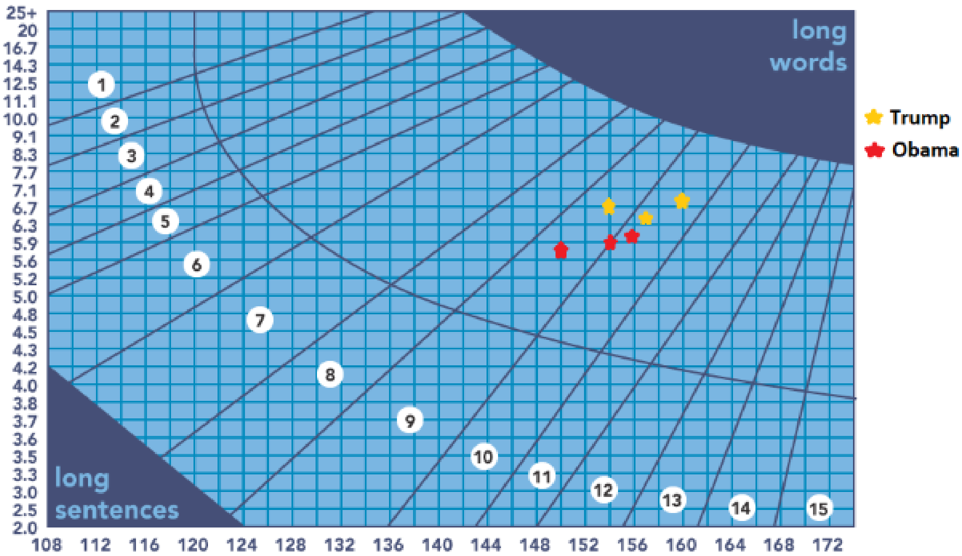

Figure 2: Grade-reading-level according to the Fry Graph.

Each SOTU address is readable at a mid ninth-grade to mid tenth-grade level. Although Obama uses significantly fewer sentences per hundred words, Trump consistently reads at a higher level than Obama and is the closer of the two subjects to reading at eleventh-grade level (16-17 years old).

L2SCA and CMA Obama produces a significantly higher Sentence Complexity Ratio (SCR) than Trump ($p < 0.0001$). Obama also produces a significantly higher Complex T-unit ratio and uses significantly more dependent clauses per T-unit (both $p < 0.0001$). Trump uses significantly more modifiers per noun phrase ($p < 0.01$). The difference in both Adjacent Sentence Syntactic Structure and Paragraph Sentence Syntactic Structure was insignificant ($p = .6612$, $p = .0868$ respectively). This corroborates the similar Fry Graph readability scores. Finally, Obama produces structures with significantly higher left-embeddedness ($p < .025$).

	Obama		**Trump**	
Measurement	Mean	σ^2	Mean	σ^2
# syllables / # words	153.0769	20.5325	157.7143	26.2041
# sentences / # words	5.9538	0.2963	6.6615	0.5839
SCR	1.9022	0.0125	1.5057	0.0174
CTR	0.4900	0.0027	0.3030	0.0027
DCT	0.7658	0.0143	0.3924	0.0073
LE	3.9029	0.2251	3.4115	0.3035
MNP	0.7695	0.0035	0.8799	0.0071
ASSS	0.1208	0.0004	0.1252	0.0006
PSSS	0.0882	0.0002	0.0992	0.0018

Dangling Preposition No examples were found in President Trump's acceptance speech, whereas President Obama used one such construct in his first acceptance speech:[6]

6. I will never forget who this victory truly belongs to.

7. I will never forget to whom this victory truly belongs.

(6) is as realized by Obama, whereas (7) is the proper construct per the American Heritage Dictionary. We noted in the previous section that this

[6] His 2008 acceptance speech was chosen as Trump has only given one acceptance speech.

was a rather experimental implementation. Evidently, one example is simply not sufficient as a sample size and a far greater dataset would need to be considered in order to use this as a measurement of structural complexity. Furthermore, a specific metric would need to be chosen and justified. We could, for example, consider # dangling prepositions / # PNPs. It would likely be more representative to consider only those cases that could be fronted: # dangling prepositions / # frontable PNPs. Finally, the implementation would need to be generalized, as dangling prepositions are not limited to only clause-end or sentence-end. These formalizations would need to be made before any meaningful results could be obtained.

4 Conclusion

Obama produced structures of significantly higher complexity than Trump with respect to all but one metric computed by the SCAs, ergo we may reject our null hypothesis. Given that several p-values were unable to be represented as a decimal,[7] we can state this to a high degree of confidence. Interestingly, however, both speakers placed similarly on the Fry readability graph and, in fact, Trump consistently read at a slightly higher intra-grade level. This was corroborated by the CMA, for which only the readability metrics were not significantly different.[8] Indeed, Rimmer warns of "danger in the over-emphasis of quantitative data in measuring complexity" (2006, p. 507). Though Obama produces structures of a higher complexity, this may not imply that his speech is more complex in qualitative terms. Trump is well known for his highly idiosyncratic speech patterns, from his use of epiphora to clausal disjunction (Sclafani, 2017, p. 21). University of Edinburgh syntactician Geoffrey Pullum has described Trump's rhetoric as lacking "any structure".[9] It is possible that the SCAs used in this analysis are simply unable to parse such disjointed structures and that this may explain the contrast between the significant difference in the quantitative metrics computed by the SCAs and the indistinguishable readability scores.

Although the Complex T-unit ratio (CTR) proved to significantly differentiate the syntactic structures of the two subjects, it is not wholly indicative of relative clause usage. Relative clauses are one of just several dependent clauses that may influence the CTR. Although visual parsing of relative clauses is a frequent study of research (Spivey-Knowlton et al., 1993; Traxler et al., 2002; Carreiras et al., 2004), no SCA presents a formal implemen-

[7] The p-values were returned using e^{-7} or e^{-8}.

[8] See PSSS/ASSS in Sections 1.2 and 3.

[9] `https://languagelog.ldc.upenn.edu/nll/?p=20490#more-20490`

tation of relative clause parsing. This is regrettable as the CTR does not distinguish sub-features of dependent clauses. The left-embeddedness (LE) computed by the CMA is perhaps a better indicator of relative clause usage, though once more it is not wholly indicative given that left-embedded clauses may adjuncts or appositional clauses. Obama did produce structure with significantly higher left-embeddedness. Although a doubly centre-embedded relative clause is considered far more complex than a right-embedded relative clause (Hakes et al., 1976, p. 283), neither LE nor the CTR distinguish between these two dependent clauses in their evaluation.
The dangling preposition was chosen as an experimental metric of syntactic complexity. The data was quite limited and we outlined steps that would need to be taken in order to formalize and generalize the analysis. It is perhaps too specific a metric, too limited in scope. Nonetheless it provided a novel approach to the evaluation of structural complexity and it would be interesting to consider the metric as part of the overall evaluation.

5 Appendices

Obama 2010: https://www.americanrhetoric.com/speeches/PDFFiles/Barack%20Obama%20-%20State%20of%20the%20Union%202010.pdf

Obama 2011: https://www.americanrhetoric.com/speeches/PDFFiles/Barack%20Obama%20-%20State%20of%20the%20Union%202011.pdf

Obama 2012: https://www.americanrhetoric.com/speeches/PDFFiles/Barack%20Obama%20-%20State%20of%20the%20Union%202012.pdf

Trump 2018: https://www.americanrhetoric.com/speeches/stateoftheunion2018.htm

Trump 2019: https://www.americanrhetoric.com/speeches/stateoftheunion2019.htm

Trump 2020: https://www.americanrhetoric.com/speeches/stateoftheunion2020.htm

References

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in l2 writing development? *Tesol Quarterly*, *45*(1), 5–35.

Brill, E. (1995). Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational linguistics*, *21*(4), 543–565.

Carreiras, M., Salillas, E., & Barber, H. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in spanish. *Cognitive Brain Research*, *20*(1), 98–105.

Casanave, C. P. (1994). Language development in students' journals. *Journal of second language writing*, *3*(3), 179–201.

Chomsky, C. (1969). The acquisition of syntax in children from 5 to 10.

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? a computational investigation of syntactic complexity in l2 learners. *Journal of Second Language Writing*, *26*, 66–79.

Demirezen, M. (2012). An analysis of the problem-causing structures of simple sentences for turkish university students. *International journal of humanities and social science*, *2*(3), 135–146.

De Villiers, J. G., Flusberg, H. B. T., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of psycholinguistic Research*, *8*(5), 499–518.

Dorgeloh, H., & Wanner, A. (2010). *Syntactic variation and genre* (Vol. 70). Walter de Gruyter.

Fry, E. B. (1977). *Elementary reading instruction.* McGraw-Hill.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, *36*(2), 193–202.

Hague, S. A., & Mason, G. E. (1986). Using the computer's readability measure to teach students to revise their writing. *Journal of Reading*, *30*(1), 14–17.

Hakes, D. T., Evans, J. S., & Brannon, L. L. (1976). Understanding sentences with relative clauses. *Memory & Cognition*, *4*(3), 283–290.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*, *30*(4), 461–473.

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the society for research in child development*, *35*(1), iii–67.

Izumi, S. (2003). Processing difficulty in comprehension and production of relative clauses by learners of english as a second language. *Language learning*, *53*(2), 285–323.

Kemper, S. (1987). Life-span changes in syntactic complexity. *Journal of gerontology*, *42*(3), 323–328.

Kemper, S., Kynette, D., Rash, S., O'Brien, K., & Sprott, R. (1989). Life-span changes to adults' language: Effects of memory and genre. *Applied Psycholinguistics*, *10*(1), 49–66.

Kidd, E., & Bavin, E. L. (2002). English-speaking children's comprehension of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of psycholinguistic research*, *31*(6), 599–617.

Kunsmann, P., Gordesch, J., & Dretzke, B. (1998). Native speakers' reactions to modern english usage. *Journal of Quantitative Linguistics*, *5*(3), 214–223.

Levy, R., & Andrew, G. (2006). Tregex and tsurgeon: tools for querying and manipulating tree data structures. In *Lrec* (pp. 2231–2234).

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International journal of corpus linguistics*, *15*(4), 474–496.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of english: The penn treebank.

McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2002). *Coh-metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension* (Tech. Rep.). Technical report, Institute for Intelligent Systems, University of Memphis ....

Miller, G. A. (1962). Some psychological studies of grammar. *American psychologist*, *17*(11), 748.

Olson, G. M. (1973). Developmental changes in memory and the acquisition of language. In *Cognitive development and acquisition of language* (pp. 145–157). Elsevier.

Rimmer, W. (2006). Measuring grammatical complexity: The gordian knot. *Language Testing*, *23*(4), 497–519.

Sclafani, J. (2017). *Talking donald trump: A sociolinguistic study of style, metadiscourse, and political identity*. Routledge.

Sekine, S., & Grishman, R. (1996). The apple pie parser. *http://www. cs. nyu. edu/cs/projects/proteus/app/*.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *47*(2), 276.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69–90.

Vyatkina, N. (2013). Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal*, *97*(S1), 11–30.
Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American philosophical society*, *104*(5), 444–466.