Computerized assessment of syntactic complexity in Alzheimer's disease: a case study of Iris Murdoch's writing

Serguei Pakhomov · Dustin Chacon · Mark Wicklund · Jeanette Gundel

Published online: 10 November 2010
© Psychonomic Society, Inc. 2010

Abstract Currently, the majority of investigations of linguistic manifestations of neurodegenerative disorders such as Alzheimer's disease are conducted based on manual linguistic analysis. Grammatical complexity is one of the language use characteristics sensitive to the effects of Alzheimer's disease and is difficult to operationalize and measure using manual approaches. In the current study, we demonstrate the application of computational linguistic methods to automate the analysis of grammatical complexity. We implemented the Computerized Linguistic Analysis System (CLAS) based on the Stanford syntactic parser (Klein and Manning, *Pattern Recognition*, 38 (9), 1407–1419, 2005) for longitudinal analysis of changes in syntactic complexity in language affected by neurodegenerative disorders. We manually validated CLAS scoring and used it to analyze writings of Iris Murdoch, a renowned Irish author diagnosed with Alzheimer's disease. We found clear patterns of decline in grammatical complexity consistent with previous analyses of Murdoch's writing conducted by Garrard, Maloney, Hodges, and Patterson (*Brain*, 128(250–260, 2005). CLAS is a fully automated system that may be used to derive objective and reproducible measures of syntactic complexity in language production and can be particularly useful in longitudinal studies with large volumes of language samples.

Keywords Syntactic complexity · Computational linguistics · Natural language processing · Dementia · Iris Murdoch

Introduction

Language is one of the most complex human behaviors, and it is sensitive to cognitive impairment resulting from neurodegenerative disorders such as Alzheimer's disease. A number of studies have investigated various aspects of language production and comprehension including sentence structure complexity, idea density, use of referring expressions and discourse coherence (Almor, Kempler, MacDonald, & Andersen, 1999; Almor, MacDonald, Kempler, Andersen, & Tyler, 2001; Bickel, Pantel, Eysenbach, & Schroder, 2000; Kemper, Marquis, & Thompson, 2001; Kempler, 1995; Kempler, Almor, Tyler, Andersen, & MacDonald, 1998; Lyons et al., 1994). With several notable exceptions, the majority of the studies that investigate language impairment in Alzheimer's disease rely on manual linguistic analysis of spoken or written samples obtained from subjects either experimentally or through observation.

One notable exception is a study by Brown et al. (Brown, Snodgrass, Kemper, Herman, & Covington, 2008) that reported on the development of Computerized Propositional Idea Density Rater (CPIDR) that automated the calculation of the idea density score using the same manual approach as in several previous longitudinal studies of Alzheimer's disease (Kemper et al., 2001; Mitzner & Kemper, 2003; Snowden et al., 1996). Another longitudinal study by Garrard et al. (Garrard, Maloney, Hodges, & Patterson, 2005) reports on the use of several computer-assisted methods for assessing syntactic complexity and lexical features of writing samples from Iris Murdoch, a renowned Irish author diagnosed with Alzheimer's disease shortly after publishing in 1995 her last book in a long, award-winning writing career. Murdoch's clinical records including brain imaging and pathology results were used to confirm both the clinical diagnosis and the physical manifestations of Alzheimer's disease in the

S. Pakhomov (✉) · D. Chacon · M. Wicklund · J. Gundel
University of Minnesota,
Twin Cities, MN, USA
e-mail: pakh0002@umn.edu

form of brain atrophy, plaques and neurofibrillary tangles, as well as gliosis and spongiosis in the temporal lobe areas. Details of these neuropsychological and pathological assessments are available in the Garrard et al. publication (Garrard et al., 2005).

Garrard et al. compared syntactic complexity and lexical measures of Murdoch's early and mid-career books with those of her last book, and found significant differences. Murdoch's books were found to be particularly suitable for linguistic analysis as she was known in literary circles for her resistance to any editing of her writing prior to publication, alleviating concerns that the published books may not be representative of her actual language production. While some of the lexical measures reported by Garrard et al. were computed in an automated fashion (i.e., word frequency, word length, ratio of word types to word tokens), the measures of syntactic complexity (e.g., number of subordinate clauses per sentence) were still calculated by hand on small sub-samples of the writings. In addition to the manual counts of subordinate clauses, Garrard et al. also approximated clause counts by dividing the number of words in the writing samples by the total number of sentence-ending markers (periods, exclamation marks and question marks). Another automated measure consisted of the proportion of times ten most common words in each text were repeated within the space of five words ('auto-collocations').

The results of Garrard's analysis revealed clear and statistically significant differences between Iris Murdoch's earlier writings and her last book on measures of lexical content such as word frequency and type-to-token ratio; however, the differences in terms of syntactic complexity were much less clear. The latter findings were inconsistent with previously reported results (Bates, Harris, Marchman, Wulfeck, & Kritchewsky, 1995; Kemper et al., 2001; Kempler, 1995) and were attributed by Garrard et al. to the possibility that their measures of grammatical complexity were not optimally operationalized and were based on relatively small sub-samples of the writings consisting of ten sentences from the first, middle and final chapters of each book.

Operationalizing grammatical (or syntactic) complexity is particularly challenging as it involves detailed linguistic analysis, which is time consuming and subject to inter-rater variability and human error. However, methods developed in the field of computational linguistics based on automated syntactic parsing techniques may be used to aid in analyzing language produced by patients with cognitive impairment. For example, several fully automated measures of syntactic complexity have been successfully used to study the language of patients with mild cognitive impairment (Roark, Mitchell, & Hollingshead, 2007). In the current study, we build on the prior work of Garrard et al. and Roark et al., and demonstrate the use of a fully automated Computerized Linguistic Analysis System

(CLAS) for longitudinal analysis of changes in syntactic complexity in language affected by Alzheimer's disease.

Background

Grammatical complexity

To measure the grammatical complexity of English sentences, we implemented three computational approaches that use the Stanford syntactic parser (Klein & Manning, 2005) to provide information on the hierarchical constituent structure as well as syntactic dependencies between lexical items. The parser produces a hierarchical tree representation of the input sentences as shown in Figs. 1 and 2. Two of the grammatical complexity scoring approaches we used rely on counting the number of branches and the lexical item's depth in the syntactic tree representation of the sentence. These approaches are illustrated in Fig. 1. The Yngve (1960) scoring method assigns a score of zero to each rightmost child under any given node in the tree and increments the score by one in a right-to-left fashion for that child node's siblings.

Thus, for example, the determiner (DT) node for the indefinite article "a" in the noun phrase "a red tail" receives an Yngve score of 2 because it is the first (leftmost) of three siblings under the noun phrase node (NP). Note, that the verb phrase (VP) node has an Yngve score of 1 because it is the second of three children under the S node. The total Yngve score for the indefinite article node in the noun phrase "a red tail" is calculated by traversing the path from the S node down to the "a" node at the lowest (terminal) level. In this case we will add the score of 1 on the VP node

Fig. 1 A simple tree diagram illustrating the computation of Yngve (left) and Frasier (right) syntactic complexity measures. (x indicates path termination for the Frasier method)

Fig. 2 A more complex *tree diagram* of a sentence from Jackson's Dilemma showing the results of an automatic syntactic parse and two scoring methods: Yngve (*left*) and Frazier (*right*). (x indicates path termination points for the Frazier method)

to the score of zero on the next NP node to the right, then another zero on the PP node, then another zero on the second NP node in the path and, finally, a score of 2 on the DT node for a total score of 3.

A more complicated and realistic example is provided in Fig. 2. In this example, the verb phrase (VP) “came upon him at times after Tim’s death” has four children under it: the verb (VBD) and three prepositional phrases (PP). The leftmost child of this VP node with four children receives a score of 3, the next two children from the left a score of 2 and 1, and the rightmost child a score of 0. The final score for each lexical item at the lowest terminal node level in Fig. 2 is the sum of the scores of each branch that dominates that lexical item’s node. For example, the lexical item “was” is dominated by the VBD node (score of 1) that is dominated by the VP node (score of 1) that is dominated by the S node (score 0). Thus, the final score for the lexical item “was” is calculated to be 2.

The Frazier (1985) approach proceeds in a bottom-up fashion and calculates the score for each word by tracing a path from the word up to the highest node that is not the leftmost descendant of another node higher up in the tree (parent). The lexical item receives one “point” for each branch in its upward path, with 1.5 “points” for branches from an S node. For example, in Fig. 1, we start with the pronoun “she” and trace it’s path to the root S node resulting in a score of 2.5. The next lexical item “found” represented by the VBD node only has a path to the VP node at which

point the path terminates because the VP node in this example is not the leftmost descendant of the root S node. Thus, the Frazier score for “found” is 1. A more complicated example of the Frazier approach is illustrated by the scores on the right side of the branches in Fig. 2.

The third scoring method we implemented in CLAS relies on computing the length of grammatical dependencies between lexical items in the sentence (Gibson, 1998; Lin, 1996; Marneffe & Manning, 2008). The grammatical dependency representation is qualitatively different from the tree representations of the Yngve and Frazier methods. A grammatical dependency reflects a specific type of a non-hierarchical relation that holds between two entities. For example, the sentence “He was beginning to experience a loss of identity which now came upon him at times after Tim’s death” in Fig. 3 contains the following set of dependencies extracted by the Stanford parser.

In the example in Fig. 3, the parser identified a nominal subject (“nsubj”) relation that holds between the noun phrase subject of the main clause (“He”) and the verb phrase head in the main clause (“beginning”). The relations with underscores in their labels represent more complex dependencies. For example, the prepositional relation (“prep_of”) is the result of combining a simple prepositional relation [“prep(loss, of)”] and a prepositional object relation [“pobj(of, identity)”]. The numbers in the example in Fig. 3 indicate the word’s serial position in the

Dependency Relation	Distance
nsubj(beginning-3, He-1)	2
xsubj(experience-5, He-1)	4
aux(beginning-3, was-2)	1
aux(experience-5, to-4)	1
xcomp(beginning-3, experience-5)	2
det(loss-7, a-6)	1
dobj(experience-5, loss-7)	2
nsubj(came-12, loss-7)	5
prep_of(loss-7, identity-9)	2
rel(came-12, which-10)	2
advmod(came-12, now-11)	2
rcmod(loss-7, came-12)	5
prep_upon(came-12, him-14)	2
prep_at(came-12, times-16)	4
poss(death-20, Tim-18)	2
prep_after(came-12, death-20)	8
Total:	45

Fig. 3 Automatically parsed syntactic dependencies and the calculation of the dependency length/distance

sentence.¹ Each dependency relation receives a distance score calculated as the absolute difference between the serial positions of the words that participate in the relation. For example, the distance for the nominal subject relation (“nsubj”) is $3-1 = 2$. Based on these dependency relations and the distances in serial positions of the constituent words, the score is calculated as the total length of the dependencies, or the sum of all dependency distances in the sentence.

The depth and the degree of branching captured by the Yngve and Frazier methods have been shown to be associated with working memory demands (Resnik, 1992). The length of grammatical dependencies was also previously shown to be predictive of the processing time in sentence comprehension tasks (Gibson, 1998; Lin, 1996). In the early stages of Alzheimer’s disease, its linguistic manifestations have more to do with the deterioration of semantic features. Semantically “empty” speech characterized by overuse of pronouns has been noted as one of the distinctive features of the disorder (Almor et al., 1999; Kempler, 1995), as well as semantic deficits affecting one’s ability to determine semantic relatedness between concepts (Aronoff et al., 2006). However, cognitive impairment in Alzheimer’s disease was also found to be associated with decreased performance on tasks that involve working memory, particularly in the more advanced stages (Almor et al., 1999; Bickel et al., 2000; Kempler, et al., 1998; MacDonald, Almor, Henderson, Kempler, & Andersen, 2001). Given the possible association with working memory and the deterioration of semantic relations, we expect these measures of grammatical complexity to be sensitive to the effects of Alzheimer’s disease on language production and comprehension.

¹ The Stanford parser relies on the standard Penn Treebank tokenization scheme in which the possessive “’s” is treated as a separate token, thus resulting in a 20-word sentence in this particular case.

Methods

Data

Three of the books selected for this study were the same as those used by Garrard et al. [“Under the Net” (1954), “The Sea, The Sea” (1978) and “Jackson’s Dilemma” (1995)]. We added another book to our analysis (“The Green Knight”) that was published in 1994, immediately prior to the last book, “Jackson’s Dilemma.” This selection was motivated by the desire to test the sensitivity of the automated syntactic and lexical analysis tools in distinguishing writing samples produced within a shorter period of time than the original corpus allowed. The order of the publication of these books is shown in Fig. 4 for reference in the interpretation of the results of this study.

The books were scanned at 600-dpi resolution, rendered as TIFF images and subsequently converted to text using Tesseract,² a freely available optical character recognition program. The errors in the output of Tesseract mainly consisted of misrecognized punctuation and line breaks, and were manually corrected prior to both automated and manual linguistic analyses. We quasi-randomly extracted 20 non-contiguous passages from each of the four books. We avoided selecting dialogue as it constitutes a different type of discourse and is not currently part of the scope of the study. Thus, most of the resulting passages consisted of descriptions of scenes and thoughts attributable to the various characters of the books.

Computerized linguistic analysis system (CLAS)

The computational approaches to syntactic parsing and subsequent determination of the Yngve, Frazier and syntactic dependency length (SDL) scores were implemented in the Java programming language using the Unstructured Information Management Architecture³ as the development platform. The system consisted of four modules arranged as a pipeline with the output of the module used as input to subsequent modules as shown in Fig. 5.

The Tokenizer module identifies individual tokens (e.g., word tokens, punctuation tokens, number tokens, etc.) in the input text and passes this information to the Sentence Detector. The Sentence Detector is based on an OpenNLP maximum entropy (Maxent)⁴ package trained to recognize sentence boundaries. These two modules are necessary to split the text into sentences and pass each sentence for subsequent syntactic parsing and complexity scoring by the downstream modules. For the purposes of this study, CLAS

² <http://code.google.com/p/tesseract-ocr/>

³ <http://uima.apache.org/>

⁴ <http://opennlp.sourceforge.net/>

Fig. 4 Timing of Iris Murdoch's books

was configured to produce the following measurements for each sentence in the sample passages:

- Mean number of words (Utterance Length)
- Mean number of clauses [count of S nodes in the parse tree (Fig. 2)]
- Total Yngve depth (Yngve Depth)
- Total Frazier depth (Frazier Depth)
- Total syntactic dependency length (SDL)

The means of these measurements were compared across the four books to determine if there was any evidence of decline in any of the measurements for the books published later in Iris Murdoch's life.

Manual validation of syntactic complexity measures

We randomly selected 10 sentences from three of the books ("The Sea, The Sea", "The Green Knight" and "Jackson's Dilemma") for a total of 30 sentences that were parsed using the Stanford parser and manually scored by a trained linguist (DC) for syntactic complexity following the algorithms described in the Background section. We compared the manually obtained scores to those produced by CLAS for the Yngve Depth, Frazier Depth and SDL approaches. This comparison was performed to test the functionality of the computerized tools and to ensure their consistency with human scores.

Statistical analysis

To compare manual and automated Yngve, Frazier and Syntactic Dependency Length scores, we calculated the

mean difference and 95% confidence interval for the two sets of scores for each approach. Confidence intervals around estimated means were calculated based on the binomial distribution. To compare syntactic complexity scores across the four books, we used one-way ANOVA with subsequent pairwise post-hoc t-tests using Tukey's Honestly Significant Differences (HSD) approach to adjust for multiple comparisons. Test results were considered significant if the p-value was less than 0.05. All statistical calculations were carried out using R statistical package (version 2.10.0).

Results

General corpus statistics

The mean length of the 80 passages from the four books was 20.71 sentences or 331 words (approximately 1 printed page in the original) per passage. The total size of the collection of passages from all four books used in this study was 26,484 words found in 1,657 sentences.

Validation of automated syntactic complexity measures

The comparison between Yngve, Frazier and Syntactic Dependency Length scores obtained automatically by CLAS and manually by a trained linguist (DC) showed a high degree of agreement. The mean difference between the manual and automatic scores for total Yngve depth was 1.97 (SD = 4.1), total Frazier depth 1.35 (SD = 1.00) and total Syntactic Dependency Length 0.02 (SD = 0.17).

TEXT >> Tokenizer >> Sentence Detector >> Syntactic Parser >> Complexity Score Generator >> OUTPUT

Fig. 5 Sequence of CLAS processing modules

Syntactic complexity

The results of the comparisons between the means in sentence length and number of clauses per sentence are shown in Fig. 6. We observe a clear pattern of decline in these two measurements. Both measurements are highest on the first book (“Under the Net”) and lowest on the last book (“Jackson’s Dilemma”). Table 1 shows the results of the utterance length and number of clauses measurements obtained in Garrard et al. study compared to those obtained in the current study. Garrard’s et al. counted only subordinate clauses in a given sentence; however, our automated parsing approach counted the entire sentence (the matrix clause) as well. Thus, to compare these counts, we subtracted the top S node (see Fig. 2) representing the matrix clause. Both counts including and excluding the top S node are presented in Table 1. The ANOVA test of the differences between the means across all four books for both the utterance length ($F = 10.2$, $p < 0.001$) and the number of clauses ($F = 9.88$, $p < 0.001$) are statistically significant.

Utterance length Post-hoc tests show significant differences between “The Green Knight” and “Under the Net” ($p < 0.001$), “Jackson’s Dilemma” and “Under the Net” ($p < 0.001$), and “Jackson’s Dilemma” and “The Sea, The Sea” ($p < 0.05$). No significant differences were found between “The Sea, The Sea” and “Under the Net,” “The Green Knight” and “The Sea, The Sea,” or “The Green Knight” and “Jackson’s Dilemma.”

Number of clauses There were significant differences between “The Sea, The Sea” and “Under the Net” ($p = 0.030$), “The Green Knight” and “Under the Net” ($p = 0.001$), and “Jackson’s Dilemma” and “Under the Net” ($p < 0.001$). No other differences were found to be significant

with the post-hoc tests including the difference between “Jackson’s Dilemma” and “The Sea, The Sea.”

The results of the comparison among the four books in terms of the three computational measures of syntactic complexity (Yngve Depth, Frazier Depth and SDL) are presented in Fig. 7. Overall, these results follow a similar pattern of decline across the four books with the earlier writing sample, “Under the Net,” having greater syntactic complexity than the later books, particularly “Jackson’s Dilemma.” The means for the SDL measure pattern differently for the mid-career book “The Sea, The Sea” and late-career “The Green Knight,” with the results for these books being slightly higher than the first and last books. Significant differences between the books were found with all three approaches using ANOVA testing: Yngve Depth ($F = 3.36$, $p = 0.018$), Frazier Depth ($F = 8.62$, $p < 0.001$), SDL ($F = 4.74$, $p = 0.003$). However, the results of the post-hoc analysis detailed below were less clear than the results for the utterance length and number of clauses per sentence.

Yngve depth Post-hoc tests revealed significant differences only between “Jackson’s Dilemma” and “Under the Net” ($p = 0.008$). No other differences were significant.

Total Frazier depth Significant differences were found between “The Green Knight” and “Under the Net” ($p = 0.017$), “Jackson’s Dilemma” and “Under the Net” ($p < 0.001$), and “Jackson’s Dilemma” and “The Sea, The Sea” ($p = 0.049$). No significant differences were found in any other pairwise comparisons including “The Sea, The Sea” and “The Green Knight.”

Syntactic dependency length (SDL) The following differences were significant: “Jackson’s Dilemma” and “Under the Net” ($p = 0.049$), “Jackson’s Dilemma” and “The Sea,

Fig. 6 Comparison of the mean sentence length and the mean number of clauses per sentence computed by CLAS across four Iris Murdoch’s books

Table 1 Sentence length comparison between the analysis reported in Garrard et al.'s and the current study

	Under the Net		The Sea, The Sea		Green Knight		Jackson's Dilemma	
	M	SD	M	SD	M	SD	M	SD
Garrard's manual analysis								
Words per sentence	19.7	11.6	22.7	14.1	NA	NA	15.3	10.8
Clauses per sentence	1.5	1.8	2.2	1.0	NA	NA	0.8	1.3
Garrard's automated analysis								
Words per sentence-boundary mark	13.9	-	15.1	-	NA	NA	14.6	-
CLAS automated analysis								
Words per sentence	18.2	12.9	16.3	11.0	15.6	11.1	13.9	10.5
Clauses per sentence								
including the top S node	2.8	2.0	2.4	1.8	2.3	1.8	2.1	1.6
excluding the top S node	1.8	2.0	1.4	1.8	1.3	1.8	1.1	1.6

The Sea" ($p = 0.022$), and "Jackson's Dilemma" and "The Green Knight" ($p = 0.003$).

Discussion

Computerized linguistic analysis system (CLAS)

In this study, we demonstrated the application of a computerized system that implements several computational linguistic approaches to measuring syntactic complexity of

English utterances in a longitudinal study of language production affected by Alzheimer's disease. We conducted functional validation of CLAS and found that the differences between the manually derived scores and the automatically computed scores were minimal. CLAS complements prior work by Brown et al. (2008) that focused on the implementation of Kintsch and Keenan's (1973) methodology for measuring propositional content (idea density) of language production. Manually computed idea density and syntactic complexity have been used extensively as part of one of the largest longitudinal studies of Alzheimer's disease ("The Nun Study") to investigate the relationship between

Fig. 7 Mean total Yngve Depth, Frazer Depth and Syntactic Dependency Length (*SDL*) scores for four Iris Murdoch's books

linguistic abilities early in life and the risk of developing Alzheimer's disease later in life (Snowdon et al., 1996), as well as for longitudinal assessment of written (Kemper et al., 2001) and oral (Mitzner & Kemper, 2003) language use. The notion of syntactic complexity, however, is difficult to operationalize, as evidenced by a prior study by Garrard (Garrard et al., 2005); it is also hard to scale to larger samples of data such as web blogs, personal writings over a lifetime, speeches, diaries and sermons. Computer-aided linguistic analysis of these voluminous longitudinal samples can enable more detailed and objective measurement of changes in linguistic abilities over time. We anticipate that these tools would be used more in research focused on understanding brain-behavior relations rather than as diagnostic instruments. Language use is highly variable and to have diagnostic utility methods like the ones described in this paper would need to track patient performance over long periods of time. Even so, the precise point that would indicate abnormal decline may be difficult to determine. However, a more immediate clinical use of such instruments may be realized in the context of developing interventions aimed at treating/reversing the causes of dementia as well as other disorders affecting language. Currently no such treatments are available for Alzheimer's disease, but in order to develop effective treatments and test them in clinical trials, one must have an objective and reliable way to track patient's cognitive performance. Tools for automated speech and language analysis may provide an ability to do so.

Syntactic complexity in Iris Murdoch's writing

In this analysis of Iris Murdoch's writings, we found clear patterns of decline by several computerized measures of syntactic complexity across the four books that we examined.⁵ First, our results with the measurements of the mean sentence length and number of clauses per sentence are consistent with those obtained by Garrard et al. (Garrard et al., 2005). The sentence length means obtained in the current study are more in-line with Garrard's automated "words per sentence-ending mark measure" obtained on larger samples than the "words per sentence measure" manually computed from smaller 30-sentence samples. The mean number of clauses computed in the current study from automated sentence parsing after excluding the top S node in the parse trees is also comparable to the manually obtained number of subordinate clauses in Garrard's study. There are minor differences between these two sets of scores; however, both display the same decreasing trend across the books.

⁵ However, we cannot rule out the possibility that decline in syntactic complexity may be attributed to the effects of aging rather than Alzheimer's disease.

Second, using syntactic complexity measures computed by CLAS, we found significant differences in complexity between Murdoch's earlier writings and those she wrote later in life. These findings are in line with another study of longitudinal written language samples obtained from a historical figure, King James I of England (1566–1625), who reportedly suffered from an illness that resulted in cognitive impairment symptoms (Williams, Holmes, Kemper, & Marquis, 2003). This study attempted to use simple measures of syntactic complexity (mean sentence length and mean number of clauses per sentence) in King James' letters and found a pattern of linguistic complexity decline atypical of normal aging, but with timing of onset more consistent with vascular dementia rather than Alzheimer's disease.

Declines in syntactic complexity of sentences among other linguistic abilities have been extensively investigated in people with Alzheimer's disease and mild cognitive impairment (Garrard et al., 2005; Harper, 2000; Kempler, 1995; Roark et al., 2007; Williams et al., 2003), as well as in healthy aging adults (Glosser & Deser, 1992; Marini, Boewe, Caltagirone, & Carlomagno, 2005). The latter studies showed relative stability of *micro*-linguistic abilities (e.g., word use, syntax, phonology at an individual utterance level) across the young adult (25–39 years old) and young elderly (60–74 years old) groups, with significant and sharp declines present in more advanced age (> 74 years old). Iris Murdoch was 75 years old when her last book "Jackson's Dilemma" was published and 74 years old during the publication of "The Green Knight," which is right at the boundary where significant declines in microlinguistic abilities including syntax were found in healthy aging adults. Thus, our findings provide additional support for the syntactic preservation hypothesis proposed by Kempler et al. (Kempler, 1995) which suggests that people diagnosed with Alzheimer's disease tend to maintain more automatic linguistic functions such as syntax until fairly advanced stages in the disease progression, while other higher level linguistic functions, including semantic memory, thematic content and reference, may be impaired at the earlier stages (Almor et al., 1999; Kempler, 1995). Our findings also underline the need for age and education-based norms for written output in order to make either manual or automated language analysis tools more effective at studying how cognition is impacted by neurodegenerative disease (Venneri, Forbes-Mckay, & Shanks, 2005). While we certainly cannot tease out the effects of normal aging, the decline in grammatical complexity seen from 1994–1995 exceeds the rate of change from 1954–1978, or from 1978–1994, indicating an acceleration that may be more attributable to the effects of Alzheimer's rather than normal aging.

Limitations and future work

Our study has a number of limitations that bear on the interpretation of the results. First, this is a study of writing samples from a single author, which limits the ability to generalize from the current findings. However, using Iris Murdoch's writings to develop and test computational methods for language analysis has a number of distinct advantages due to the availability both of detailed neuropsychological, imaging and pathology results confirming the diagnosis of Alzheimer's disease and of longitudinal samples of language production. Furthermore, documentation of the fact that Iris Murdoch resisted any editorial intervention alleviates the attribution concerns that would typically be associated with studies that use published works to investigate the effects of neurodegenerative disease on language. Another limitation is that our system is currently trained on English text only. While this limitation does not affect the interpretation of the current results, it does limit the applicability of our approach to analyzing speech from speakers of other languages. Automated parsers have been developed and trained for a number of other languages; however, additional development of syntactic complexity scoring, as well as further validation and testing will be required to extend CLAS to other languages. Also, the current implementation of CLAS is designed specifically for written discourse. In future work, we will introduce a number of modifications that will enable processing of spontaneous speech samples. These modifications will include prosody-based utterance segmentation, dysfluency and repair detection, and robust shallow parsing to process incomplete sentences.

References

- Almor, A., Kempler, D., MacDonald, M., & Andersen, E. (1999). Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain and Language*, 67, 202–227.
- Almor, A., MacDonald, M., Kempler, D., Andersen, E., & Tyler, L. (2001). Comprehension of long distance number agreement in probable Alzheimer's disease. *Language and Cognitive Processes*, 16(1), 35–63.
- Aronoff, J., Gonnerman, L., Almor, A., Arunachalam, S., Kempler, D., & Andersen, E. (2006). Information content versus relational knowledge: Semantic deficits in patients with Alzheimer's disease. *Neuropsychologia*, 44, 21–35.
- Bates, E., Harris, M., Marchman, V., Wulfeck, B., & Kritchewsky, M. (1995). Production of complex syntax in normal aging and Alzheimer's disease. *Language and Cognitive Processes*, 10, 487–544.
- Bickel, C., Pantel, J., Eysenbach, K., & Schroder, J. (2000). Syntactic comprehension deficits in Alzheimer's disease. *Brain and Language*, 71(3), 432–448.
- Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, 40(2), 540–545.
- Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing*. Cambridge, UK: Cambridge University Press.
- Garrard, P., Maloney, L., Hodges, J., & Patterson, K. (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain*, 128(250–260).
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 61(1), 1–76.
- Glosser, G., & Deser, T. (1992). A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. *Journal of Gerontology*, 47(4), 266–272.
- Harper, L. (2000). *Sentence production in Individuals with Alzheimer's disease*. Calgary: University of Calgary.
- Kemper, S., Marquis, J., & Thompson, M. (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. *Psychology and Aging*, 16(4), 600–614.
- Kempler, D. (1995). Language changes in dementia of the Alzheimer's type. In Lubinski (Ed.), *Dementia and communication: Research and clinical implications* (pp. 98–114). San Diego: Singular.
- Kempler, D., Almor, A., Tyler, L. K., Andersen, E. S., & MacDonald, M. C. (1998). Sentence comprehension deficits in Alzheimer's disease: A comparison of off-line vs. on-line sentence processing. *Brain and Language*, 64(3), 297–316.
- Kintsch, W., & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, 5(3), 257–274.
- Klein, D., & Manning, C. D. (2005). Natural language grammar induction with a generative constituent-context model. *Pattern Recognition*, 38(9), 1407–1419.
- Lin, D. (1996). *On the Structural Complexity of Natural Language Sentences*. Paper presented at the Computational Linguistics (COLING) Conference.
- Lyons, K., Kemper, S., LaBarge, E., Ferraro, R., Balota, D., & Storandt, M. (1994). Oral language and Alzheimer's disease: A reduction in syntactic complexity. *Aging and Cognitive*, 1(4), 271–281.
- MacDonald, M., Almor, A., Henderson, V., Kempler, D., & Andersen, E. (2001). Assessing working memory and language comprehension in Alzheimer's disease. *Brain and Language*, 78, 17–42.
- Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research*, 34(5), 439–463.
- Marneffe, M., & Manning, C. (2008). *The Stanford typed dependencies representation*. Paper presented at the Workshop on Cross-Framework and Cross-Domain Parser Evaluation.
- Mitzner, T. L., & Kemper, S. (2003). Oral and written language in late adulthood: Findings from the Nun Study. *Experimental Aging Research*, 29(4), 457–474.
- Resnik, P. (1992). *Left-corner parsing and psychological plausibility*. Paper presented at the Computational Linguistics (COLING) Conference.
- Roark, B., Mitchell, M., & Hollingshead, K. (2007). *Syntactic complexity measures for detecting Mild Cognitive Impairment*. Paper presented at the ACL 2007 Workshop on Biomedical Natural Language Processing (BioNLP), Prague, Czech Republic.
- Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. *JAMA*, 275(7), 528–532.
- Venneri, A., Forbes-McKay, K. E., & Shanks, M. F. (2005). Impoverishment of spontaneous language and the prediction of Alzheimer's disease. *Brain*, 128(Pt 4), E27.
- Williams, K., Holmes, F., Kemper, S., & Marquis, J. (2003). Written language clues to cognitive changes of aging: An analysis of the letters of King James VI/I. *Journal of Gerontology and Psychological Sciences*, 58B(1), 42–44.
- Yngve, V. (1960). *A model and an hypothesis for language structure*. Paper presented at the American Philosophical Society.