Evaluating the Level of Semantic Complexity in Prose Argumentation

Pamela Toman
353 Serra Mall
Stanford, CA 94305
ptoman@stanford.edu

Pujun Bhatnagar
353 Serra Mall
Stanford, CA 94305
pujun@cs.stanford.edu

Zhiyang He
353 Serra Mall
Stanford, CA 94305
hzyjerry@stanford.edu

1 Introduction

Integrative complexity is a construct introduced by Tetlock and Suedfeld (see [19, 21]) to study the intellectual style used by individuals or groups in processing information, problem solving, and decision making. It measures the extent to which an author entertains multiple perspectives and integrates them into coherence: discourse with low integrative complexity admits the existence and legitimacy of only a single perspective, whereas discourse with high integrative complexity allows the simultaneous correctness of multiple perspectives. Table 1 in the Appendix gives examples of low and high integrative complexity.

Political and clinical psychology motivates integrative complexity. For example, Suedfeld & Tetlock [20] show that declines in integrative complexity in diplomacy during times of crises is a leading indicator of war, while increases in integrative complexity is a leading indicator of reaching compromise agreements that avert war; Winter [23] finds that police officers who exhibit primarily low integrative complexity are more inclined to act violently under stress; and Liht and Savage [9] use it as a good measure of effectiveness of an intervention program designed to reduce extremism among young UK Muslims.

Researchers interested in using the integrative complexity metric must currently invest in laborious manual scoring, as automatic coding for this construct is in its infancy. Conway et al. [3] develop the first automated system based on detection of surface linguistic markers. They achieve modest relationships with human-scored integrative complexity (average Cronbach's alpha = .62; average r = .46) [3], where expert human performance is assessed to be an alpha of at least 0.85 [1]. Through leveraging and extending techniques used in related research in integrative complexity, automated essay scoring, and argument mining, in addition to work covered in class on natural language inference, we aim to improve on this performance.

2 Automated Integrative Complexity

Integrative complexity is traditionally measured on a 7-point scale with heavy human involvement. The training requirement is not insubstantial: would-be coders take a training course with extensive discussion on a large number of examples,[1] and must pass a test with at least 0.85 agreement with expert coders before coding new texts.

The unit of analysis is usually the paragraph, though the technique can be applied to larger work. Scores of 1 reflect single-idea thinking. Scores of 3 reflect *differentiation*, or multiple alternatives or perspectives, but do not express how those alternatives can be simultaneously true. Scores of 5 express *integration* of multiple ideas: the unit might state the way in which distinct statements can be simultaneously true through synthesizing perspectives, admitting mutual influence, or leveraging causation. Scores of 7 are rare; they are assigned to paragraphs that complicate even the way in which multiple ideas are integrated. The even-numbered scores enable shades of difference between these landmark scores.

The rest of this section discusses automated formalizations of integrative complexity. Two approaches, both published in 2014, attempt to build systems that automatically score this human-insight-intensive metric: Conway et al. [3] develop a system using a set of human-developed weights, and Kannan-Ambili [7] explores machine learning methods and proposes a semantic coherence feature based in WordNet that improves performance over baselines. We also discuss two additional approaches that measure slightly different semantic complexity constructs from integrative complex-

[1] Materials and workshop information is available at http://www2.psych.ubc.ca/ psuedfeld/index2.html.

ity, one from Pennebaker [14] and one from Hermann [27].

2.1 Approach of Conway et al.

Conway et al. [3] build a system that emulates the 7-point scale used in human-coded work, and find correlations between 0.34 and 0.57 and alphas between 0.30 and 0.72 to a human gold standard on out-of-sample datasets. To do so, they generate two large lists of words and phrases relevant to differentiation and integration. They generate these lists through identifying a large number of seed words by hand and expanding the set repeatedly through documented synonymy relations. Conway et al. then combine this bag-of-words through weighting each observed keyword by statistically-informed-then-hand-tweaked probabilities of indicating differentiation or integration. They develop final scores with a two-staged approach: first the system amasses a score for differentiation, which is capped at 3, and then the system adds additional points for keywords associated with integration. They train on three datasets and test on six others, totaling about 1300 scoreable paragraphs.[2]

Conway et al. report the best performance on this problem to date. However, the model used is very simple. Its use of surface features and hand-estimated weights for keywords require a level of human involvement that means the approach cannot easily be transferred to new languages or genres – and in fact they note that language changes over time pose a challenge to their system, finding that the system performs much worse on the Nixon-Kennedy debates from the 1960s than on more recent, personal writing. Additionally, the limited amount of human-labeled texts poses an ongoing challenge for this problem: 1300 paragraphs limit the extent to which meaningful deep features can be associated with their contributions to correct labeling.

Our primary take-away from this work, as in much of the other work that we will discuss, is in two parts: there is a need for work that goes beyond hand-engineered bags of words, and successful work will need to leverage alternative external data sources to improve performance on this problem.

[2]Luke Conway has shared most of these datasets with this team, excluding datasets that have IRB restrictions.

2.2 Approach of Kannan-Ambili

The second published approach is a Masters thesis from Aardra Kannan-Ambili [7]. Kannan-Ambili explores a variety of machine learning algorithms, including logistic regression, support vector machines, and multi-class classifiers. The thesis offers only minimal discussion of the features used, but does propose a "semantic coherence" measurement that averages a function of the path length and path depth in WordNet between words in the first (assumed to be topic) sentence and the words in each succeeding sentence.

After rebinning the 1-7 scale to a 1-3 scale, Kannan-Ambili achieves accuracies of 0.75 to 0.83 given 10-fold cross-validation. She does not report accuracies on the 7-point test, nor does she report correlation or alpha scores. Additionally, it appears that Kannan-Ambili's train and test data is limited to the 167-item practice test made available by Suedfeld's supporting materials website.

Kannan-Ambili's semantic coherence metric is the only feature in the literature specifically designed for estimating integrative complexity. We are enticed by the idea of measuring semantic coherence to identify whether a paragraph addresses exclusively the same theme or multiple variants on a theme. However, we question whether the assumption that the first sentence of a paragraph is the "topic" sentence is valid, and would have preferred to see a multi-stage model with tagging of topic sentences or experimental results exploring the effect of assuming the first sentence is a topic sentence.

We are intrigued by the potential for working on semantic coherence in a vector space: Is it possible to build an alternative semantic coherence feature that is based on locating clauses in a high dimensional space and then tracking the relative locations and trajectory patterns that the author takes in forming a paragraph? Such an approach is attractive, though it would require some additional thought to identify whether and how integration relationships as well as differentiation relationships could be captured in such a framework.

2.3 Other Semantic Complexity Measures

Other proposed semantic complexity measures outside integrative complexity include Pennebaker's "cognitive complexity" [15] and Hermann's "conceptual complexity" [5]. Pennebaker was one of the first people to work in the realm of semantic complexity. He developed his cognitive complexity as a conglomeration of features in the early writing assessment tool Linguistic Inquiry and Word Count [14]. Pennebaker's approach is a purely bag-of-words-using-counts model, and it does not address the extent to which concepts are integrated together.

Slightly later, Hermann and others [5] developed a metric for conceptual complexity as part of work formalizing and measuring leadership traits in politics. The Hermann and Young approach [27] has rules that incorporate broad range of possible word groupings and applies a variety of algorithms to determine the likelihood of language being complex, based on the positioning of typical complexity indicators in relation to other words and indicators. Hermann's conceptual complexity again focuses primarily on differentiation [5].

We note that both of these formalizations of semantic complexity and their automated counterparts are distinct from integrative complexity: Conway et al. [3] reports that Hermann's automated complexity score correlates with human-scored integrated complexity at 0.21 (with human-scored paragraphs having a slight negative correlation of -0.05), and that Pennebaker's complexity score correlates at 0.14. However, their approaches help contextualize the literature on approaches to measuring the semantic complexity of texts.

2.4 Overall Synthesis

Overall, we find that there are three major approaches to this problem:

Bags of words combined heuristically as used by Pennebaker [14], Conway et al.[3], Houck et. al. [6], where the heuristics are derived through substantial human insight

Rule-based multiple-pass approach as discussed in Young and Hermann [27], where the rules are derived through substantial human insight

Machine learning approaches as used by Kannan-Ambili [7], where the relation of features to predictions is not generated by humans but rather learned using supervised methods

In general, we note that all touchstone work and models in this realm have come out of psychology and political science: Pennebaker, Suedfeld, and Conway are psychologists; Hermann and Tetlock are political scientists. Unsurprisingly, the work thus far has focused almost exclusively on simple surface-level features such as the presence of phrases like "but", "nevertheless", and "on the other hand" combined via counts and rule-based approaches, with a few forays into the light machine learning used in the computational social sciences such as logistic regression. Kannan-Ambili's Masters thesis in Artificial Intelligence from the University of Georgia is an exception, but it too touches only briefly on the realm of non-surface feature extraction.

The machine learning approaches, however, hold substantial unpursued promise. We are hopeful that the features from the related literature as well as those that address semantic coherence, the relation of clauses to each other, and potentially deeper underlying representations like framing metaphors, may improve performance. We focus the rest of the literature review on two literatures not covered in class that have potential in understanding and scoring argumentation structure, automated essay scoring and argumentation mining, before turning to an integration of ideas from these sources and class.

3 Automated Essay Scoring

Although the existing systems perform reasonably well at identifying the presence of multiple perspectives (differentiation), most struggle to identify integration. Their challenges on integration suggest that perhaps integration is flagged less by syntactic structures and more by semantic content or non-surface indicators of argument structure. Since differentiation and integration are usually embedded in the structures of the text, argument discourse structure extraction, widely applied in automatic essay scoring (AES), offers potential improvements.

3.1 Approaches to Automated Essay Scoring

Several techniques are employed by the Educational Testing Service (ETS) arm of the AES community in order to do automated essay evaluation. Most of the state of the art techniques, as per Deane [4], use 11 primary features: organization, development, grammar, usage, mechanics, style, average word length, median word frequency, positive features and two content features.

Many of these are simple correlates of good writing (like checks for grammar, usage, mechanics, style, markers of the presence of positives features of good writing, topic-specific vocabulary usage, and lexical complexity). Other features, such as those for organization and development, measure text structure through classifying sentences in into essay-disclosure categories: introductory material/background, thesis, main ideas, supporting ideas and conclusion. Some researchers such as Shermis et al. [17] have shown that a template based approach, where each sentence is matched to a template before scoring and then analyzed based on the matching template, show promise.

Outside the realm of ETS's focus on correlates of argument strength, Stab and Gurevych [18] developed a system that identifies argumentative structure without matching explicit discourse markers as strings. They aim to study the relationship between different argument components and score the essay based on how well different arguments work with each other. They use a machine learning approach for classification that uses structural, lexical, syntactic and contextual features. They first apply multiclass classification on the discourse to identify different argument components. Then for each pair of related arguments, they use binary classification to decide whether each argument is supportive or non-supportive of the thesis. Since the approach does not rely on discourse markers, it has very good performance on extracting implicit and non-adjacent arguments. This inspires us to develop a structure-based method to identify and study integrative complexity.

In another study, Persing and Ng [16] built a computational model for scoring argument strengths based on a new finer-grained argument strength metric using scores of 1-4. Different from scheme-based argument studies like that of Stab and Gurevych, the new model incorporates multiple linguistic metrics, including POS n-grams, semantic frames, transitional phrases, coreference, prompt agreement, argument component predictions (ACP) and argument errors (ARE) to label argument scores. After labeling all the sentences in corpus using these metrics, they then convert a set of heuristic scoring rules into features for training a regressor. Since this method is much more in depth semantically, its result is very promising: it outperforms baseline models by 16.1% on 10,000 human labeled essays from the ICLE corpus. The semantic methods applied in this model tap into deeper semantic and natural language inference levels, and inspire us to go beyond syntactic study of integrative complexity: perhaps we can measure IC by comparing how much the arguments oppose/support each other.

3.2 Synthesis

Although grading of grammar and fluency through syntax have been achieved in AES, grading based on intellectual complexities long remains a challenge. Traditional complexity analysis for AES determines semantic and argument complexities based on set of predefined templates, but this approach is limited by those same templates: variations on a known template can produce both increase/decrease in complexity. Recent studies dig into the underlying argumentative structure and suggest computational models for scoring essays. These works have achieved promising results in contextualizing and analyzing arguments.

Our take away from studying AES is that it is possible to apply machine learning and regression to extract and study different argumentative components in the text. We are optimistic about experimenting with potential syntactic correlates of argument strength as well as applying semantic features mentioned in the AES studies.

4 Argumentation Structure

Fundamental to integrative complexity is the notion of argument structure: paragraphs that make straightforward arguments are low in complexity, and paragraphs that embrace nuance are complex. Based on this insight, we seek out literature specifically related to transforming paragraphs into mathematical structures of argument structure, potentially either for preprocessing or for insights into potentially useful features.

Abundant previous linguistic studies address formalizing argumentation. Fields spawned include rhetorical structure theory, intentional structure theory, and graph-based representations of discourse. Within the frameworks and datasets developed in this realm, recent research like Palau and Moens [12] and Madinani et al. [10] apply machine learning to automate argument detection.

4.1 Frameworks for Argumentation Structure

Rhetorical structure theory (RST) posits that all texts can be decomposed into a composition of relations. For instance, two phrases in a text might have an "antithesis", "motivation", or "justification" relation hold between them. This approach to understanding and the features that are useful in producing these annotations are discussed in Carlson and Marcu [2]. The RST relations are applied recursively, similar to traditional parsing. After analysis under RST, a paragraph has a tree-like structure that describes its argumentation approach; a treebank is available.

However, there is limited recent work automating rhetorical structure theory. Today, many researchers seem to prefer formalizing argument structure as a graph. The prime thrust of this area is from Wolf and Gibson [25, 24], who argue that tree structures are inappropriate for representing discourse structure because they require too much of a deviation from natural representation. They give the example paragraph "[Farm prices in October edged up 0.7% from September] [as raw milk prices continued their rise,] [the Agriculture Department said.] [Milk sold to the nation's dairy plants and dealers averaged $14.50 for each hundred pounds,] [up 50 cents from September and up $1.50 from October 1988,] [the department said.]", noting multiple elaborative, cause-effect, similarity, and attribution relations between each element in the text. They also provide a corpus of graph-based discourse structures and an annotation guide that does not assume hierarchical relationships.

4.2 Selected Modeling Efforts

Moving into work that specifies models within these frameworks, Palau and Moens [12] provide a context-free grammar (CFG) approach to finding the sections of text that represent premises, support, conclusions, and other argument structure, and they compare it to a statistical machine learning approach. Their CFG is a basic proof of concept; it uses only surface features like the presence of the phrases "for these reasons" and "in the light of all the material". Their machine learning approach is more complex, and uses features like semantic relatedness (using both WordNet distances and word vectors), tense of main verb, and context-sensitive phrase matching for rhetorical patterns. Their set of features draws on Teufel's work on argumentative zoning of regions like "background" and "aims" in scientific papers [22] using features like the presence of a citation, linguistic features of the first finite verb, and the presence of named entity types. Palau and Moens achieve F_1 scores near 0.74 for the statistical approaches and near 0.67 for the context-free-grammar approach. We note that the relatively high score they achieve for a very simple CFG approach suggests that a tree-based argumentation formalism, even if not "correct", may have value as an early approximation.

Madnani et al. [10] explore an automated system for detecting organizational elements necessary for arguments. The authors distinguish structural "shell", such as "The argument states that ..." from specific claims, which they refer to as "meat". In the research they develop a hybrid detection method specifically looking for these "shells". Their hybrid method has three components: (i) A linguistic rule-based system that extracts structural patterns and trains n-grams, (ii) A supervised-sequence model that uses simple patterns to calculate shell probabilities, and (iii) a lexical baseline that weighs word phrases based on their frequency in persuasive essays. The system reaches high argument detection accuracy of F_1 = 0.736 and Cohen's $kappa = 0.699$ on human labeled data. Compared to other studies based on detection of discourse markers, which are single words, shell detection captures longer sequences that express more complex relationships between the components of an argumentative discourse. Because of its ability to capture more complex relations, this approach to feature engineering holds substantial promise for classifying integrative complexity.

4.3 Synthesis

We are comfortable with tree-based representations of thesis-and-support in Western writing with a thesis. In fact, we suspect that a paragraph like the motivating example from Wolf and Gibson would be "unscoreable" within integrative complexity framework because it is purely descriptive. Even from a theoretical perspective, then, we are comfortable using and benefiting from the computationally tractable representation of a tree.

One seemingly unrecognized assumption of the argumentation mining literature is that all meaningful relations are explicit in the text as base phrases or clauses. This is likely a problematic assumption, though potentially not particularly damaging at this early stage in tool development, since one can easily imagine a paragraph that has meaningful units or relations represented implicitly. For instance, in a paragraph with an integrative complexity score of 4 that hints at integration through juxtaposition, there would be no explicit thesis on which to derive argument structure. Alternatively, a paragraph might be so one-dimensional that the author doesn't realize the lack of an explicit thesis. We wonder whether a vector space model that focuses on relations between explicit statements might be able to uncover this implicit structure.

5 Future Work

Given this literature review, we are leaning toward two distinct approaches. The first uses neural networks to build out a sentence-level vector space. By projecting the sentences of newly observed paragraphs into that space, we hope to derive a features related to the way in which the sentences relate to each other. The second implements and then extends the hand-engineered feature work already present in the realms of automated essay scoring, argumentation structure, and natural language inference.

5.1 Vector-Space Sentence Coherence Model

A vector space model of paragraph semantic coherence for integrative complexity is related to work on building sentence or paragraph-level representations from word vectors. Given a paragraph-based vector space, we suspect it may be possible to measure the directionality and/or distance of change between nearby sentences in a paragraph to assess the ways in which the sentiments expressed are related to each other. Basic compositions of word vectors to form paragraph units were explored by Mitchell & Lapata [11], and recent highlights in the realm of paragraph representations include Yessenalina & Cardie [26], who learn semantic compositions in the sentiment realm, and Le and Mikolov [8], who build paragraph vectors based on predictions of the next word in a stream with the implementation doc2vec. We are excited about this approach, as building the underlying vector space from a huge quantity of data is possible independent from the approximately 1000 hand-coded paragraphs.

Given the focus on hierarchical multi-pass features in the psychology literature, as well as the lack of clear guidance from theory regarding features beyond bags of words, we suspect that a neural network approach that learns its own features may succeed in building out a vector space, and that a classifier based on the relative relationship of the sentences in that space to each other – potentially in tandem with surface and deeper features – holds potential for making progress in automating measures of integrative complexity.

5.2 Deeper Hand-Engineered Features

As an alternative and more traditional approach, we imagine a three-staged effort to hand-engineering features:

1. As a baseline aligned with most existing work, implement simple surface phrase-matching features; improve performance by using ordinal regression methods as per Pedregosa-Izquierdo [13]

2. Add features drawn from the literature on automated essay scoring, argumentation mining, and natural language inference (e.g., [10, 12, 16, 22], as well as papers not discussed here but covered in class)

3. Develop novel features that have not been mentioned in the literature, such as a pre-processing system for argumentation mining, semantic coherence features, sentiment parsing, and patterns in the use of determiners that reflect the introduction of new material

We hope that work feature engineering realm can be supported by a variety of existing datasets, thereby reducing our reliance on the approximately 1000 paragraphs generously shared by Luke Conway. In particular, the PPDB, SICK, and SNLI might allow feature engineering for identifying restatements of existing knowledge or distinguishing restatements from new statements, the Carlson and Marcu argument structure treebank and Wolf and Gibson graphbank might allow feature engineering for arguments.

Due to the limited gold standard data and lack of sustained interest in the fields of AES and argumentation structure, we suspect that sentence-level vector approach has more potential than a pure feature engineering approach based in those literatures, as well as greater likelihood of failure.

References

[1] G. Baker-Brown et al. “Coding Manual for Conceptual/Integrative Complexity”. In: *Cambridge University Press* (1992).

[2] L. Carlson and D. Marcu. “Discourse Tagging Reference Manual”. In: *ISI Technical Report* (2001), pp. 30–117.

[3] L. Conway et al. “Automated Integrative Complexity”. In: *Political Psychology* (2014), pp. 603–624.

[4] P. Deane. “On the relation betwrrn automated essay scoring and modern views of the writing construct.” In: *Accessing Writing* 18 (2013), pp. 7–24.

[5] M. G. Hermann. “Assessing leadership style: A trait analysis”. In: *The Psychological assessment of political leaders*. Ed. by Jerrold Post. Ann Arbor: University of Michigan Press, 2005.

[6] S. C. Houck, L. C. Conway, and L. J. Gornick. “Automated Integrative Complexity: Current Challenges and Future Directions”. In: *Political Psychology* 35 (2014), pp. 647–659.

[7] A. Kannan-Ambili. “Automated Scoring of Integrative Complexity Using Machine Learning and Natural Language Processing”. MA thesis. University of Georgia, Dec. 2014.

[8] Q. Le and T. Mikolov. “Distributed Representations of Sentences and Documents”. In: *Proceedings of the 31st International Conference on Machine Learning* (2014), pp. 32–70.

[9] J. Liht and S. Savage. “Preventing Violent Extremism through Value Complexity: Being Muslim Being British”. In: *Journal of Strategic Security* 6.4 (2013), pp. 44–66.

[10] Nitin Madnani et al. “Identifying high-level organizational elements in argumentative discourse”. In: *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics. 2012, pp. 20–28.

[11] J. Mitchell and M. Lapata. “Composition in Distributional Models of Semantics”. In: *Cognitive Science* 34 (2010), pp. 1388–1429.

[12] R. M. Palau and M. Moens. “Argumentation mining: the detection, classification and structure of arguments in text”. In: *Proceedings of the 12th international conference on artificial intelligence and law*. ACM. 2009, pp. 98–107.

[13] F. Pedregosa-Izquierdo. “Feature extraction and supervised learning on fMRI : from practice to theory”. Theses. Université Pierre et Marie Curie - Paris VI, 2015.

[14] J. W. Pennebaker, R. J. Booth, and M. E. Francis. “Linguistic Inquiry and Word Count: LIWC”. In: *Mahwah NJ: Erlbaum Publishers* (2011).

[15] James W Pennebaker and Laura A King. “Linguistic styles: language use as an individual difference.” In: *Journal of personality and social psychology* 77.6 (1999), p. 1296.

[16] I. Persing and V. Ng. “Modeling Argument Strength in Student Essays”. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* 1 (2015), pp. 543–552.

[17] M. D. Shermis, A. Shneyderman, and Y. Attali. “How importatnt is content in the ratings of essay assessments?” In: *Assessments in Education: Principles, Policy and Practice* 15 (2008), pp. 91–105.

[18] C. Stab and I. Gurevych. “Identifying Argumentative Discourse Structures in Persuasive Essays.” In: *EMNLP*. 2014, pp. 46–56.

[19] S. Streufert, P. Suedfeld, and M. J. Driver. “Conceptual structure, information search, and information utilization”. In: *Journal of Personality and Social Psychology* (1965), pp. 736–740.

[20] P. Suedfeld and P. Tetlock. “Integrative complexity of communications in international crises”. In: *Journal of Conflict Resolution* (1977), pp. 169–184.

[21] P. Suedfeld, P. Tetlock, and S. Streufert. “Conceptual/integrative complexity. In C. Smith (Ed.), Handbook of thematic content analysis”. In: *Cambridge University Press* (1992), pp. 393–401.

[22] S. Teufel. “Argumentative zoning: Information extraction from scientific text”. PhD thesis. University of Edinburgh, 1999.

[23] D. A. Winter. “Slot rattling from law enforcement to law-breaking: A personal construct theory exploration of police stress”. In: *International Journal of Personal Construct Psychology* 6 (1993), pp. 253–267.

[24] F. Wolf and E. Gibson. “Representing discourse coherence: A corpus-based study”. In: *Computational Linguistics* 31.2 (2005), pp. 249–287.

[25] F. Wolf and E. Gibson. *The descriptive inadequacy of trees for representing discourse coherence*. MIT. 2003.

[26] A. Yessenalina and C. Cardie. “Compositional matrix-space models for sentiment analysis”. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (2011), pp. 172–182.

[27] M. D. Young and M. G. Hermann. “Increased Complexity Has Its Benefits”. In: *Political Psychology* 35 (2014), pp. 30–43.

Syntactic - Semantic Complexity	Example
Low - Low	Soviet agriculture is a disaster and for an obvious reason. Fifty years ago they collectivized all their farms and made farmers work not for themselves but for the government. Individual incentives were lost. Farmers had to work for the glory of the state. And ever since, the Soviets have not been able to produce enough food to feed their people. This dismal performance will continue as long as the leaders in the Kremlin remain committed to the silly notion that people will work as hard for others as for themselves.
Low - High	Some view abortion as a civil liberties issue; others see abortion as murder. How you view abortion depends on a complicated mixture of legal, moral, philosophical and perhaps scientific judgments. For example, is there a constitutional right to abortion? If there is, what criteria should be used to determine when human life begins? And, a question that must be answered before any of the others can be, who possesses the authority to resolve these issues?
High - Low	Renunciation of thinking is a declaration of spiritual bankruptcy. Where there is no longer a conviction that men can get to know the truth by their own thinking, skepticism begins. Those who work to make our age skeptical in this way, do so in the expectation that, as a result of denouncing all hope of self-discovered truth, men will end by accepting as truth what is forced upon them by authority and by propaganda.
High - High	Their experiences with war and depression during the thirties created in many members of our parents' generation a drive to create some form of security for the future that was not available for them to enjoy in earlier years. By continuously building upon their gradually increasing assets while still maintaining the conservative lifestyles they had been pressed to follow during hard times, they created economic stability for themselves. This economic stability, enjoyed by many approaching old age, lends greater power to seniors' increasingly vocal demands for an improved quality of life for the elderly. Their offspring, not having faced the same hardships as their parents, have had opportunity and cause to be somewhat reflective about issues pertaining to the quality of life in general, including the plight of the elderly.

Table 1: Examples of varying levels of syntactic and semantic complexity, taken from Baker-Brown et al. [1]. Assessments of complexity are based on defensible scores and intend to illustrate the space of integrative complexity.