**COCHRANE — Quantifying**

<1 figure >

**Quantifying Ambiguity by Gamifying the Writing Process: A Case Study on William Blake’s ‘The Sick Rose’**

Cochrane, E. and Milstein, D.

Both the challenge and the pleasure of poetry analysis originate in nocuous and polysemous ambiguity: Different readers interpret linguistic expressions in a poem based on different contexts. The layers of ambiguity grow exponentially when a reader applies semiotic analysis to a poem to examine linguistic expressions for modalities, paradigms, syntagmatic structures, rhetorical tropes, intertexual elements, and semiotic codes. Poetry scholars might argue, in fact, that to eliminate nocuous and polysemous ambiguity would diminish the poem and defeat the purpose of poetic interpretation.

To date, ambiguity analysis generally manifests itself in domains like requirements gathering documents, where ambiguity can lead to misinterpretation, lost money and time, and poor or stunted project outcomes. Approaches that attempt a probabilistic prediction of ambiguity in such documents can be useful for indicating ambiguity when writing these texts, so that authors can be reminded to clarify their writing where it *might* be necessary. In poetry analysis, however, it may be less useful for interpretation purposes to probabilistically predict which expressions are ambiguous or not. Orthography petrifies meaning in poetry, some argue; nevertheless, a poem may be infused with multiple meanings by virtue of ambiguities created through wordplay and rhetorical devices that can be discerned only through close reading and puzzle solving to decode the linguistic expressions and structures.

We are curious to discover if reducing ambiguity will limit poetic analysis and what new types of questions may emerge as a result of disambiguation. To that effect, in this presentation we aim to present three facets of what we hope will become a large project on quantifying ambiguity:

a. Method as Applied to Research in the Humanities: This algorithmic tool will be further refined to provide a mechanism to quantify the ambiguity of a given text by analyzing its lexical and syntactic ambiguity to produce a number that represents the overall ambiguity of the text. As well as gamifying the process of communicating unambiguously (by enabling the dynamic scoring of the author’s progress as she writes), the tool enables the ambiguity of diverse texts to be objectively compared and opens new avenues for text comparison and analysis. Furthermore, the tool should identify ambiguity *within* a text by quantifying the ambiguity of individual words and phrases inside the text. An immediate application of this will be to produce an extension that will highlight terms or phrases in a text in colors representing their ambiguity, providing an ‘ambiguity map,’ the equivalent of a ‘heat map’ (see Chen and Wang, 2012) for ambiguity within texts:

**

Figure 1. Sample output of the tool in current prototype.

b. Critical Assessment of the Application: Application of the tool to William Blake’s ‘The Sick Rose’. Initially, we set our ambitions too high—to apply the tool to what many consider the most difficult poem in the world due to its high level of ambiguity, Mallarmé’s ‘Coup de dés’ (CDD). Obstacles arose because of the structure and variants in the Littré Dictionary (which Mallarmé used to write the poem), which made it difficult to apply the algorithm in its current form. CDD’s length and visual formatting also render it too difficult to parse manually in order to compare it to the algorithm’s output. Therefore, we decided to work with William Blake’s ‘The Sick Rose’, for its brevity, morphological simplicity, and its ambiguity (scholars refer to it as one of Blake’s ‘gnomic triumphs’ [Bloom, 1987]). We anticipate that this 34-word poem will be more manageable both for the algorithm to analyze, so that we can manually code the poem and compare our effort to the algorithmic output, and to build a proof-of-concept multimodal dictionary with word processor.

c. Impact in Formulating and Addressing Research Questions: An outcome of our analysis is the realization that dictionaries (whether those used by the poet or current dictionaries used by readers of poetry) provide definitions that may themselves be ambiguous, primarily due to circularity (for example, the definition of a word in a dictionary may contain a word, the definition for which contains the original word), but also due to the general effect of nocuous or polysemic ambiguity in the dictionaries themselves. Additionally, by hand coding one poem we will show that a computer-aided coder (e.g., a special word processor) would greatly speed up the task of specifying the semantics of a text. Therefore, we intend to build and showcase a multimodal dictionary for the poem, which we will discuss in conjunction with a proposed, longer-term project to build a type of word processor that enables authors to specify the meanings of words in-line while writing a poem or to mark up a poem post-hoc.

As dictionaries often provide multiple meanings for a single word, the proposed new dictionary will provide identifiers for concepts as well as linking terms to those concepts (in much the same way that WordNet does [see Miller, 1995]), enabling authors to use the word processor to specify the concept they are referring to when they use a term or phrase (rather than simply including a multi-meaning term in their text), and minimizing the ambiguity of their writing. The dictionary will require all concepts to be defined using data (images, audio, video, other sensor data), semantic primitives / semantic primes, or some combination of these, thus ensuring that the definitions in the dictionary are minimally ambiguous (Wierzbicka, 1972). In order to eventually enable *semantic* ambiguity to be quantified, the dictionary will also not limit the referrers that may refer to concepts to only terms or phrases but will allow any text to link to a concept. This effort aligns with current ambiguity analysis theory, where authors state that ambiguity can be avoided by using controlled languages (Fuchs and Schwitter, 1996; Gause and Weinberg, 2011), style guides (Kovitz, 1998; Ryser and Glinz, 2000), and lexicons (Chantree et al., 2006) In this section, we also address concerns that creating a lexicon of domain-specific terminology, even for a 34-word poem, will be a sizeable task. If time permits, we hope to solicit feedback from the audience on the proposed word processor, which would enable authors to create texts with minimal ambiguity by enabling them to specify the meanings of words and phrases as they type (in much the same way as predictive text software works). The word processor would also enable deterministic semantic analysis (in contrast to the mainstream approaches to semantic analysis, which are, for the most part, probabilistic [see Hoffman, 2001]) by capturing a machine-readable version of the meaning of the text being written.

**References**

**Blake, W.** (1789). The Sick Rose. In *Songs of Innocence and Experience*.

**Bloom, H.** (ed.). 1987. *Introduction to William Blake’s Songs of Innocence and of Experience*. New York: Chelsea House, pp. 1–28.

**Chantree, F., Kilgarriff, A., de Roeck, A. and Willis, A.** (2005). Disambiguating Coordinations Using Word Distribution Information. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, Borovets, Bulgaria.

**Chantree, F., Nuseibeh, B., de Roeck, A. and Willis, A.** (2006). Identifying Nocuous Ambiguities in Natural Language Requirements. In *Proceedings of the 14th IEEE International Conference*, Minneapolis, Minnesota.

**Chen, L. and Wang, Y.** (2012). Heat Map Displays for Data from Human Abuse Potential Crossover Studies. *Therapeutic Innovation and Regulatory Science,* **46**(6) (November): 701–7.

**Empsom. W.** (2004). *Seven Types of Ambiguity*. Random House.

**Fuchs, N. and Schwitter, R.** (1996). Attempto Controlled English (ACE). In *Proceedings of the First International Workshop on Controlled Language Applications*. Leuven, Belgium.

**Gause, D. and Weinberg, G.** (2011). *Exploring Requirements: Quality before Design*. Dorset House, New York.

**Harrison, A.** (1963). Poetic Ambiguity. *Analysis,* **23**(3) (January).

**Hoffman, T.** (2001). Unsupervised Learning by Probabilistic Latent Semantic Analysis. *Machine Learning,* **42**: 177–96.

**Kovitz, B. L.** (1998). *Practical Software Requirements: A Manual of Content and Style.* Manning, Greenwich, CT.

**Miller, G.** (1995). WordNet: A Lexical Database for English. *Communications of the ACM,* **38**(11): 39–41.

**Norvig, P.** (1998). Interpretation under Ambiguity. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*.

**Poesio, M.** (1996). Semantic Ambiguity and Perceived Ambiguity. In Van Deemter, K. and Peters, S. (eds), *Semantic Ambiguity and Underspecification*. Stanford, CA: Center for the Study of Language and Information. CSLI Lecture Notes 55.

**Resnik, P.** (1999). Semantic Similarity in a Taxonomy: An Information-Based Measure and Its Application to Problems of Ambiguity in Natural Language. *Journal of Artificial Intelligence Research,* **11**: 95–130.

**Rogers, R.** (1973). On the Metapsychology of Poetic Language: Modal Ambiguity. *International Journal of Psychoanalysis,* **54**(1): 61–74.

**Ryser, J. and Glinz, M.** (2000). Scent—A Method Employing Scenarios to Systematically Derive Test Cases for System Test. Technical Report 2000.03, University of Zurich, Switzerland.

**Schull, F., Rus, I. and Basili, V.** (2000). How Perspective-Based Reading Can Improve Requirements Inspections. *IEEE Computer,* **33**(7) (July): 73–79.

**Wasow, T., Perfors, A. and Beaver, D.** (2003). The Puzzle of Ambiguity. In Orgun, O. and Sells, P. (eds). *Morphology and the Web of Grammar: Essays in Memory of Steven G. Lapointe*. CSLI Publications.

**Wierzbicka, A.** (1972). *Semantic Primitives*. Athenäum, Frankfurt.