

The KWIC-step: A Dance for 2 or More

Susan L. Wiesner (dap2sw@surrey.ac.uk)
University of Surrey

This cross-discipline paper considers the use of corpus linguistics as it applies to a study of dance writing.

Argument/Premise

Linguists and Computational Linguistic Engineers who have studied language for special purposes (LSP) generally agree that the lexicon used by those in a specialist discipline offers insight into the concepts and ideologies of that discipline. As Dance is a specialist discipline, it should then follow that the lexicon of dance writing can demonstrate aesthetic concepts and theoretical approaches and beliefs through generalized language patterns detected using KWIC methods. This research supports that contention in that the empirical approach (specifically key words in context), when used against a corpus of 1.4 million words from written dance texts, does offer a means to develop an ontology of the discipline of dance. One ontological example supported by the empirical data is that of both hierarchical and rhizomatic structures of relationships between participants in the dance event. Through this example, this paper will demonstrate both the process and analytical product of using corpus linguistics methods and models to develop an ontology of Dance.

Analytical Method

Due to a dearth of research into dance writing, especially in the area of corpus linguistics, there were no tested, dance-specific models on which to base this study. Therefore, this research contains a variety of methodologies and approaches in the disciplines of corpus linguistics and dance analysis. The corpus methodologies are based on the work of Sinclair (2003), Ahmad (2002), and Biber (1998). Linguistic analyses based on KWIC (key words in context) such as word frequencies, collocations, concordances, and POS tagging were performed on the data generated by computing tools (e.g. System Quirk, Unitex, CLAWS). Additional methods proposed by Traboulsi et al (2003) for identifying a local grammar were used against samples of critical texts (i.e. dance performance reviews). Finally, through a two-pronged approach, the data was analysed

using the analytical modes of description, interpretation, and evaluation proposed for dance by Adshead et al (1988).

After establishing the written language of Dance as a specialist language (or language for special purposes) through the frequency of open class words in the top 100 most frequently used words, weirdness factors, and Z-scores, the data was analysed in a variety of ways. For example, for the purposes of this particular study into ontological relationships between participants, general concepts were identified by focusing on those open class words (OCWs) within the top 100 most frequently used words in a general dance corpus (the Surrey Dance Corpus, or SDC). Finding that the OCWs included several conceptual possibilities which appeared at times to cross conceptual boundaries, it became apparent that a method for categorisation was required to provide distinctions between the concepts. Therefore, the analytical model introduced by Adshead et al (1988) was used for the categorisation. This model calls for a deconstruction of a dance work via description, interpretation, and evaluation based upon identifiable characteristics of the individual work. The identified distinctions include not only the movement and production components, but also concepts of form, concepts through which we view a dance, and concepts particular to a dance. After determining the conceptual categories, the analysis went beyond the top 100 most frequently used words, and considered the relative frequencies of the concept words in the subsequent groups of most frequently used words (top 200, 300, etc.). Word counts also were performed on personal pronouns and conceptual metaphor (using Charteris-Black's method). Finally, collocations and concordances using the concept words as target words were generated and analysed.

Each step was repeated using additional corpora: a group of sub-corpora defined by content of the texts and/or writers' particular activity within the discipline (e.g. critic, scholar, choreographer); and a corpus focused on one choreographer (used as a case study throughout the research).

Conclusion

Through this research into dance writing, Dance can be shown to be a LSP with a lexicon specific to the discipline. So, too, have several unexpected patterns been detected that reflect various concepts and ideologies, one of which is that of the relationships between and among the participants in the dance event (e.g. dancer, choreographer, audience, critic, company). The relational connections are further supported by a general belief in the collective, as shown through word counts and concordances. Additional methods used in dance analysis provide a structure for studying and further clarifying the empirically generated patterns. The choreographic and analytic concepts in Dance as they are perceived and expressed through

written dance texts demonstrate not only the importance of these concepts to specialists in the discipline of Dance, but also the benefit of using the corpus linguistics approach to studying dance writing.

Bibliography

Adshead, J., V. A. Bringinshaw, P. Hodgins, and M. Huxley. Ed. J. Adshead. *Dance Analysis: Theory and Practice*. London: Dance Books, 1988.

Ahmad, K. "The Role of Specialist Terminology Management in Artificial Intelligence and Knowledge Acquisition." *Applications Oriented Terminology Management*. Ed. S. E. Wright and G. Budin. Amsterdam: John Benjamins, 2001. 809-844.

Biber, D, S. Conrad, and R. Reppen. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press, 1998.

Charteris-Black, J. . *Corpus Approaches to Critical Metaphor Analysis*. New York: Palgrave MacMillan, 2004.

Sinclair, J. *Reading Concordances*. London: Pearson Education Limited, 2003.

Trabousli, H., D. Cheng., and K. Ahmad. 2003.