Linking Rhetoric and Methodology in Formal Scientific Writing

Shlomo Argamon (argamon@iit.edu)
Illinois Institute of Technology
Dept. of Computer Science, 10 W 31st Street
Chicago, IL 60616, USA
Jeff Dodick (jdodick@vms.huji.ac.il)
The Hebrew University of Jerusalem
Department of Science Teaching, Givat Ram
Jerusalem 91904, ISRAEL

Abstract

Studying the communication patterns of scientists can give us insight into how science actually works. We argue that methodological differences between different scientific fields should lead to recognizable differences in how scientists in these fields use language to communicate with one another. This paper reports on a corpus-based study of peer-reviewed journal articles in paleontology and physical chemistry which used techniques of computational stylistics to compare the rhetorical styles used in the two fields. We found that indeed the two fields are readily distinguishable based on the stylistic character of their articles. As well, the most significant linguistic features of these distinctive styles can be connected directly to differences posited by philosophers of science between 'historical' (such as paleontology) and 'experimental' (such as physical chemistry) sciences.

Introduction

It has become clear in recent years that communication among different scientists working in a laboratory is critical for scientific success (Dunbar 1995). The particular uses of language by scientists serve to create a "collaborative space", whose worldview makes possible communication about complex observations and hypotheses (Goodwin 1994). Linguistic analysis has also been shown to elucidate features of scientific problem solving, as in Ochs et al.'s (1994) study of physicists' metaphoric talk of travel in a variety of graphical spaces.

At the same time, philosophers of science are increasingly recognizing that the classical model of a single "Scientific Method" (usually based on that of experimental sciences such as physics) does a disservice to sciences such as geology and paleontology, which are no less scientific by virtue of being historically oriented. Instead, it is claimed, differences in method may stem directly from the types of phenomena under study (Cleland, 2002). *Experimental* science (such as physics) attempts to formulate general predictive laws, and so relies heavily on repeatable series of controlled experiments which test hypotheses (Latour & Woolgar 1986). *Historical* science, on the other hand, deals with *contingent* phenomena, studying specific phenomena in the past in an attempt to find unifying explanations for effects caused by those phenomena (Mayr 1976). Because of this, reasoning in historical sciences consists largely of *reconstructive* reasoning, as compared to the *predictive* reasoning from causes to possible effects characteristic of experimental science (Gould 1986; Diamond 1999).

In this paper, we take some first steps towards analyzing the linguistic features of scientific writing in experimental and historical science, using several types of linguistically-motivated document features together with machine learning methods. Our goal is to examine if linguistic features that are indicative of different classes of scientific articles may be usefully correlated with the rhetorical and methodological needs of historical and experimental sciences. This paper describes a corpus-based study of genre variation between articles in a historical science (paleontology) and an experimental science (physical chemistry), with methodological differences as mentioned above. We hypothesize that corresponding rhetorical differences between articles in the respective fields will also be found. Standard methods of computational stylistics were used, confirming this hypothesis. Further, we defined a set of linguistically-motivated features for use in genre classification, based on systemic functional principles. These features enable a more nuanced examination of the rhetorical differences, allowing us to correlate these linguistic differences with the methodological differences posited by philosophers of science.

We note that the work reported here is only a first step, and more extensive studies of larger and more varied corpora of scientific papers will need to be undertaken in order to more firmly determine the links between scientific rhetoric, methodology, and cognition.

Hypotheses

Based on prior work in the philosophy and history of science we thus formulate our main hypothesis:

> ***H1*****:** *Stylistic features will distinguish more strongly between articles from different kinds (historical or experimental) of science than between articles from different journals in the same kind of science.*

We also formulate more detailed hypotheses regarding what sorts of rhetorical features we expect to be most significant in distinguishing articles in the different fields, based on posited methodological differences between historical and experimental sciences, as follows. First, a key element of *historical* reasoning is the need to differentially weight the evidence. Since any given trace of a

past event is typically ambiguous as to its possible causes, many pieces of evidence must be combined in complex ways in order to form a confirming or disconfirming argument for a hypothesis (termed *synthetic* thinking by Baker (1996)). Such thinking is, as Cleland (2002) argues, a necessary commitment of historical science (as opposed to experimental science), due to the fundamental asymmetry of causation. A single cause will often have a great many disparate effects, which if taken together would specify the cause with virtual certainty. Since all the effects cannot actually be known (as some are lost in the historical/geological record), evidence must be carefully weighed to decide between competing hypotheses (the methodology sometimes known as "multiple working hypotheses"). Experimental sciences tend, on the other hand, to adhere more or less to a "predict and test" methodology, in which manipulative experiments are used to confirm or disconfirm specific hypotheses (Cleland 2002). We therefore hypothesize:

***H2a*:** *Writing in historical science has more features expressing the weight, validity, likelihood, or typicality of different assertions or pieces of evidence*

***H2b*:** *Writing in experimental science has more features typical of explicit reasoning about predictions and expectations.*

Note that the presence or absence of linguistic features that can be linked to reasoning of a particular type is not by itself evidence of such reasoning. However, a consistent pattern of many of these features (as shown below) together aligned with the dichotomy proposed in *H2* strongly argues for such differences, which future research will attempt to elucidate in greater detail.

The Corpus

The study reported here was performed using a corpus of recent (2003) articles drawn arbitrarily from four peer-reviewed journals in two fields: *Palaios* and *Quaternary Research* in paleontology, and *Journal of Physical Chemistry A* and *Journal of Physical Chemistry B* in physical chemistry (chosen in part for ease of electronic access). *Palaios* is a general paleontological journal, covering all areas of the field, whereas *Quaternary Research* focuses on work dealing with the quaternary period (from roughly 1.6 million years ago to the present). The two physical chemistry journals are published in tandem but have separate editorial boards and cover different subfields of physical chemistry, specifically: studies on molecules (*J. Phys Chem A*) and studies of materials, surfaces, and interfaces (*J. Phys Chem B*). The numbers of articles used from each journal and their average (preprocessed) lengths in words are given in Table 1.

Table 1. Journals used in the studies with number of articles and average words per article.

Journal	# Art.	Avg. Words
Palaios	116	4584
Quaternary Res.	106	3136
J. Phys. Chem. A	169	2734
J. Phys. Chem. B	69	3301

Study 1: Distinctiveness

Methodology

We first test hypothesis *H1* by testing on our corpus whether paleontological and physical chemistry articles are stylistically distinctive from each other. The method was to represent each document as a numerical vector, each of whose elements is the frequency of a particular lexical feature of the text. We then applied the SMO learning algorithm (Platt 1998) as implemented in the Weka system (Witten & Frank 1999), using a linear kernel, no feature normalization, and the default parameters. (Other options did not appear to improve classification accuracy, so we used the simplest option.) SMO is a support vector machine (SVM) algorithm; SVMs have been previously applied successfully to text categorization problems (Joachims 1998). Generalization accuracy was measured using 20-fold cross-validation[1].

Features

For this first study, we used a set of 546 functions words taken en masse from the stop-word list of the popular research information retrieval system AIRE (Grossman & Frieder 1998); this procedure ensured task and theory neutrality. The set of function words used are similar to those used in many previous studies, such as Mosteller and Wallace's (1964) seminal stylometric work[2]. Each document was thus represented as a vector of 546 numbers between 0 and 1, each the relative frequency of one of the function words.

Results and Discussion

Table 2 shows results for binary classification between each pair of journals in our corpus, giving the percentage of test articles erroneously classified (in 20-fold cross-validation) using linear SMO learning and function-word frequencies as features. We first note that average accuracy on test documents from different fields (historical vs. experimental) was at least 97%, indicating excellent discriminability (far above chance). At the same time, the two physical chemistry journals are quite indistinguishable, as 34% is slightly *greater than* the error of always choosing the majority class (since 69/238=29% of those

[1] In *k*-fold cross-validation (Mitchell 1997) the data is divided into *k* subsets of equal size. Training is performed *k* times, each time leaving out one of the subsets, and then using the omitted subset for testing, to estimate the classification error rate; the average error rate over all *k* runs is reported. This gives quite a stable estimate of the expected error rate of the learning method for the given training size (Goutte 1997).

[2] Relative frequencies of function words, such as prepositions, determiners, and auxiliary verbs, have been shown in a number of studies to be useful for stylistic discrimination, since they act as easily extracted proxies for the frequencies of different syntactic constructs, and also tend not to covary strongly with document topic.

Table 2. Error rates for linear SMO using function word features for pairs of journals using 20-fold cross-validation.

	Historical		Experimental	
	P	***QR***	***PCA***	***PCB***
Palaios	--	10%	**0.4%**	**1%**
Quat Res	10%	--	**2%**	**3%**
Ph Ch A	**0.4%**	**2%**	--	34%
Ph Ch B	**1%**	**3%**	34%	--

articles are from *Phys. Chem. B*). In the case of *Palaios* vs. *Quat. Res.* we get an average error rate of 10%, an order of magnitude higher than any error rate in the cross-disciplinary case. Hence these results support *H1*, in that articles across disciplines are more easily distinguished than articles within a single discipline (from different journals). Of course, the 10% error rate obtained for distinguishing the two paleontology journals is far less than the 48% we would get by majority class classification, which points to a subsidiary distinction between these two journals. This is not unreasonable, given that *Quat. Res.* deals with a specific subset of the topics in *Palaios*[3]. We leave this question, however, for future research.

Study 2: Systemic Variation

Methodology

In order to more precisely analyze the rhetorical differences between articles in the two fields a follow-up study used as features the relative frequencies of sets of keywords and phrases derived from consideration of notions of systemic functional linguistics (Halliday 1994).

Systemic functional linguistics (SFL) construes language as a set of interlocking choices for expressing meanings: "either this, that, or the other", with more general choices constraining the possible specific choices. For example: "A message is either about doing, thinking, or being; if about doing, it is either standalone action or action on something; if action on something it is either creating something or affecting something pre-existent," and so on. A *system* is a set of options for meanings to be expressed, with *entry conditions* denoting when that choice is possible – for example, if a message is not about doing, then there is no choice possible between expressing standalone action or action on something. Each option has also a *realization specification*, giving constraints (lexical, featural, or structural) on statements expressing the option. Options serve as entry conditions for more specific subsystems.

[3] This may be related to the fact that *Quat. Res.* contains more articles than *Palaios* using chemical and radiochemical assaying, since such techniques are only applicable to younger remains from the Quaternary Period; such tools in fact are similar to the experimental techniques seen in physical chemistry. Indeed this is corroborated by the fact that the error rate between Quat Res and the PC journals was higher than Palaios and the same PC journals. More detailed study of the specific articles will be needed to test and refine this hypothesis.

By viewing language as a complex of choices between mutually exclusive options, the systemic approach is particularly appropriate to examining variation in language use. A systemic specification allows us to ask the following type of question: In places where a meaning of general type A is to be expressed in a text (e.g., "a message about action"), what sorts of more specific meanings (e.g., "standalone action" or "action on a thing") are most likely to be expressed by different types of people or in different contexts? A general preference for one or another option, when not dictated by specific content, is indicative of individual or social/contextual factors. Such preferences can be measured by evaluating the relative probabilities of different options by tagging their realizations in a corpus of texts (Halliday 1991).

As features, then, in the absence of a reliable systemic parser, we use keywords and phrases as proxy *indicators* for various systems. For example, an occurrence of the word "certainly" usually indicates that the author is making a high-probability modal assessment of an assertion. The drawback of this approach is lexical ambiguity, since the meaning of such keywords can depend on context. We reduce the effect of ambiguity, however, by using as complete a set of such *systemic indicator* keywords/phrases as possible for each system we represent, and also by using only measures of *comparative* frequency between the aggregated features. In addition, since we use very large sets of indicators for each system, it is unlikely that such ambiguity would introduce a systematic bias, and so such noise is more likely to just reduce the significance of our results instead of biasing them. Preprocessed articles in our corpus were each converted into a vector of 101 feature values (relative frequencies of system options) and the same learning protocol (using SMO) was used as in Study 1.

Features

The systemic features we used are based on options within three main systems, following Matthiessen's (1995) grammar of English, a standard SFL reference. Indicator lists were constructed by starting with the lists of typical words and phrases given by Matthiessen, and expanding them to related words and phrases taken from Roget's Interactive Thesaurus[4] (manually filtered for relevance). Keyword lists were constructed entirely independently of the target corpus. We used systems and subsystems within: CONJUNCTION, linking clauses together (either within or across sentences); MODALITY, giving judgments regarding probability, usuality, inclination, and the like; and COMMENT, expressing modal assessments of attitude or applicability. MODALITY and COMMENT relate directly to how propositions are assessed in evidential reasoning (e.g., for likelihood, typicality, consistency with predictions, etc.), while CONJUNCTION is a primary system by which texts are constructed out of smaller pieces, and so may be expected

[4] http://www.thesaurus.com

Table 3. Average error rates for linear SMO using systemic features for pairs of journals using 20-fold cross-validation.

	Historical		Experimental	
	P	***QR***	***PCA***	***PCB***
Palaios	--	26%	**9%**	**9%**
Quat Res	26%	--	**17%**	**14%**
Ph Ch A	**9%**	**17%**	--	32%
Ph Ch B	**9%**	**14%**	32%	--

Table 4. Strong features (see text) for Paleontology or Physical Chemistry, using SMO.

System	**Hist.**	**Exper.**
CONJUNCTION	Extension	Enhancement
COMMENT	Validative	Predictive
MODALITY/**Type**	Modalization	Modulation
Modalization: Manifestation	Implicit	Explicit
Modulation: Manifestation	Explicit	Implicit

to reflect possible differences in overall rhetorical structure[5]. These systems and the indicators we used are described more fully in the Appendix.

Results and Discussion

We first check inter-class discriminability (*H1*), testing the results of Study 1 above. Table 3 presents classification error rates averaged over 20-fold cross-validation. In all four **cross**-disciplinary cases, error rates are 17% or less, while in the two **intra**-disciplinary cases, accuracy is noticeably lower; *Palaios* and *Quat. Res.* are significantly less distinguishable at 26% error, while *J. Phys. Chem. A* and *J. Phys. Chem. B* are entirely undistinguishable[6]. This further supports hypothesis *H1*, as above. Moreover, consistency with Study 1 results helps to validate the approach taken in this study.

We now consider what consistent picture, if any, emerges of the rhetorical difference between the two classes of scientific articles (paleontology and physical chemistry) from the patterns of feature weights in the learned models. To do this, we ran SMO on the entire corpus (without reserving test data) for each of the four pairs of a paleontology with a physical chemistry journal, and ranked the features according to their weight for one or the other journal in the weight vector. We call a feature *strong*, if it was among the 30 with the highest absolute weights out of 101 features for the same class in models learned for all journal pairs. Among strong features, some striking patterns emerge, shown in Table 4.

[5] Other textual/cohesive systems, such as PROJECTION, TAXIS, THEME, and INFORMATION cannot be easily addressed, if at all, using a key-word-based approach.

[6] Error rates are higher for this feature set than for the function words due to the smaller number of features—clearly there are some stylistic differences that our systemic features do not capture.

First, in COMMENT, we see a preference for Validative comments by paleontologists and one for Predictive comments by physical chemists. This linguistic opposition directly supports both hypotheses *H2a* and *H2b*, related to methodological differences between historical and experimental sciences. As noted, the historically-oriented paleontologist has a rhetorical need to explicitly delineate the scope of validity of different assertions, as part of synthetic thinking (Baker 1996) about complex and ambiguous webs of past causation (Cleland 2002). This is not a primary concern, however, of the experimentally-oriented physical chemist; her main focus is prediction: the predictive strength of a theory and its predictive consistency with the evidence.

Next, we consider the (complicated) system of MODALITY. At the coarse level represented by the simple features, we see a primary opposition in Type. The preference of the (experimental) physical chemist for Modulation (assessing what 'ought' or 'is able' to happen) is consistent with a focus on prediction and manipulation of nature, and supportive of hypothesis *H2b*. The (historical) paleontologist's preference for Modalization (assessing 'likelihood' or 'usuality') is consistent with the outlook of a "neutral observer" who cannot directly manipulate or replicate outcomes, and is thus supportive of hypothesis *H2a*.

This same pattern is also seen within the complex paired features combining values for modality **Type** and **Manifestation**. Implicit variants are more likely to be used for options that are well-integrated into the expected rhetoric, while Explicit realizations are more likely to be used for less characteristic types of modal assessment, as more attention is drawn to them in the text. Keeping this in mind, note that Modalization is preferably Implicit in paleontology but Explicit in physical chemistry; just the reverse holds for Modulation. This shows that Modalization is integrated smoothly into the overall environment of paleontological rhetoric, and similarly Modulation is a part of the rhetorical environment of physical chemistry.

Finally, in the textual system of CONJUNCTION, we see a clear opposition between Extension, indicating paleontology, and Enhancement, indicating physical chemistry. This implies that paleontological text has a higher density of discrete informational items, linked together by extensive conjunctions, whereas in physical chemistry, while there may be fewer information items, each is more likely to have its meaning deepened or qualified by related clauses. This may be indicative that paleontological articles are more likely to be primarily descriptive in nature, requiring a higher information density, while physical chemists focus their attention more deeply on a single phenomenon at a time. At the same time, this linguistic opposition may also reflect differing principles of rhetorical organization: perhaps physical chemists prefer a single coherent 'story line' focused on enhancements of a small number of focal propositions, whereas paleontologists may prefer a multifocal 'landscape' of connected propositions. Future work will

include interviews and surveys of the two types of scientists to investigate these hypotheses.

Related Work

Previous work has investigated the relationship between choice probabilities and contextual factors. For example, Plum & Cowling (1987) demonstrate a relation between speaker social class and choice of verb tense (past/present) in face-to-face interviews. Similarly, Hasan (1998) has shown, in mother-child interactions, that the sex of the child and the family's social class together have a strong influence on several kinds of semantic choice in speech. These previous studies involved hand-coding a corpus for systemic-functional and contextual variables and then comparing how systemic choice probabilities vary with contextual factors via multivariate analysis. By contrast, this study uses large numbers of neutral features and machine learning to automatically build accurate classification models.

Further, by examining differences between systemic preferences across scientific genres, we are quantitatively analyzing differences in register. *Register* denotes functional distinctions in language use related to the context of language use (Eggins & Martin 1997), and may be considered to comprise: *mode*, the communication channel of the discourse; *tenor*, the effect of the social relation between the producer and the audience; and *field*, the domain of discourse. We focus in this paper on the field-related distinction between historical and experimental science, with mode and tenor held relatively constant, by using articles written by working scientists drawn from peer-reviewed journals. Our results indicate that the difference in the types of reasoning needed by historical and experimental sciences leads to correlated differences in rhetorical preferences (perhaps best understood as 'functional tenor' (Gregory 1967)).

Conclusions

We have shown how machine learning techniques together with linguistically-motivated features can be used to provide empirical evidence for rhetorical differences between writing in different scientific fields. Further, by analyzing the models output by the learning procedure, we can see what features realize the differences in register that are correlated with different fields. This provides indirect evidence for methodological variation between the sciences, insofar as rhetorical preferences can be identified which are linked with particular modes of reasoning. This study thus provides empirical evidence for those philosophers of science who argue against a monolithic "scientific method".

Future work will include validating these results against a larger corpus of articles including more scientific fields, as well as incorporating more involved linguistic processing—the rhetorical parsing methods developed by Marcu (2000) are an important step in this direction. Methods for discovering rhetorically important features such as the subjectivity collocations of Wiebe et al. (2001) may also be helpful. Further, the current study treats each article as an indivisible whole. However, as noted by Lewin et al. (2001) in their analysis of social science texts, the rhetorical organization of an article varies in different sections of the text—future work will include studying rhetorical variation across different sections of individual texts, by incorporating techniques such as those of Teufel and Moens (1998).

References

Baker, V.R. (1996). The pragmatic routes of American Quaternary geology and geomorphology. *Geomorphology* **16**, pp. 197-215.

Cleland, C.E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science*.

Diamond, J. (1999). *Guns, Germs, & Steel.* (New York: W. W. Norton and Company).

Dodick, J. T., & N. Orion. (2003). Geology as an Historical Science: Its Perception within Science and the Education System. *Science and Education*, **12**(2).

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R.J. Sternberg, & J. Davidson (Eds.). *Mechanisms of Insight*. (Cambridge MA: MIT Press). pp. 365-395.

Eggins, S. & J. R. Martin, (1997). Genres and registers of discourse. In T. A. van Dijk, *Discourse as structure and process. A multidisciplinary introduction*. Discourse studies 1 (London: Sage), pp. 230–256.

Goodwin, C. (1994). Professional Vision. *American Anthropologist*, **96**(3), pp. 606-633.

Gould, S. J. (1986). Evolution and the Triumph of Homology, or, Why History Matters, *American Scientist* (Jan.-Feb. 1986): pp. 60-69.

Goutte, C. (1997) Note on free lunches and cross-validation, *Neural Computation*, 9(6):1246-9.

Gregory M., (1967). Aspects of varieties differentiation, *Journal of Linguistics* **3**, pp. 177-198.

Grossman, D. and O. Frieder (1998). *Information Retrieval: Algorithms and Heuristics* , Kluwer Academic Publishers.

Halliday, M.A.K. (1991). Corpus linguistics and probabilistic grammar. In Karin Aijmer & Bengt Altenberg (ed.), *English Corpus Linguistics: Studies in honour of Jan Svartvik*. (London: Longman), pp. 30-44.

Halliday, M.A.K. (1994). *An Introduction to Functional Grammar*. (London: Edward Arnold).

Hasan, R. (1988). Language in the process of socialisation: Home and school. In J. Oldenburg, Th. v Leeuwen, & L. Gerot (ed.), *Language and socialisation: Home and school*. North Ryde, N.S.W.: Macquarie University.

Joachims, T. (1998). Text categorization with Support Vector Machines: Learning with many relevant features. In *Machine Learning: ECML-98, Tenth European Conference on Machine Learning*, pp. 137-142.

Latour, B. & S. Woolgar, (1986). *Laboratory Life: The Construction of Scientific Facts* (Princeton: Princeton Univ. Press).

Lewin, B.A., J. Fine, & L. Young (2001). *Expository Discourse: A Genre-Based Approach to Social Science Research Texts* (Continuum).

Marcu, D. (2000). The Rhetorical Parsing of Unrestricted Texts: A Surface-Based Approach. *Comp. Ling.*, **26**(3), pp. 395-448.

Martin, J. R. (1992). *English Text: System and Structure*. (Amsterdam: Benjamins).
Matthiessen, C. (1995). *Lexicogrammatical Cartography: English Systems*. (Tokyo, Taipei & Dallas: International Language Sciences Publishers).
Mayr, E. (1976). *Evolution and the Diversity of Life.* (Cambridge: Harvard University Press).
Mitchell, T. (1997) *Machine Learning.* (McGraw Hill).
Mosteller, F. and Wallace, D. L. (1964). *Inference and Disputed Authorship: The Federalist Papers,* (Reading, MA: Addison Wesley).
Ochs, E., S. Jacoby, & P. Gonzales, (1994). Interpretive journeys: How physicists talk and travel through graphic space, *Configurations* **1**:151-171.
Platt, J. (1998), *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines,* Microsoft Research Technical Report MSR-TR-98-14.
Plum, G. A. & A. Cowling. (1987). Social constraints on grammatical variables: Tense choice in English. In Ross Steele & Terry Threadgold (ed.), *Language topics. Essays in honour of Michael Halliday*. (Amsterdam: Benjamins).
Teufel, S., and Moens, M. (1998). Sentence extraction and rhetorical classification for flexible abstracts. In *Proc. AAAI Spring Symposium on Intelligent Text Summarization*.
Wiebe, J., T. Wilson and M. Bell. (2001). Identifying Collocations for Recognizing Opinions. In *Proc. ACL/EACL '01 Workshop on Collocation, Toulouse, France, July 200.*
Witten, I.H. and Frank E. (1999). *Weka 3: Machine Learning Software in Java:* http://www.cs.waikato.ac.nz/~ml/weka.

Appendix: Systems and Features

CONJUNCTION

On the discourse level, the system of Conjunction serves to link a clause with its textual context, by denoting how the given clause *expands* on some aspect of its preceding context. Similar systems also operate at the lower levels of noun and verbal groups, while denoting similar logico-semantic relationships, e.g., "*and*" usually denotes "additive extension". Options within Conjunction are as follows:

- Elaboration: Deepening the content of the context
  - Appositive: Restatement or exemplification
  - Clarifying: Correcting, summarizing, or refocusing
- Extension: Adding new related information
  - Additive: Adding new content to the context
  - Adversative: Contrasting new information with old
  - Verifying: Adjusting content by new information
- Enhancement: Qualifying the context
  - Matter: What are we talking about
  - Spatiotemporal: Relating context to space/time
    - Simple: Direct spatiotemporal sequencing
    - Complex: More complex relations
  - Manner: How did something occur
  - Causal/Conditional:
    - Causal: Relations of cause and effect
    - Conditional: Logical conditional relations

Note that the actual features by which we represent an article are the frequencies of each subsystem's indicator features, each measured relative to its siblings. So, for example, one feature is Elaboration/*Appositive*, whose value is the total number of occurrences of Appositive indicators divided by the total number of occurrences of Elaboration indicators (Appositive + Clarifying). The relative frequencies of Elaboration, Extension, and Enhancement within Conjunction are also used as features.

COMMENT

The system of Comment is one of modal assessment, comprising a variety of types of "comment" on a message, assessing the writer's attitude towards it, or its validity or evidentiality. Comments are generally realized as adjuncts in a clause (and may appear initially, medially, or finally). Matthiessen (1995), following Halliday (1994), lists eight types of Comment, which we give here along with representative indicators for each such subsystem.

- Admissive: Message is assessed as an admission
- Assertive: Emphasizing the reliability of the message
- Presumptive: Dependence on other assumptions
- Desiderative: Desirability of some content
- Tentative: Assessing the message as tentative
- Validative: Assessing scope of validity
- Evaluative: Judgment of actors behind the content
- Predictive: Coherence with predictions

MODALITY

The features for interpersonal modal assessment that we consider here are based on Halliday's (1994) analysis of the Modality system, as formulated by Matthiessen (1995). In this scheme, modal assessment is realized by a simultaneous choice of options within four systems[7]:

- Type: What kind of modality?
  - Modalization: How 'typical' is it?
    - Probability: How likely is it?
    - Usuality: How frequent/common is it?
  - Modulation: Will someone do it?
    - Readiness: How ready are they (am I)?
    - Obligation: Must I (they)?
- Value: What degree of the relevant modality scale?
  - Median: In the middle of the normal range.
  - High: More than normal
  - Low: Less than normal
- Orientation: Is the modality expressed as an Objective attribute of the clause or as Subjective to the writer?
- Manifestation: Is the assessment Implicitly realized by an adjunct or finite verb, or Explicitly by a projective clause?

The cross-product of these subsystems gives many modality assessment types, each realized through a subset of indicators. *Simple* features are each option in each system above (e.g., Modalization/*Probability* opposed to Modalization/*Usuality*), while *complex* features are pairwise combinations of such simple features. The indicator set for each such feature is the intersection of the indicator sets for the two component features. Frequencies were normalized by the total set of occurrences of both primary systems (Modalization and Value in the previous example).

[7] Note that we did not consider the system of POLARITY, since it cannot be properly addressed without more sophisticated parsing.