The Languages of Science: A Corpus-Based Study of Experimental and Historical Science Articles

Shlomo Argamon (argamon@iit.edu)
Paul Chase (chaspau@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 works in practice. We have been studying the language of peer-reviewed journal articles in several scientific fields, and argue that methodological differences between different scientific fields are reflected in differences in how scientists in these fields rhetorically organize their texts. Techniques of computational stylistics enable us to compare rhetorical styles between the *historical* and the *experimental* sciences, and thus relate differences in language use to methodological differences posited by philosophers of science. A previous study (Argamon & Dodick 2004) demonstrated a clear difference between writing style in paleontology and that in physical chemistry; this paper extends those results to peer-reviewed articles from twelve journals in six different fields. Our current results demonstrate more generally that historical science writing is readily distinguishable from that of experimental science. Moreover, the most significant linguistic features of these distinctive styles are directly related to differences posited by philosophers of science between the historical and experimental sciences.

Introduction

It has become clear in recent years that communication among scientists is critical for scientific success (Latour & Woolgar 1986; Dunbar 2001; Goodwin 1994). Key to communication, of course, is effective use of language, so it is unsurprising that idiosyncratic features of scientific prose have evolved to be useful in constructing complex scientific arguments (Halliday & Martin 1993). The way scientific language is organized, then, should reflect scientific reasoning and methodology.

Recently, philosophers of science have begun to recognize that the classical model of a single "Scientific Method" is overly simplistic, but rather, different scientific methods may be used in different fields with varying epistemological needs (Cleland, 2002). *Experimental* science (such as physics) attempts to formulate general predictive laws, and so relies heavily on repeatable series of controlled experiments that test specific hypotheses (Diamond 2002). *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 2002).

Our contention is that such methodological differences should lead to discernable differences in language usage; analysis of such differences will both provide empirical support for the "multiple methodologies" position (as opposed to a monolithic "scientific method"), as well as possibly enable a more nuanced examination of those methodological differences.

Overview

We have been applying techniques of computational stylistics to this problem. In this paper, we generalize our earlier results (Argamon & Dodick 2004), analyzing the linguistic features of scientific writing in experimental and historical science as exemplified in peer-reviewed articles in a number of journals in various fields. The larger corpus also enables a more detailed analysis of the features that distinguish the two classes of scientific articles. We apply stylistic text categorization methods, as in our earlier study, to determine if the article classes are significantly distinctive from each other, and what features distinguish them. (We only consider topic-*independent* (or `stylistic') features, since each field will be easily (and trivially) distinguished by its own jargon, such as `fossil' in paleontology, or `quantum entanglement' in physics.)

To summarize, we find that differences in language style between articles in different fields do exist, corresponding in broad outline to differences in methodology between experimental and historical science. Further, a set of linguistically well-motivated features for genre classification (based on Systemic Functional Linguistics) enable a nuanced examination of the rhetorical differences, demonstrating a clear connection between linguistic differences classes of articles and methodological differences between kinds of science, as posited by philosophers of science.

Our work in this vein can be seen as a computational extension of previous work examining the relationship between linguistic choice 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 (1988) has shown, in mother-child interactions, that the child's sex and the family's social class together have a strong influence on several kinds of semantic choice in speech.

Hypotheses

Based on the above-mentioned work in the philosophy and history of science we 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 consider more detailed hypotheses regarding what sorts of stylistic features we expect to be most significant in distinguishing articles in the different fields, based on posited methodological differences. 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 (Cleland 2002; Baker 1996). Since all the effects of any given past cause cannot actually be known (as some are lost in the historical/geological record), evidence must be carefully weighed to decide between competing 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 predictions formulated as 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.*

Further, historical science primarily studies complex and unique entities via observation and comparison, while experimental science studies uniform and interchangeable entities via manipulative experimentation. The consequent manipulative nature of information gathering in the experimental sciences will tend towards highly focused and controlled experimentation (indeed, this is considered a key quality of good methodology). Historical scientific methodology differs here, in that (a) the context of uncontrolled observations may be critical for proper interpretation, and (b) that the uniqueness of the objects studied means that comparison among many different observations is critical.

Argumentation in the historical sciences thus may be expected to require a higher density of separate `information units' than the more focused argumentation we expect in the experimental sciences. Thus we may also hypothesize:

H3a: *Writing in historical science has features indicating a greater variety of information units in the text, as well as a concern with contextualizing information about varied and unique objects of study; this reflects the methodology of uncontrolled observations on unique historical entities.*

H3b: *Writing in experimental science has features indicating more focused attention to a single (or small number of) `storylines', reflecting a more focused approach of controlled manipulation of nature to test specific hypotheses about uniform objects.*

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

Journal	#Art	Avg. Words
J. Geology	93	4891
J. Metamorphic Geol.	108	5024
Biol. J. Linnean Society	191	4895
Human Evolution	169	4223
Palaeontologia Electronica	111	4132
Quaternary Research	113	2939
Physics Letters A	132	2339
Physical Review Letters	114	2545
J. Physical Chemistry A	121	4865
J. Physical Chemistry B	71	5269
Heterocycles	231	3580
Tetrahedron	151	5057

Note that the presence or absence of one or two 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.

Methodology

The Corpus

The study reported here was performed using a corpus of recent (2003) articles drawn from twelve peer-reviewed journals in six fields, as given in Table 1. This triples the number of journals and articles surveyed when compared with our previous study (Argamon and Dodick, 2004).

Textual Features

First, we used a set of 546 functions words (termed FW) 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 on the Federalist Papers[1]. 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.

In order to more precisely analyze the rhetorical differences between articles in the two fields a follow-up experiment used as features the relative frequencies of sets of keywords and phrases derived from consideration of notions of Systemic Functional Linguistics (SFL) (Halliday 1994). This theory 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: "If clause A is to be extended by clause B, then clause B may *extend* A by adding new information, *elaborate* on the meaning already in A, or *enhance* understanding of A by qualifying circumstances; if B is enhancing, then it may do so with regards to spatiotemporal circumstances, manner, or causes/conditions," 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. Options serve as entry conditions for more specific subsystems.

By viewing language as a complex of choices among 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., "one clause expands on another"), what sorts of more specific meanings (e.g., "extension" or "enhancement") are most likely to be expressed in different contexts? 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. We use 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. The use of large sets of indicators for each system makes it unlikely that ambiguity will introduce a systematic bias.

We focus on two large-scale aspects of a text related to its rhetorical organization which enable it to realize complex conceptual structures; for space reasons we only give a brief overview here (details in the Appendix; also see (Argamon and Dodick 2004)).

Cohesion refers to linguistic resources that enable language to connect to its larger context, both textual and extratextual (Halliday & Hasan 1976). Such resources include a wide variety of referential modalities (pronominal reference, deictic expressions, ellipsis, and more), as well as lexical repetition and variation, and different ways of linking clauses together. How an author uses these various cohesive resources is an indication of how the author organizes concepts and relates them to each other. Within cohesion, our computational work considers just conjunctions, for feasibility of automated extraction.

In SFL, clausal conjunctions are organized via the system of EXPANSION, which describes features which serve to link clauses together causally or logically. The first of the three main types of EXPANSION is Extension, which links clauses giving different information together, realized by words such as ``and'', ``but'', and ``furthermore''. The second type is ENHANCEMENT, which qualifies information in one clause by another (e.g., ``similarly...'' or ``therefore...''). Third is ELABORATION, which deepens a clause by clarification or exemplification (e.g., ``in other words...'' or ``more precisely'').

Assessment we define as qualification of events or assertions in the text according to their rhetorical or epistemic properties. One key system for assessment is MODALITY in which the likelihood, typicality, or necessity of an event are indicated, usually by a modal auxiliary verb or an adjunct adverbial group. There are two main types of modality: MODALIZATION, which quantifies levels of likelihood or frequency (e.g., ``probably'', ``might'', ``usually'', ``seldom''), and MODULATION, which quantifies ability or necessity of performance (e.g., ``ought to...'', ``should...'', ``allows...'').

A second important system for assessment is COMMENT, which 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).

Machine Learning

Each document was represented by two numerical vectors, one for each of the above feature sets. In each vector, the elements were the relative frequency of the lexical features of the text. We 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 10-fold cross-validation.

[1] 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.

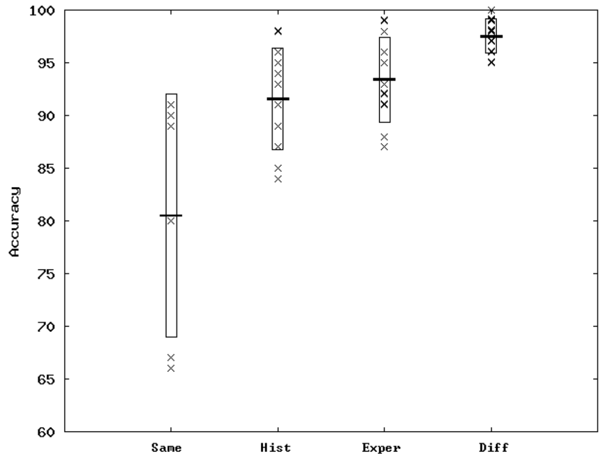

Figure 1: Learning accuracy for distinguishing articles in different pairs of journals using FW features. (Full explanation of this figure appears in text).

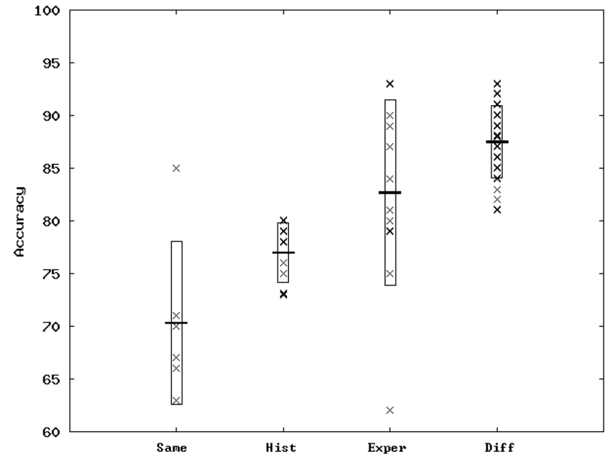

Figure 2: Learning accuracy for distinguishing articles in different pairs of journals using SC features. (Full explanation of this figure appears in text).

Results

Distinctiveness

We first test hypothesis $\mathbf{H_1}$ by testing on our corpus whether experimental and historical journal articles are stylistically distinctive from each other. For each pair of two different journals, we used SMO (as above) to learn classifiers distinguishing articles in one journal from those in the other; the estimated learning accuracy is a measure of how distinguishable the journals are from one another (in terms of the features we have defined). These pairwise classification results are shown for feature sets FW and SC in Figures 1 and 2, respectively. Note that journal pairs from a single field are the least distinguishable, followed by those from two different (out of three) historical fields, then from two different experimental fields, and finally, most distinctive are those journal pairs where one journal is historical and the other experimental. While there is some variance within each group, the difference between the ``Diff'' pairs and the pool of all ``Similar'' pairs (i.e., the union of ``Same'', ``Exper'' and ``Hist''), for both feature sets, is quite noticeable and statistically significant (FW: Mann-Whitney U=940.5, p<0.0001; SC: Mann-Whitney U=915.5, p<0.0001).

Linguistic characterization

We now consider what consistent picture, if any, emerges of the linguistic differences between the two classes of scientific articles (historical and experimental) 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 thirty-six pairs of a historical with an experimental journal, and ranked the features according to their weight for one or the other journal in the weight vector. The sign of a feature's weight (positive or negative) denotes which class of documents it indicates. To simplify presentation of the results, we will call a feature *consistently indicative*, if it was indicative at all for one class of document in at least 25 of the 36 trials.

Table 2 shows the consistently indicative features found. Most important for understanding these results are **oppositions**, in which an option in a particular system is strongly indicative of one article class (either experimental or historical science) while a different option of that same system is indicative of the other class. Such an opposition indicates a meaningful linguistic difference between the classes of articles, in that each prefers a distinctive way (its preferred option) of expressing the same general meaning (the parent system). We now discuss the most salient and how they relate to our hypotheses.

First, consider the system of EXPANSION. At the top level, we see an opposition between Extension, which indicates historical science, and Enhancement, which indicates experimental science. This implies that historical science articles generally have a higher density of discrete informational items, whereas experimental science articles tend to have fewer information items, which may have their meaning deepened or qualified by informationally related clauses. This appears to reflect differing principles of rhetorical organization—experimental scientists seem to prefer a single coherent `story line' focused on enhancements of a small number of focal propositions, whereas historical scientists seem to prefer a multifocal `landscape' of connected propositions. This supports hypotheses $\mathbf{H3_a}$ and $\mathbf{H3_b}$, respectively, comparing the contextual examination of various and highly unique entities by historical science with the more universalist, hence focused, examination of generic entities by experimental science.

Two other oppositions are visible in the `subsystems' of Extension and Enhancement as well, which, while not bearing directly on our hypotheses, seem related. Within Extension, we see an experimental preference for Additive, compared with a historical preference for Adversa-

System	*Historical*	*Experimental*
Expansion	Extension(26)	Enhancement(31)
Elaboration		Apposition(28)
Extension	Adversative(30)	Additive(26)
Enhancement	Matter(29)	Spatiotemporal(26)
Comment	Admissive(30) Validative(32)	Predictive(36)
Modality Type	Modalization(36)	Modulation(35)
Modulation	Obligation(29)	Readiness(26)
Modality Value		High(27)
Modality Orientation	Objective(31)	Subjective(31)

Table 2. Consistent indicator features within each of the systems used in the study. Numbers in parentheses show in how many paired-classification tests the feature names was an indicator for the given class of documents.

tive. This further supports the notion that historical scientists use a ``multiple working hypotheses" method, in which comparison of alternatives is key (**H3$_a$**). Further, we see that SpatioTemporal relationships form a core type of Enhancement for experimental scientists (presumably describing forms and relationships of experimental subjects and results), while historical scientists do Enhancement more via Matter and so contextualize their statements more (**H2$_a$**).

Posited methodological distinctions between the kinds of science are further supported by our results for COMMENT. Here we see preference for Validative and Admissive comments by historical scientists compared to a very strong consistent preference for Predictive comments by experimental scientists. The latter result is a straightforward consequence (hypothesis **H2$_b$**) of the experimental scientist's focus on predictive consistency. The historical scientist, on the other hand, evinces a rhetorical need (via Validation comments) to explicitly delineate the scope of validity of different assertions (hypothesis **H2$_a$**), likely due to synthetic thinking (Baker 1996) about complex and ambiguous webs of past causation (Cleland 2002). An Admissive comment marks a clause as the opinion (perhaps strongly held) of the author; this too appears indicative of a more hedged and explicitly comparative argumentation style.

Finally, we consider MODALITY, starting with simple (non-conjunctive) features. The primary opposition is in modality Type. The preference of experimental science writing for Modulation (assessing what `must' or `is able' to happen) is supportive of hypothesis **H2$_b$** and consistent with a focus on prediction and manipulation of nature. Concurrently, historical science writing shows a preference for Modalization (assessing `likelihood' or `usuality'), support hypothesis **H2$_a$** and consistent with the outlook of an observer who usually cannot directly manipulate or replicate outcomes, and therefore (i) cannot make unqualified statements of what must (or must not) happen, and (ii) uses therefore the method of ``multiple working hypotheses".

Within Modulation, we see also a more delicate opposition, where experimental science articles prefer Readiness (possibility), and historical science articles prefer Obligation (necessity). To understand this, note first that either kind of assessment opens up a clause for value-based consideration; if a statement is to be simply accepted, on the other hand, modal assessment will typically be avoided (compare ``John went to the store" with ``John should have gone to the store"; the first is either true or false, whereas the second makes a value judgment that can be disputed). Thus experimental scientists open Readiness up for consideration and possible disputation by the reader, while the default is for events to be necessary (and so need not be noted explicitly as such). For historical scientists, on the other hand, necessity cannot be assumed, rather Obligation is noted openly in the text, allowing dispute, while the possible and contingent is assumed as the default (and nuanced levels of probability are expressed by Modalization, as noted above).

Conclusions

Our study of the languages of science provides two intertwined results. First, application of machine learning techniques to linguistically motivated textual features provides empirical evidence for rhetorical differences between writing in historical and experimental sciences, thus supporting recent claims of philosophers against a monolithic "scientific method". Second, analysis of the models output by this technique gives insight into which language features most consistently realize the difference in functional text type corresponding to different types of science. Such linguistic preferences are directly linked with the particular modes of reasoning posited by philosophers for these different kinds of science, pointing towards a more nuanced analysis of methodological variation among different scientific fields while lending further weight to the argument for a multiplicity of scientific methods.

Future work includes validating these results against a larger corpus of articles including more scientific fields. Moreover, we intend to study linguistic variation across different sections of individual texts within and across fields, as the stylistic and discourse-structural character of an article varies strongly among its different sections (e.g., Introduction vs. Methods vs. Results), as analyzed by Lewin et al. (1986) for social science texts.

This work may also have meaningful applications towards improving science education. By delineating some of the language skills necessary for effective scientific communication, it might be possible to develop methods for teaching such skills to scientific novices.

References

Argamon, S., and Dodick, J.T. 2004. Linking rhetoric and methodology in formal scientific writing. In Proc. 26th Annual Cognitive Science Society meetings, Chicago IL.

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. (2002). *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. (2001). What scientific thinking reveals about the nature of cognition. In *Designing for Science* (pp. 115–140). Mahwah, NJ: Lawrence Erlbaum Associates.

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.: pp. 60-69.

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).

Halliday, M. A. K., & R. Hasan. (1976). *Cohesion in english*. London: Longman.

Halliday, M. A. K., & J.R. Martin. (1993). *Writing science: Literacy and discursive power*. London: Falmer

Hasan, R. (1998). 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 *ECML-98, 10th 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 (1986). *Expository discourse: A genre based approach to social science research texts*. Continuum

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).

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

EXPANSION

On the discourse level, the system of Expansion serves to link a clause with its textual context, by denoting how the given clause *expands* on some aspect of its preceding context. Options within Expansion 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

COMMENT

The system of Comment comprises 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 (1992), following Halliday (1994), lists the following types of Comment:

- *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

Modal assessment per Halliday (1994) is realized by a simultaneous choice of options within four systems[2]:

- *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?

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