

# WHAT MAKES LANGUAGE FORMAL?

Eduard H. Hovy <sup>1</sup>

Information Sciences Institute  
of the University of Southern California  
4676 Admiralty Way  
Marina del Rey, CA 90292-6695  
Telephone: 213-822-1511  
ARPA address: HOVY@VAXA.ISI.EDU

Topic Area: Natural Language Generation, Text Planning

## Abstract

This paper addresses part of the question “how do we say the same thing in different ways in order to communicate non-literal, pragmatic information?”. Since the style of the text can communicate much information — it may be stuffy, slangy, prissy — generators that seek to satisfy pragmatic, hearer-related goals in addition to simple informative ones must have rules that control how and when different styles are used. But what is “style”? In this paper, formal and informal language is analyzed to provide stylistic rules that enable a program to produce texts of various levels of formality.

## 1 Introduction

When we produce language, we tailor our text to the hearer and to the situation. This enables us to communicate more information than is contained in the literal meanings of our words; indeed, the additional information often has a stronger effect on the hearer than the literal content has. This information is carried by both the content and the form of the text. As speakers and hearers, we associate various interpretations of the speaker, his goals, the hearer, and the conversational circumstances, with the various ways of expressing a single underlying fact or idea.

The level of formality of text is one of the strongest carriers of additional information. This level reflects the level of formality of the conversational setting (for instance, a burial or a party) and of the interpersonal distance between the interlocutors. But what does it mean for language to “seem relaxed” or to “be formal”? No single item in the language defines the level of formality; rather, text seems to contain a number of little clues that cumulatively create a certain impression. What are these little clues? Where do they appear in language and how do we decide to use them?

To answer this question, handbooks of writing are of little use: typically, they describe styles in terms of the characteristics of complete paragraphs of text (see, say, [Birk & Birk 65] and [Hill 1892]), which is not useful for a practical, generator-oriented approach. Instead, a functional approach is to describe styles in terms of the decisions a generator has to make: decisions such as sentence content, clause order and content, and word selection.

---

<sup>1</sup>This paper was written while the author was at Yale University Computer Science Department, 2158 Yale Station, New Haven, CT 06520-2158, U.S.A. This work was supported in part by the Advanced Research Projects Agency monitored by the Office of Naval Research under contract N00014-82-K-0149. The work was also supported by AFOSR contract F49620-87-C-0005.

## 2 Formality

The level of textual formality is probably the pre-eminent stylistic aspect; it comes into play along the whole range of generator decisions (from the initial sentence topic selection and organization down to the final word selection). All language users have rules for making their text more or less formal. The best way to illustrate these rules is to dissect a piece of text:

*Yesterday, December 7, 1941 — a date which will live in infamy — the United States of America was suddenly and deliberately attacked by naval and air forces of the Empire of Japan.*

*The United States was at peace with that nation and, at the solicitation of Japan, was still in conversation with its Government and its Emperor looking forward to the maintenance of peace in the Pacific.*

*Indeed, one hour after Japanese air squadrons had commenced bombing Oahu, the Japanese Ambassador to the United States and his colleague delivered to the Secretary of State a formal reply to a recent American message. While this reply stated that it seemed useless to continue the existing diplomatic negotiations, it contained no threat or hint of war or armed attack.*

*It will be recorded that the distance of Hawaii from Japan makes it obvious that the attack was deliberately planned many days or even weeks ago. During the intervening time, the Japanese Government has deliberately sought to deceive the United States by false statements and expressions of hope for continued peace.*

[“We Will Gain the Inevitable Triumph — So Help Us God”, war address by F.D. Roosevelt to joint session of Congress of the United States, December 8, 1941.]

What characteristics make this address formal? Certainly, one factor is the use of formal verbs and nouns instead of more common ones, such as “solicitation” instead of “request”. Another factor is the use of full names and titles instead of their common abbreviations. Accordingly, we replace words and phrases in the address by less formal equivalents ((a) below) and use the common names for entities (b).

The result, however, is definitely not informal. The sentences still seem long and involved. In order to simplify them, we (c) remove conjunctions and multi-predicate phrases, and (d) remove adverbial clauses, or place them toward the ends of sentences. Now, however, the text seems odd; for example, phrases such as “it will be recorded” do not blend with phrases such as “deliberately tried to cheat” (introduced by (a)). To improve this, we (e) eliminate the use of passive voice, and (f) refer to the involved parties — speaker, hearer, and others — directly.

Now some phrases sound flowery and out of place. To simplify, some nominalized verbs can be converted to verbs (g); noun groups can be simplified by dropping redundant adjectives and nouns (h); pronominalization can be increased (i). Finally, a few finishing touches: simplified tenses (j); colloquial phrases (k); complete elision of redundant words where grammatical (l):

We were<sub>(f)</sub> suddenly and deliberately attacked by naval and air forces of Japan<sub>(b)</sub> yesterday, December 7, 1941<sub>(a)</sub>. We’ll never forget this date<sub>(c,d,k,l)</sub>.

We were<sub>(f)</sub> at peace with them<sub>(i)</sub>. [and,]<sub>(c)</sub> At Japan’s request<sub>(a,h)</sub> we were<sub>(f)</sub> still talking to<sub>(a)</sub> their<sub>(e)</sub> Government. [and its Emperor.]<sub>(h)</sub> We were<sub>(f)</sub> looking forward to having<sub>(a,g)</sub> peace in the Pacific.

[Indeed]<sub>(i)</sub> One hour after Japanese air squadrons [had]<sub>(j)</sub> started<sub>(a)</sub> bombing Oahu, their Ambassador<sub>(a)</sub> [and his colleague]<sub>(i)</sub> gave<sub>(a)</sub> our<sub>(f)</sub> Secretary of State a formal reply to a recent message. [While]<sub>(c)</sub> [This reply said<sub>(a)</sub> that]<sub>(i)</sub> They<sub>(f,i)</sub> thought it was<sub>(e)</sub> useless to continue negotiating<sub>(a)</sub>. [there was<sub>(a)</sub>]<sub>(c)</sub> But they<sub>(i)</sub> didn't<sub>(k)</sub> [threaten or]<sub>(h)</sub> talk about<sub>(a)</sub> war. [or armed attack.]<sub>(h)</sub>

[Note<sub>(e,f)</sub> that]<sub>(i)</sub> The distance of Hawaii from Japan makes it obvious that they<sub>(f)</sub> deliberately planned<sub>(e)</sub> the attack a while<sub>(k)</sub> [or even weeks]<sub>(i)</sub> ago. [In<sub>(a)</sub> the intervening time,]<sub>(d)</sub> The Japanese Government [has]<sub>(j)</sub> deliberately tried<sub>(a)</sub> to cheat<sub>(a)</sub> us<sub>(f)</sub> by [false statements and]<sub>(h)</sub> pretending<sub>(a)</sub> [expressions of hope for continued]<sub>(i)</sub> to hope for peace in the mean time<sub>(k)</sub>.

### 3 Rules for Creating Formal Text

A number of texts, ranging from politicians' speeches and writings to discussions with friends, were analyzed in the manner above. The transformation steps were stated as rules that provide criteria by which PAULINE<sup>2</sup> makes appropriate choices at decision points. One of the program's rhetorical goals, the goal controlling formality, takes one of the values *highfalutin*, *normal*, *colloquial*. In order to make text more formal, the program examines its options at decision points and applies the strategies paraphrased here:

- **topic inclusion:** to make long sentences, select options that contain causal, temporal, or other relations to other sentence topics
- **topic organization:** to make complex sentences, select options that are subordinated in relative clauses; that conjoin two or more sentence topics; that are juxtaposed into relations and multi-predicate enhancer and mitigator phrases
- **sentence organization:** make sentence seem weighty by including many adverbial clauses; by placing these clauses toward the beginnings of sentences; by building parallel clauses within sentences; by using passive voice; by using more "complex" tenses such as the perfect tenses; by avoiding ellipsis, even though it may be grammatical (such as "Joe got more than Pete [did]", "When [I was] 20 years old, I got married")
- **clause organization:** make weighty, formal clauses, by including many adjectives and adjectival clauses in noun groups; by doubling nouns in noun groups ("Government and Emperor", "statements and expressions"); by including many adverbs and stress words in predicates; by using long, formal phrases; by nominalizing verbs and adverbs ("their flight circled the tree" instead of "they flew round the tree"); by pronominalizing where possible; by not referring directly to the interlocutors or the setting
- **phrase/word choice:** select formal phrases and words; avoid doubtful grammar, slang, and contractions (say "man" rather than "guy" and "cannot" rather than "can't")

In contrast, by following inverted strategies, PAULINE makes its text less formal.

<sup>2</sup>PAULINE (Planning And Uttering Language In Natural Environments) is a generator program that produces various texts from a single story representation under various settings that model pragmatic circumstances. PAULINE consists of over 12,000 lines of T, a Scheme-like dialect of LISP developed at Yale University. It generates over 100 variations of a description of an episode that occurred at Yale in April 1986 (see below and [Hovy 87a, 87b]), as well as different versions of texts in two other domains (see [Hovy 86b], [Bain 86], and [Hovy 86a]).

## 4 Determining Appropriate Levels of Formality

Knowing how to make formal text is not enough. The generator must also know when it is appropriate. Since the level of formality is not actually measurable, it is most apparent only when the level is suddenly changed or is inappropriate. In order to determine the pragmatic effects of formality, then, the important question is: *what does the speaker achieve by altering the level of formality?*

First, if you become less formal, you signal a perceived or desired decrease in the interpersonal distance between yourself and the hearer. In any relationship, the participants maintain a certain distance (say, from intimate to aloof) which is mirrored by a corresponding level of formality. Which interpersonal distance corresponds to which level of formality depends, of course, on social convention and on the interlocutors and their relationship; for example, colloquial or informal language is often used to discuss relatively intimate topics, and more formal language often indicates that you feel, or wish to feel, more distant than the conversation had been implying (perhaps after you are offended or become uncomfortable with the topic). See [Brown & Levinson 78] on the use of formal honorifics and [Kuno 73] and [Harada 76] on Japanese deictic honorifics.

Second, if you alter the level of textual formality, you may perturb the tone or atmosphere of the conversation. Since the conversational atmosphere is also mirrored by textual formality, a serious conversation (a burial speech or a conference talk) requires more formality than an everyday conversation (a report to the family of the day's events). An inappropriate level of formality can affect the hearer's emotion toward you: if you are too informal, you may seem cheeky or irreverent; if you are too distant, you may seem snooty or cold. A large amount of work by sociologists, anthropologists, and psycholinguists describes the characteristics of various settings and the appropriate levels of formality in various cultures (see, for example, [Irvine 79] and [Atkinson 82] on formal events; [Goody 78] and [R. Lakoff 77] on politeness).

Based on these considerations, after PAULINE is given values for the parameters that characterize the conversational setting, the speaker, and the hearer (in boldface), it uses the following rules to activate its rhetorical goal of formality:

1. set the rhetorical goal of formality to
  - *colloquial* when the **depth of acquaintance** is marked *friends*, or when the **relative social status** is marked *equals* in an atmosphere (tone) marked *informal*
  - *normal* when the **depth of acquaintance** is marked *acquaintances*
  - *highfalutin* when the **depth of acquaintance** is marked *strangers*
2. then, reset the goal value one step toward *colloquial* if **desired effect on interpersonal distance** is marked *close* or if **tone** is marked *informal*
3. or reset the goal value one step toward *highfalutin* if **desired effect on interpersonal distance** is marked *distant* or if **tone** is marked *formal*
4. and invert the value if **desired effect on hearer's emotion toward speaker** is marked *dislike* or if **desired effect on hearer's emotional state** is marked *angry*

## 5 The Rules at Work

PAULINE uses these rules to produce the following two texts when it is being *highfalutin* (say, writing for a newspaper) and *colloquial* (say, talking to a friend). (This episode is represented in a property-inheritance network such as described in [Charniak, Riesbeck & McDermott 80] using elements based on Conceptual Dependency (see [Schank 82] and [Schank & Abelson 77]). Approximately 130 representation elements denote the events, actors, locations, props, and their relationships (temporal, intergoal, causal, etc.).)

HIGHFALUTIN	COLLOQUIAL	Decision Type
[IN EARLY APRIL], A SHANTYTOWN -- NAMED WINNIE MANDELA CITY -- [WAS [ERECTED] BY] [SEVERAL] STUDENTS ON BEINECKE PLAZA,  [SO THAT]  YALE UNIVERSITY WOULD [DIVEST FROM] COMPANIES DOING BUSINESS IN SOUTH AFRICA.  [LATER, AT 5:30 AM ON APRIL 14], THE SHANTYTOWN [WAS DESTROYED] BY OFFICIALS; [ALSO, AT THAT TIME.] THE POLICE ARRESTED 76 STUDENTS.  SEVERAL LOCAL POLITICIANS AND FACULTY MEMBERS [EXPRESSED CRITICISM] OF [YALE'S] ACTION.  [FINALLY], YALE [GAVE] THE STUDENTS [PERMISSION] TO [REASSEMBLE] THE SHANTYTOWN THERE [AND, CONCURRENTLY], THE UNIVERSITY [ANNOUNCED] THAT A COMMISSION WOULD GO TO SOUTH AFRICA IN JULY TO [INVESTIGATE] THE SYSTEM OF APARTHEID.	[ ] STUDENTS [PUT] A SHANTYTOWN, [ ] WINNIE MANDELA CITY, UP  [ON BEINECKE PLAZA] [IN EARLY APRIL].  THE STUDENTS WANTED YALE UNIVERSITY TO [PULL THEIR MONEY OUT OF] COMPANIES DOING BUSINESS IN SOUTH AFRICA.  [ ] OFFICIALS [TORE [IT] DOWN] AT 5:30 AM ON APRIL 14, [AND] THE POLICE ARRESTED 76 STUDENTS.  SEVERAL LOCAL POLITICIANS AND FACULTY MEMBERS [CRITICIZED] THE [ ] ACTION.  [LATER,] YALE [ALLOWED] THE STUDENTS TO [PUT [IT] UP] THERE [AGAIN]. [ ] THE UNIVERSITY [SAID] THAT A COMMISSION WOULD GO TO SOUTH AFRICA IN JULY TO [STUDY] THE SYSTEM OF APARTHEID.	clause position verb formality ellipsis  mode, verb formality adjective inclusion clause position  conjunction  verb formality  clause position mode, verb formality conjunction  verb formality adjective inclusion  word formality verb formality verb formality  conjunction verb formality  verb formality

## 6 References

1. Atkinson, J.M., *Understanding Formality: the Categorization and Production of 'Formal' Interaction*, in *British Journal of Sociology*, vol 33 no 1, 1982.

2. Bain, W.M., *A Case-based Reasoning System for Subjective Assessment*, AAAI Proceedings, 1986.
3. Birk, N.P. & Birk, G.B., *Understanding and Using English*, fourth edition, Odyssey Press, New York, 1965.
4. Brown, P. & Levinson, S.C., *Universals in Language Usage: Politeness Phenomena*, in *Questions and Politeness: Strategies in Social Interaction*, Goody E. (ed), Cambridge University Press, Cambridge, 1978.
5. Charniak, E., Riesbeck, C.K. & McDermott, D.V., *Artificial Intelligence Programming*, Lawrence Erlbaum Associates, Hillsdale, 1980.
6. Goody, E. (ed), *Questions and Politeness: Strategies in Social Interaction*, Cambridge University Press, Cambridge 1978.
7. Harada, S.I., *Honorifics*, in *Syntax and Semantics 5: Japanese Generative Grammar*, Shibátani M. (ed), Academic Press, New York, 1976.
8. Hill, A.S., *The Foundations of Rhetoric*, Harper & Brothers, New York, 1892.
9. Hovy, E.H., 1986a, *Some Pragmatic Decision Criteria in Generation*, in *Natural Language Generation: Recent Advances in Artificial Intelligence, Psychology, and Linguistics*, Kempen, G. (ed), Kluwer Academic Publishers, Dordrecht, Boston, 1987.
10. Hovy, E.H., 1986b, *Putting Affect into Text*, Proceedings of the Eighth Conference of the Cognitive Science Society, 1986.
11. Hovy, E.H., 1987a, *Generating Natural Language under Pragmatic Constraints*, in *Journal of Pragmatics*, vol XI no 6, 1987, forthcoming.
12. Hovy, E.H., 1987b, *Interpretation in Generation*, Proceedings of the AAAI conference, 1987.
13. Irvine, J.T., *Formality and Informality of Speech Events*, in *American Anthropologist*, vol 81 no 4, 1979.
14. Kuno, S., *The Structure of the Japanese Language*, Harvard University Press, 1973.
15. Lakoff, R., *Politeness, Pragmatics, and Performatives*, in *Proceedings of the Texas Conference on Performatives, Presuppositions, and Implicatures*, 1977.
16. Schank, R.C., *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*, Cambridge University Press, Cambridge, 1982.
17. Schank, R.C. & Abelson, R.P., *Scripts, Plans, Goals and Understanding*, Lawrence Erlbaum Associates, Hillsdale, 1977.