

**Sublanguages, Complexity, Genre, Domain,
Controlled Languages, Authoring Systems,
General Translation, Random Text,
"Try-Anything" Systems:
A Plea for Definitions and Research Priorities**

**Muriel Vasconcellos
Washington, D.C.
1993**

"Vasconcellos attacks sublanguages."

— H. Somers, "Forum Issue," *MT*, 1992

**MT-ites are super-careful in defining their methods;
not so careful in defining the data they operate on.**

— Vasconcellos, November 1993

SUBLANGUAGE: A subset of natural language.

Possible defining criteria:

- **Intervention in its production**
- **Complexity**
- **Genre**
- **Domain**

Characteristics of COMPLEXITY:

- **Length:** sentences and words, av. # syllables per 100 words (Flesch)
- **Vocabulary:** # single-word terms, % recurring SWT, # phrases, % recurring phrases; potential for ambiguity
- **Syntax:** noun phrases (modification structure), verbs (tense, aspect, modality), complementation, connectives, negatives, quantifiers. POTENTIAL FOR AMBIGUITY. SPECIAL PROBLEMS: conjoined clauses, anaphora, incomplete sentences, anacolutha, repairs
- **Embedding:** av. # clauses per sentence, average # clauses per information block, av. depth of embedding (Cook); integration vs. fragmentation (Chafe)
- **Argument structure (case frames):** pattern, variety (Cook)
- **Networks of semantic concepts:** # of nodes, linkages (Meyer; Nirenburg)
- **Information structure:** markedness, rate of delivery (Prague School [Mathesius]; Halliday; Chafe; Clark & Haviland)
- **Cohesive devices:** reference, substitution/ellipsis, lexical cohesion, conjunction (Halliday & Hasan)
- **Thematic structure:** markedness (Prague School [Trávníček]; Halliday)

EXAMPLE OF COMPLEXITY

En otros (Jamaica, Cuba, México y Panamá), sin embargo, la prevalencia inicial ha declinado, lo cual además se ha asociado a una temprana introducción, a veces previa a la primera tetada, de agua o soluciones glosadas y de fórmulas de leche en biberones, que interfieren con una lactancia existosa, que hacen que no sólo la prevalencia inicial de LME sea baja, sino que ésta descienda rápidamente en los días subsiguientes, como se observa en Brasil, República Dominicana, Colombia y Ecuador, donde los ID al mes son muy elevados.

1 sentence: 88 words; 7 clauses

Figure 7. Style complexity ranking of selected texts.²

Grade I: Simple style, below 170

3.1	1.29	1.26	Mark Twain, Tom Swayer
2.1	1.40	1.29	Ernest Hemingway, For Whom the Bell Tolls
2.1	1.50	1.48	James Joyce, A Painful Case
2.3	1.53	1.33	Saul Bellow, Herzog
2.5	1.56	1.40	Updike, Rabbit, Run
2.4	1.60	1.42	D. H. Lawrence, Lady Chatterly's Lover
2.1	1.62	1.52	Hans Konig, Death of a Schoolboy
2.0	1.67	1.45	Dressman Taylor, Address Unknown
2.7	1.69	1.44	John O'Hara, Assembly

Grade II: Medium style, below 200

2.6	1.73	1.42	A. Solzhenitsyn, Cancer Ward
2.6	1.73	1.50	Dostoevsky, The Idiot
2.4	1.77	1.61	Graham Greene, I Spy
2.5	1.79	1.60	James Baldwin, This Morning, This Evening, So Soon
2.9	1.81	1.52	Frank Norris, McTeague
2.0	1.82	1.52	Robert Crichton, Camerons
2.4	1.84	1.54	Theodore White, The Making of a President
2.3	1.92	1.70	Ramsey Clarke, Crime in America
2.3	1.92	1.70	Kurt Vonnegut, Slaughterhouse Five

Grade III: Complex style, below 250

3.0	2.00	1.83	Ellery Queen, A Fine and Private Place
2.9	2.07	1.72	Albert Camus, The Plague
3.2	2.13	1.84	James N. Hall, Lost Island
3.0	2.14	1.60	Mark Twain, Innocents Abroad
2.4	2.18	1.75	Joseph Heller, Catch-22
3.4	2.27	1.68	C. S. Lewis, Out of the Silent Planet
3.1	2.36	2.16	Andre Malraux, Man's Fate
2.9	2.42	1.93	Arthur Clarke, Rendezvous with Rama
2.7	2.45	1.79	John O'Hara, Waiting for Winter

Grade IV: More complex style, 250 and over

2.8	2.55	1.79	Graham Greene, The Basement Room
3.5	2.70	2.06	Thoreau, Walden
4.7	2.76	1.88	Franz Kafka, The Trial
3.4	2.83	1.88	William Faulkner, Light in August
4.5	3.00	2.11	F. Scott Fitzgerald, The Great Gatsby
4.2	3.23	2.10	Richard Taylor, Good and Evil
5.7	3.35	2.21	H. P. Lovecraft, Cool Air
5.7	3.35	2.23	A. Solzhenitsyn, August, 1914
4.2	3.82	2.19	Dostoevsky, White Nights

(From Cook 1975)

THE FOUR GENRES OF DISCOURSE (Longacre 1976)

	- PROJECTED	+ PROJECTED
S U C C + E S S I O N	<p><u>Narrative</u></p> <ul style="list-style-type: none"> - First or third person - Agent-oriented - Accomplished time - Chronological linkage 	<p><u>Procedural</u></p> <ul style="list-style-type: none"> - Nonspecific person - Patient-oriented - Projected time
S U C C - E S S I O N	<p><u>Expository</u></p> <ul style="list-style-type: none"> - No necessary person reference - Subject-matter oriented - Time not focal - Logical linkage 	<p><u>Hortatory</u></p> <ul style="list-style-type: none"> - Second person - Addressee-oriented - Mode, not time - Logical linkage

Characteristics of GENRE:

- **Form:** business letter, grant proposal, legal agreement; maintenance manual, weather report, textbook, essay, script/play, novel, poem, etc.
- **Aim (Kinneavy)**
- **Involvement vs. detachment (Chafe)**
- **Register: formality vs. informality**
- **Projection: shorter constructions, person not specified (Longacre)**
- **Succession: longer constructions, person specified (Longacre)**
- **Coherence (Vasconcellos)**

THE ADMS OF DISCOURSE (Kinneavy 1971)

SCHOOL	Aristotle & Aquinas	Cassirer	Morris	Miller	Russell	Reichenbach	Richards	Bühler Jakobson Kinneavy
PRINCIPLE	Level of probability	Historical sequence in Greece	Reactions of animals to stimuli	Sociopsychological motives	Kinds of sentences	Faculty addressed	Reference vs. emotion	Communication process
	Scientific—certain	Metaphysical, representative of the world	Informative	Informative—information more uniform	Informative—declarative	Communicative (thoughts to be believed)	Scientific—pure reference	Reference—Informative —Scientific—Exploratory
	Dialectical—probable		Valuative	Opinion—more uniform	Questioning—interrogative			
	Rhetorical—seemingly probable	Pragmatic—practical use by Sophists	Incitive	Status change	Imperative	Promotive—actions to accomplish	Rhetorical—mixed reference and emotion	Persuasive
	Poetic—internally probable						Poetic—pure emotion	Literary
		Mythological—expressing aspiration Interjectional	Systemic	Emotive	Expressive of emotion—exclamatory	Suggestive—emotions aroused		Expressive

DOMAIN (*Webster III*, 3a and b):

- **A distinctly limited sphere of knowledge or of intellectual, institutional, or cultural activity**
- **A circumscribed realm of human concern**

It follows that a DOMAIN may be delimited by the network of concepts to which it refers. In referring to and expressing these concepts, it has its own SUBLANGUAGE.

Characteristics of a DOMAIN SUBLANGUAGE:

- **Vocabulary:** specific words, with specific meanings, and not others; special meanings assigned to words from general vocabulary
- **Syntactic structures:** noun phrases—modification structure;
- verbs—person, tense, aspect, modality, conjoined clauses;
- types of complementation, connectives, negatives, quantifiers; typical ambiguities; typical problem structures—*anaphora, fragments, etc.*
- **Argument structure (case frames):** pattern, variety
- **Detachment vs. involvement**
- **Information structure:** markedness, rate of delivery, integration vs. fragmentation
- **Cohesion:** reference, substitution/ellipsis, lexical cohesion, types of conjoining
- **Thematic structure:** markedness

The "TRY ANYTHING" MT SYSTEM

will have a go at any natural language thrown at it, albeit with varying degrees of success. The obvious example is SYSTRAN.

- ***V. Lawson, Practical Experience of Machine Translation, 1982***

A SHORT BIBLIOGRAPHY ON THE CHARACTERISTICS OF TEXT

- Bühler, Karl. 1933 [1982]. *The Axiomatization of the Language Sciences*. Translation of *Die Axiomatik der Sprachwissenschaften*, by Robert E. Innis. In: *Karl Bühler: Semiotic Foundations of Language Theory*. New York & London: Plenum, 77-164.
- Chafe, Wallace. 1982. "Integration and Involvement in Speaking, Writing, and Oral Literature." In: *Spoken and Written Language: Exploring Orality and Literacy*, ed. Deborah Tannen. Norwood (N.J.): Ablex. *Advances in Discourse Processes*, 9. 35-53.
- Clark, Herbert H., & Susan E. Haviland/ 1977. "Comprehension and the Given-New Contract." In: *Discourse Production and Comprehension*, ed. Roy O. Freedle. Norwood (N.J.): Ablex. *Advances in Discourse Processes*, 1.
- Cook, Walter A. 1975. "Stylistics: Measuring Style Complexity." In: *Georgetown University Working Papers on Languages and Linguistics*, 11. Washington, D.C.: Georgetown University Press, 106-120.
- Cook, Walter A. 1978. "A Case Grammar Matrix Model." In: *Georgetown University Working Papers on Languages and Linguistics*, 11. Washington, D.C.: Georgetown University Press, 200-207.
- Fleisch, Rudolph. 1949. *The Art of Readable Writing*. New York: Harper & Row. Also: *The Art of Plain Talk* (1946) and *The Art of Clear Thinking* (1951), Harper & Row.
- Grimes, Joseph P. 1975. *The Thread of Discourse*. The Hague: Mouton.
- Halliday, M.A.K. 1967-68. "Notes on Transitivity and Theme in English." Parts 1 and 2, *Journal of Linguistics* 3, 37-81; 199-244; Part 3, *Journal of Linguistics* 4, 179-215.
- Halliday, M.A.K., and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Jakobson, Roman. 1960. "Closing Statement: Linguistics and Poetics." In: *Style in Language*, ed. Thomas A. Sebeok. Cambridge (Mass.): MIT Press; New York & London: Wiley, 350-377.
- Kinneavy, James L. 1971. *A Theory of Discourse: The Aims of Discourse*. New York & London. W.W. Norton.
- Longacre, Robert E. 1976. *The Anatomy of Speech Notions*. Lisse (Belgium): Peter de Ridder.
- Meyer, Ingrid. 1993. "Knowledge Management for Terminology-Intensive Applications: Needs and Tools." In *Lexical Semantics and Knowledge Representation*, eds. J. Pustejovsky & S. Bwergler. Berlin: Springer Verlag, 21-37.
- Nirenburg, Sergei. 1993. *Knowledge-based Language Generation*. Norwood (NJ): Ablex.
- Vasconcellos, Muriel. 1989. "Cohesion and Coherence in the Presentation of Machine Translation Products." In: *Georgetown University Round Table on Languages and Linguistics 1988*. Washington, D.C.: Georgetown University Press, 89-105.