GRAMMATICAL AND UNGRAMMATICAL STRUCTURES IN USER-ADVISER DIALOGUES: EVIDENCE FOR SUFFICIENCY OF RESTRICTED LANGUAGES IN NATURAL LANGUAGE INTERFACES TO ADVISORY SYSTEMS.

Raymonde Guindon
Microelectronics and Computer Technology Corporation
P.O. Box 200195
Austin, Texas 78720
guindon@mcc.com

Kelly Shuldberg[1]
University of Texas, Austin & MCC

Joyce Conner
Microelectronics and Computer Technology Corporation

ABSTRACT

User-adviser dialogues were collected in a typed Wizard-of-Oz study ("man-behind-the-curtain study"). Thirty-two users had to solve simple statistics problems using an unfamiliar statistical package. Users received help on how to use the statistical package by typing utterances to what they believed was a computerized adviser. The observed limited set of users' grammatical and ungrammatical forms demonstrates the sufficiency of a very restricted grammar of English for a natural language interface to an advisory system. The users' language shares many features of spoken face-to-face language or of language generated under real-time production constraints (i.e., very simple forms of utterances). Yet, users also appeared to believe that the natural language interface could not handle fragmentary or informal language and users planned or edited their language to be more like formal written language (i.e., very infrequent fragments and phatics). Finally, users also appeared to believe in poor shared context between users and computerized advisers and referred to objects and events using complex nominals instead of faster-to-type pronouns.

INTRODUCTION

It has been argued that natural language interfaces with a very rich functionality are crucial to the effective use of advisory systems and that interfaces using formal languages, menus, or direct manipulation will not suffice (Finin, Joshi, and Webber, 1986). Designing, developing, and debugging a rich natural language interface (its parser, grammar, recovery strategies from unparsable input, etc.) are time-consuming and labor-intensive. Nevertheless, natural language interfaces can be quite brittle in the face of unconstrained input from the user, as can be found in applications such as user-advising. One step toward a solution to these problems would be the identification of a subset of grammatical and ungrammatical structures that correspond to the language generated by users in any user-advising situations, irrespective of the domain. This subset could be used to design a core grammar, strategies to handle ungrammatical input, and some parsing heuristics portable to any natural language interface to advisory systems. This strategy would increase the habitability of the natural language interface (Watt, 1968; Trawick, 1983) and reduce its development cost.

An important feature of this restricted subset is its independence from a particular domain (e.g., statistics, medicine), making it portable. This is in contrast with another strategy which also capitalizes on restricted subsets of English, the use of sublanguages. There are naturally occurring subsets of English, usually associated with a particular domain or trade that have been called sublanguages (Harris, 1968; Kittredge, 1982). Sublanguages are characterized by distinctive specialized syntactic structures, by the occurrence of only certain domain-dependent word subclasses in certain syntactic combinations, and by the inclusion of specific ungrammatical forms (Sager, 1982). However, the association of a sublanguage with a particular domain and the emphasis on syntactic-semantic co-restrictions reduce the portability of a grammar defined on such a sublanguage.

This paper presents an empirical characterization of users' language in an user-advising situation for the purpose of defining a domain-independent restricted subset of grammatical and ungrammatical structures to help design more habitable natural language interfaces to advisory systems. This paper also presents an interpretation of the factors that cause users to naturally limit themselves to a very restricted subset of English in typed communications between users and computerized advisers. We believe these factors will be found in any typed communications between users and advisers for the purposes of performing a primary task. Hence, the restricted subset of English should be general to any such situations.

A STUDY OF USER-ADVISER DIALOGUES IN A WIZARD-OF-OZ SETTING

METHOD & PROCEDURE

Thirty-two graduate students with basic statistical knowledge were asked to solve up to eleven simple statistics problems. Participants had to use an unfamiliar statistical package to solve the problems. The upper window of the participants' screen was used to perform operations with the statistical package and the lower window was used to type utterances to the adviser. The participants were instructed to ask help in English from what they believed was a computerized adviser by typing in the help window. The participants' and adviser's utterances were sent to each other's monitor and the utterances were recorded and time-stamped automatically to files.

[1]Now at Automated Language Processing Systems, Provo, Utah

RESULTS AND COMPARISON TO OTHER STUDIES

We are reporting only a small subset of our results, those to be compared to the results of Thompson (1980) and of Chafe (1982). The comparison is to identify the grammatical and ungrammatical specializations specific to users' language with advisory systems and to help determine what features of user-advising situations might encourage or cause such specializations of structures. Chafe (1982) investigated Informal Spoken language (i.e., dinner table conversations) and formal written language (i.e., academic papers). Thompson (1980), in her second study, compared three types of dialogues, Spoken Face-to-Face, Typed Human-Human (terminal-to-terminal) with both conversants knowing their counterpart was human, and Human-Computer using the REL natural language front-end. The task was information retrieval.

The data table report two sets of data, the percentage of utterances with a particular form (e.g., one or more fragments, one or more phatics) to compare to Thompson's results and the corresponding number of occurrences of this form per 1000 words to compare to Chafe's results. When numbers are omitted from the tables, the corresponding data were not collected by Thompson or Chafe. Note that the reported data are only about users' utterances, and not the adviser's utterances. We will use *typed user-adviser dialogues* and *Wizard-of-Oz condition* to refer to the data of our study.

Completeness and Formality of Users' Utterances

As can be seen in Table 1, for completeness (i.e., fragments) and formality (i.e., phatics and and-connectors) users' utterances with advisory systems are more like Human-Computer dialogues and Formal Written language than Spoken Face-to-Face or Typed Human-Human dialogues.

Table 1: Completeness and Formality

		Chafe		Thompson		
Form	Wizard-of-Oz	Formal Written	Informal Speech	Human-Computer	Typed Human-Human	Spoken Face-Face
Fragment	24%	—	—	19% 27	74% 74	59% 62
Phatics	2% 2	0	23	4% 6	59% 59	100% 116
And-Connector	1% 1	14	56			

Users avoided casual forms of language since they produced only 24% of fragmentary utterances, as opposed to 74% in Typed Human-Human dialogues, but similar to 19% in the Human-Computer condition. Similarly, we found 2% of utterances with phatics, as opposed to 59% in the Typed Human-Human dialogues, but similar to 4% in the Human-Computer dialogues. Likewise, Chafe (1980) found no phatics in Formal Written discourse, but about 23 per 1000 words in informal speech. There is a similar finding for and-connectors.

Users in the typed user-adviser dialogues seem to expect the interface to be unable to handle fragmentary input such as found in Informal Spoken language and planned or edited their language to be as complete and formal as in the Human-Computer dialogues, and more complete and formal than the language in Typed Human-Human dialogues. This is the case even though the Wizard in our study hardly ever rejected or misunderstood any users' utterance, no matter how fragmentary or ungrammatical it was. However, when conversants know that their counterpart is another human, their language contains a large percentage of fragments and phatics, even when typed. So it appears that a priori beliefs about the nature and abilities of the adviser (i.e., this is not a human) can determine the characteristics of the language produced by the user, even when task and linguistic performances by the adviser were not negatively affected by fragmentary language from the user.

Ungrammaticalities

Even though users seemed to attempt to edit or plan their utterances to be more complete and formal, 31% of the utterances contained one or more ungrammaticalities (excluding spelling and punctuation mistakes, if included about 50% of utterances were ungrammatical). The most frequent ungrammaticalities were fragments (13% of utterances with part(s) of the utterance being one or more fragments), missing constituents (14% of utterances with one or more determiners missing), and lack of agreement between constituents (5% of utterances). While users seemed to plan or to edit their language to be as complete and formal as in the Human-Computer dialogues, certain types of ungrammaticalities were produced. Two possible interpretations of this finding are: 1) Certain types of ungrammaticalities do not seem to be easily under the conversant's control and edited or planned to be avoided during the dialogue; 2) They correspond to a telegraphic language assumed to be understood by the interface.

It would be interesting to find whether really two types of ungrammaticalities exist, some that can be avoided under some planning and others that cannot be so easily avoided. However, it is unclear whether the purposeful avoidance of some ungrammaticalities by users can be capitalized upon to reduce the need for sophisticated robust parsing as we do not know the cost from the users of avoiding certain types of ungrammaticalities. On the other hand, knowing the nature and frequency of the actual ungrammaticalities produced by users, as they are provided by this study, facilitates realizing robust parsing.

General Syntactic Features

As can be seen in Table 2, users' utterances in typed user-adviser dialogues resemble more spoken informal discourse than written formal discourse. The difference in number of occurrences per 1000 words between the Wizard-of-OZ condition and the Informal Spoken condition is much less than the same difference between the Wizard-of-OZ and the Formal Written conditions.

Table 2: Occurrences per 1000 Words of Various Syntactic Features

	Wizard-of-Oz	Chafe Informal Speech	Chafe Formal Written
Sentence Length	9	7	17-25
Passive Voice	1.0	5.0	25.4
Coordinate Conjunctions	6.7	5.8	23.8
Attributive Adjectives	32.8	33.5	134.9
First Person Referents	49.0	61.5	4.6
Nominalizations	11.4	9.7	55.8
Object of Nominalized Verb	.7	.01	12.3
Subject of Nominalized Verb	1.2	.01	4.1
Series	0	.9	7.2
Relative Clauses	3.8	9.7	15.8

Short, simple (95% of our utterances were simple), active sentences, with few coordinations, few subordinations, few relative clauses, few nominalizations, (and deletion of determiners and unmarked agreements, see the section on Ungrammaticalities) characterize the language in typed user-adviser dialogues observed in our study. These same features are features of unplanned language, which are also features of child language, which are also features of language produced under real-time production constraints (Ochs, 1979; Givon, 1979).

While formality and completeness of typed user-adviser dialogues resemble more Formal Written language, the general syntactic features of typed user-adviser dialogues resemble more Informal Spoken language. Formality and completeness appear to be independent properties of users' language from the general syntactic features, possibly planned independently.

More important for the design of natural language interfaces, the observation that typed user-adviser dialogues resemble language produced under real-time production constraints indicates that users are strained by typing utterances to request help to perform a primary task. This constrains the usability of natural language interfaces as interfaces to advisory systems. One needs to identify the conditions under which the benefits of obtaining help outweight the costs of typing in utterances to determine when natural language interfaces are effective interfaces to advisory systems. On the other hand, the natural restrictions on the language produced by the users appear generalizable to any situation where real-time production constraints exist, of which, we believe, any typed interaction to an advisory system for the purpose of performing a primary task is an instance.

Features Due Specifically to the User-Advising Application

As can be seen in Table 3, there are less imperatives in user-advising dialogues because the user cannot request the adviser to perform a statistical operation. Moreover, we also observe a goal-directed language with frequent *to* infinitives (*I want/need to ...*) and *to* purpose clauses (*What is the command to compute ...*), much more frequent than in Informal Spoken or Formal Written languages. We believe this is the only feature that appears to be specific to the advisory application, as opposed to be specific to communications under real-time constraints. However, the goal-directedness of the language may be specific to advisory systems for procedural tasks as opposed to more generaln information retrieval tasks. Of course, we are here excluding lexical restrictions because they are expected and uninteresting and syntactic-semantic co-restrictions because of the desire for easy portability.

Table 3: Features Specific to Advising

	Wizard-of-Oz	Information Retrieval Human-Computer	Chafe Informal Speech	Chafe Formal Written
Imperative	5.3%	19.0%		
To Complement	17.4		2.9	8.8

Complexity of Referring Expressions

In our study, users produced mostly very simple sentence constructions, as if under real-time production constraints (e.g., users' utterances were short and 95% of them were simple (see the section on General Syntactic Features)). Nevertheless, very few pronouns occurred, 3% of utterances contained pronouns, similar to what was found in Formal Written Language, Human-Computer dialogues, and in Cohen, Pertig, & Starr (1982) in their typed terminal-to-terminal condition. This is surprising because pronouns are very short to type. However, there were very frequent complex nominals with prepositional phrases (e.g., *a record of the listing of the names of the features*). At least 50% of the user-adviser utterances had one or more prepositional phrases. As can be seen in Table 4, most of the structurally ambiguous prepositional attachments are to NPs, in fact, mostly to the most contiguous/nearest NP. So, users prefer longer to type complex nominals with explicit relations between contiguous NPs over faster to type pronouns, even though there is evidence that they are operating under real-time production constraints. Because pronominal noun phrases (and also deictic expressions) are so rare, it appears that users rely little on spatial context (i.e., the screen), linguistic context (i.e., the utterances produced so far), and task context (i.e., statistical commands typed so far) in producing referring expressions. One interpretation of this finding is that users believe that there is poor shared context between user and adviser when they do not share physical context (as in Formal Written language) or do not know the linguistic capabilities of the conversant (as in Human-Computer dialogues). So, while in unplanned discourse speakers rely more on the context to express propositions and use more pronouns than in planned discourse (Ochs, 1979) and while user-adviser dialogues exhibit many features of unplanned discourse, users did not capitalize on context in producing referring expressions. It appears that the referential functions in language can be planned independently of and are not necessarily subject to the same real-time production constraints than the predicative and other functions of language. Again we are finding that typed user-adviser dialogues have some features of planned, Formal Written language but also have features of unplanned, Informal Spoken language.

Table 4: Distribution of Propositional Attachments

Complex NP	71	NP/VP	120	Complex NP/VP	204
Nearest NP	67	NP	72	Complex NP–nearest	131
Other NPs	4	VP	31	Complex NP–others	8
		Ambiguous	17	VP	44
				Ambiguous	21

Nevertheless, not only are most prepositional attachments to NPs to create precise description of objects, they are mostly to the most contiguous NP. This observation suggests that real-time production constraints nevertheless play some role in the production of referential expressions. Users appear to minimize resources allocated to the production of referential expressions by reducing short-term memory load by attachments to the lowest, most recent NP. This interpretation is supported by studies that show that it is easier to process right-branching structures than left-branching ones (Yngve, 1960).

The finding that most prepositional phrases attach to NPs rather than VPs and moreover attach most often to the lowest, nearest NP is important for the semantic interpretation of sentences because of the combinatorial explosion of possible attachments of prepositional phrases.

DISCUSSION

Users' utterances in typed user-adviser dialogues, when the users believe that the adviser is computerized, resemble Informal Spoken speech, except for referring expressions (i.e., frequent complex nominals) and for completeness and formality (i.e., few phatics and and-connectors, and relatively few fragments), in which case they resemble more Formal Written language. We would like to hypothesize that the grammatical and ungrammatical forms observed occur because the communicative context and the application induce certain user's beliefs and goals and induce certain processing constraints which determine the most effective syntactic forms to communicate verbally. The communicative context describes dimensions of the situation in which the discourse is generated that are believed to affect the form of the discourse. Examples of dimensions are: interaction, the extent to which user and adviser can quickly interact, respond to each other; involvement, the extent to which the communication is directed specifically to one person as opposed to an anonymous class of persons; spatial commonality, the degree to which the conversants see each other, see the same physical environment, and know that they share this environment perceptually. As can be seen in Table 5, typed user-adviser dialogues in a Wizard-of-Oz setting are more similar to Informal Spoken language on dimensions of interaction and in-

volvement, but more similar to Formal Written language on the dimension of spatial commonality. We would like to hypothesize that different values on these dimensions are associated with different restricted languages produced by the users. Findings from Biber (1986) help support this hypothesis. He performed a factor analysis on 545 text samples. He uncovered the following three dimensions:

- INTERACTIVE vs. Edited: High personal involvement and real-time constraints.
- SITUATED vs. Abstract contexts: Reliance on external situation, concrete vs. detached and deliberate.
- IMMEDIATE vs. Reported: Reference to a current situation vs. removed or past situation.

Table 5: Communicative Context Parameters

	Informal Spoken (Face-to-Face)	Terminal-to-Terminal (Human-Human)	Wizard of Oz	Formal Written
Modality	spoken	written	written	written
Interaction	high	high	high	low
Involvement	high	high	high	low
Shared Knowledge	low to high	medium to high	low to medium	low to high
Spatial Commonality	high	low	low	low

From the set of features reported by Biber that loaded highly on the three dimensions, user-adviser dialogues had features of both Interactive texts (e.g., many Wh-questions, many first person references, final prepositions) and edited texts (e.g., few phatics). This is because user-adviser dialogues, while written by users uncertain about the interface's ability to handle fragmentary and informal input, have a high degree of interaction and involvement of the conversants. The syntactic features observed in user-adviser dialogues overlapped greatly with the features of situated texts (e.g., few passives and nominalizations), except for the frequent use of complex nominals and unfrequent use of pronouns an deictic expressions, and of Immediate texts (e.g., use of present tense, few third person pronouns). The complexity of referring expressions uncovers a dimension not revealed in Biber's work: the degree of believed shared knowledge by the conversants. Our users seemed to assume poor shared knowledge and relied on complex referring expressions to insure successful communication. Another dimension is the conversants' belief in the ability of their counterpart to handle fragmentary or informal language.

Informal Spoken face-to-face language is often unplanned, interactive, situated, immediate, and subject to real-time production constraints. So are users' typed utterances to advisory systems. However, unlike Informal Spoken face-to-face language, users believe that there is poor shared context between conversants and rely little on context in producing referring expressions and users do not assume that the interface can handle fragmentary or informal language.

We would like to conclude by making the hypothesis that any typed terminal-to-terminal user-adviser dialogues will be similar to Informal Spoken language, as was observed in our study, because they are under the same communicative context and application. This provides a subset of grammatical and ungrammatical forms that can be used to define a core grammar portable to most user-advising situations, irrespective of the domain. On the other hand, the complexity of referring expressions and the degree of completeness and formality of language may differ according to the users' beliefs about the linguistic capabilities of the interface.

ACKNOWLEDGEMENTS

We wish to thank Elaine Rich, Kent Wittenburg, and Gregg Whittemore for useful comments on this research project. We also thank Sherry Kalin, Hans Brunner, and Gregg Whittemore for their help in collecting or analyzing the dialogues between users and adviser.

REFERENCES

Biber, D. (1986). Spoken and written textual dimensions in English. *Language, 62* (2), 384-414.

Chafe, W.L. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy.*. Norwood, NJ: Ablex.

Cohen, P.R., Pertig, S., & Starr, K. (1982). Dependencies of discourse structure on the modality of communication: Telephone vs. teletype. *Proceedings of the 20th Annual Meeting of the Association for Computational Linguistics.* University of Toronto, Ontario, Canada.

Finin, T.W., Joshi, A.K., & Webber, B.L. (1986). Natural language interactions with artificial experts. *Proceedings of the IEEE, 74*, 7, 921-938.

Givon, T. (1979). From discourse to syntax: Grammar as a processing strategy. In T. Givon (Ed.), *Syntax and Semantics: Discourse and syntax.* New York: Academic Press.

Grishman, R., Hirshman, L., & Nhan, N.T. (1986) Discovery Procedures for Sublanguage Selectional Patterns: Initial Experiments. *Computational Linguistics, 12*(3).

Harris, Z.S. (1968). *Mathematical Structures in Language.* New York: Wiley (Interscience).

Kittredge, R. (1982). Variation and Homegeneity of Sublanguages. In R. Kittredge & J. Lehrberger (Eds.), *Sublanguage: Studies of Language in Restricted Semantic Domains.* New York: Walter de Gruyter & Co.

Ochs, E. (1979). Planned and unplanned discourse. In T. Givon (Ed.), *Syntax and Semantics: Discourse and syntax.* New York: Academic Press.

Sager, N. (1982). Syntactic Formatting of Science Information. In R. Kittredge & J. Lehrberger (Eds.), *Sublanguage: Studies of Language in Restricted Semantic Domains.* New York: Walter de Gruyter & Co.

Thompson, B.H. (1980). Linguistic analysis of natural language communication with computers. *Proceedings of the 8th International Conference on Computational Linguistics.* Tokyo, Japan.

Trawick, D.J. (1983). *Robust Sentence Analysis and Habitability.* Doctoral Dissertation, California Institute of Technology, Pasadena.

Watt, W.C. (1968). Habitability. *American Documentation, 19*(3), 338-351.

Yngve, V. (1980). A model and an hypothesis for language structure. *Proceddings of the American Philosophical Society.*