HUMAN-TO-MACHINE INTERACTION IN NATURAL LANGUAGE: EMPIRICAL RESULTS
OF FIELD STUDIES WITH AN ENGLISH AND A GERMAN INTERFACE

J. Krause
IBM Scientific Center and Linguistic Information Science
Heidelberg, West Germany
University of Regensburg, West Germany

Abstract: Questions of feasibility and desirability of natural language interfaces for human-machine interactions gain more and more interest in empirical research. There is growing consensus that field studies provide valuable leads with respect to design decisions.

One of the most important aspects of user friendly interfaces are the restriction rules. The problem is how to restrict the diversity of communicative use of the human native language without losing the advantage of this form of communication with computers.

The paper tries to answer this question on the basis of several empirical studies which investigate the same domain-independent natural language query system, using various applications in two different natural languages - English and German. The studies were performed in cooperation with the IBM Heidelberg Scientific Center (West Germany), the New York University (Advanced Language Project), and the Department of Linguistic Information Science at the University of Regensburg (West Germany). Altogether, these experiments involved about 100 subjects and over 12,000 queries, constituting the bulk of empirical evaluations of natural query language systems reported to date. Results of these experiments are presented and one of the most recent successful software packages Q&A is discussed with respect to the selected restriction rules.

1. INTRODUCTION

At the moment interesting developments can be observed in two areas which influence each other, namely in natural language processing and in the field of graphical interfaces. Both are part of the concept of fifth-generation computers, the culture of which is intended to be "user natural in its human factors" (Gaines and Shaw, 1986; Hirose and Fuchi, 1984). Both are in the headlines these days, but the reasons for it are different.

a) Natural Language Interfaces to Databases (NLI).

In the 1970's and early 1980's NLI's were the undisputed favorite application for computational natural language research. But this broadly accepted consensus no longer
exists. At COLING 1984, Karen Sparck-Jones invited the members of a panel "Natural Language and Databases" to "speak to the proposition ... that database query is no longer a good, let alone the best, test environment for language processing research, because it is insufficiently demanding in its linguistic facts and too idiosyncratically demanding in its non-linguistic ones."

(Sparck-Jones, 1984, p. 182). The comments given at the panel clearly showed that there was a shift in research interest in the last years. Sparck-Jones (1984, p. 183) claimed "that the database application is an inadequate test environment for natural language understanding systems." Others (McKeown, 1984; Flickinger, 1984; Carbonell, 1984) supported a more balanced view.

"the database query task was an excellent paradigmatic problem for computational linguistics ... it is now time for the field to abandon its protective cocoon and progress beyond this rather limiting task."

At this time increasing public interest in NLI comes from another area; people are fascinated in a series of commercially available NLI. In 1985 three new NLI's appeared on the market: NLMenu, Language Craft, and Question and Answer (Q&A). The latter was supplemented by a German version in 1986. People are giving much attention to Q&A, which is cheap ($299 list price) and runs on a PC with 512K RAM.

b) The second area mentioned before concerns the domain of graphical interfaces, that is icons, mouse, and direct manipulation (DM).

The fascination of interacting with the computer in the DM mode did not lose its attraction both for expert and novice users, as it can be detected for instance in Shneidermann (1983) and Hutchins (Hutchins et al., 1986). With tools like MS-WINDOWS these positive feelings will also reach the mass of PC-users in the near future. The fascination with DM reminds one of the enthusiasm with which the first attempts at natural language communication were caught up.

Both NLI and DM are today of special interest, because they now can be realized on PC's. Both claim to substantially improve human computer interaction (HCI), and both have the image of being so convincingly better than formal query languages. But this is not enough. Can we validate the superiority of NLI against alternatives like formal query languages or DM, or of DM against NLI? And can we decide between design alternatives of NLI?

In this paper I want to demonstrate which design criteria of an NLI are validated by empirical testing, and I want to take a first glance at a differentiation of areas where DM will be preferred to NLI and vice versa.

2. NLI FOR THE REQUEST OF FACTS, ORGANIZED IN TABLES.

Those NLI's which are now commercially available have much in common with those which produced significant empirical results from field studies,