

# **RAPID PORTING OF THE PARLANCE<sup>tm</sup> NATURAL LANGUAGE INTERFACE**

Madeleine Bates  
BBN Systems and Technologies Corp.  
10 Moulton Street  
Cambridge, Mass. 02138

## **ABSTRACT**

Developing knowledge bases for AI systems takes too long and costs too much. Even a "portable" system may be expensive to use because its installation takes a long time or requires the labor of scarce, highly-trained people. BBN has recently created a tool for acquisition which dramatically reduces the time and cost of installing a natural language system.

During 1988, BBN used its Learner<sup>tm</sup> tool to configure the Parlance<sup>tm</sup> database interface to two different versions of a large Navy database. The configuration process was performed primarily with development versions of the Learner, which is a software tool for creating the knowledge bases, vocabulary, and mappings to the database that enable the Parlance interface to understand questions addressed to a particular database. The Learner reduced the time required to create Parlance configurations from months to weeks, and demonstrated that the Learner works effectively on databases with many hundreds of fields.

## **INTRODUCTION**

### **THE NAVY'S IDB DATABASE**

The IDB (hereafter called the Navy database) is a large, evolving database being used in the Fleet Command Center at the Navy's Pacific Fleet headquarters in Pearl Harbor, Hawaii [Ceruti, 1988]. It has dozens of tables and hundreds of fields containing information about hundreds of U.S. ships, planes and other units, as well as more limited data on foreign units.

Examples of the kind of information that may be available for a particular unit are: its home port, current location, current employments (an employment is a complex concept including destination, projected arrival time, purpose, etc.), type and amount of equipment on board, various types of readiness status (personnel readiness, equipment readiness, overall readiness, etc.), and operating characteristics (average cruising speed, maximum speed, fuel capacity, etc.). Other data in this database include detailed information about the characteristics of various types of equipment (e.g., the firing rate of guns) and properties of geographic entities (e.g., for ports, the country they are in, and whether they have a deep channel).

The Navy database provides basic data for systems under development at the Fleet Command Center. This database offers a rich environment for a natural language interface, because the need to explore the database with ad hoc queries occurs frequently.

## THE PARLANCE INTERFACE

The Parlance interface from BBN Systems and Technologies Corporation is an English language database front end. It has a number of component parts: a graphical user interface, a language understander that translates English queries into database commands for relational database systems such as Oracle and VAX Rdb, a control structure for interacting with the user to clarify ambiguous queries or unknown words, and a dbms driver to call the database system to execute database commands and to return retrieved data to the user.

The Parlance system uses several domain-dependent knowledge bases:

1. A domain model, which is a class-and-attribute representation of the concepts and relationships that the Parlance user might employ in queries.
2. A mapping from this domain model to the database, which specifies how to find particular classes and attributes in terms of the database tables and fields of the underlying dbms.
3. A vocabulary, containing the lexical syntax and semantics of words and phrases that someone might use to talk about the classes and attributes.
4. Miscellaneous additional information about how information is to be printed out (for example, column headers that are different from field names in the database).

The Learner is used to create these knowledge bases.

The following queries illustrate the kinds of questions that one can ask the Parlance system after it is configured for the Navy database:

*What's the maximum beam of the Kitty Hawk?*  
*Show me the ships with a personnel resource readiness of C3.<sup>1</sup>*  
*List the ships that are C1 or C2.*  
*Is the Frederick conducting ISE in San Diego?*  
*How many ships aren't NTDS capable?*  
*Which classes have a larger fuel capacity than the Wichita?*  
*How many submarines are in each geo region.*  
*Are there any harpoon capable C1 ships deployed in the Indian Ocean whose ASW rating is M1?*  
*List them.*  
*Show the current employment of the carriers that are C3 or worse, sorted by overall readiness.*  
*Where is the Carl Vinson?<sup>2</sup>*  
*What are the positions of the friendly subs?*

---

<sup>1</sup> Readinesses of various types are categorized C1 (most ready) to C5 (unready), M1 to M5, and so on.

<sup>2</sup> This query is ambiguous. It may be asking for a geographical region or for a latitude and longitude. The Parlance system recognizes the ambiguity and asks the user for clarification.

## THE LEARNER

The Learner is a software tool that creates the domain-dependent knowledge bases that the Parlance system needs. It "learns" what Parlance needs to know from several sources:

1. The database system itself (i.e., the dbms catalogue that describes the database structure, and the values in various fields of the database).
2. A human teacher (who is probably a database administrator, someone familiar with the structure of the database, but who is not a computational linguist or AI expert).
3. A core domain model and vocabulary that are part of the basic Parlance system.
4. Inferences (about such things as morphological and syntactic features) that the Learner makes (subject to correction and modification by the teacher).

Figure 1 shows the input and output structure of the Learner. We call the process of using the Learner *configuring* Parlance for a particular application.

The human teacher uses the Learner by stepping through a series of menus and structured forms. The Learner incrementally builds a structure that can be output as the knowledge bases shown in Figure 1.

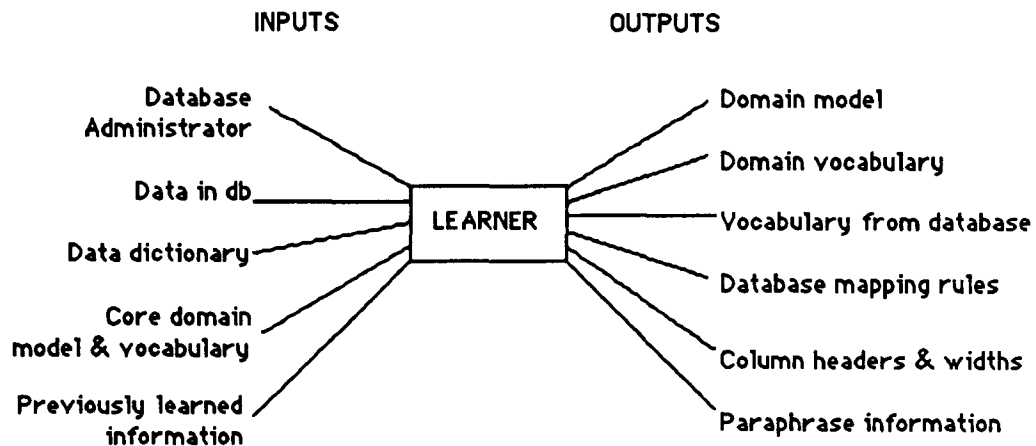


Figure 1. The Structure of the Learner

The teacher chooses particular actions and is led through steps which elicit related information that Parlance must know. For example, when the teacher designates that a particular table or set of tables belong to a class named "ship", the Learner immediately allows the teacher to give synonyms for this class, such as "vessel". The Learner will then infer that the plural form of the synonym is "vessels", instead of making the teacher supply the plural form, although the teacher can easily correct the Learner if the word has an irregular plural.

Whenever information is optional, the teacher can decline to specify it at the first opportunity, and can later initiate an action to provide it. Both required and optional information can be changed by the teacher using the Learner's editing capabilities.

The ability to assign names freely, the freedom to do many operations in the sequence that makes the most sense to the person using the Learner, and the fact that the Learner expresses instructions and choices in database terms wherever possible, make it easy for database administrators who are not computational linguists or AI experts to configure the Parlance interface.

## CONFIGURING PARLANCE

Before the Learner existed, Parlance configurations were created "by hand". That is, highly skilled personnel had to use a separate set of programs (including a Lisp editor) to create the appropriate configuration files.

Figure 2 compares this by-hand configuration process with the first experience using the Learner on the Navy database. The two examples used different databases, but in each case we began with a large set of sample queries in the target domain, and periodically tested the developing configuration by running those queries through the Parlance system. We measured our progress by keeping track of the number of those queries the system could understand as the configuration process went on. Figure 2 actually considerably understates the productivity enhancement realized with the Learner, because the personnel database used for the by-hand configuration was much smaller and less complex than the Navy database.

The Navy database used to test the first version of the Learner was considerably restructured and enlarged, and we had an opportunity to configure Parlance for the newer database. Since we had a new, improved version of the Learner, we chose to configure Parlance to the second version of the Navy database "from scratch", rather than by building on the results of the first configuration. This gave us an opportunity to measure the effort required to use the Learner to do a much larger system configuration, since the size of the target database (measured in terms of the number of fields) had nearly tripled.

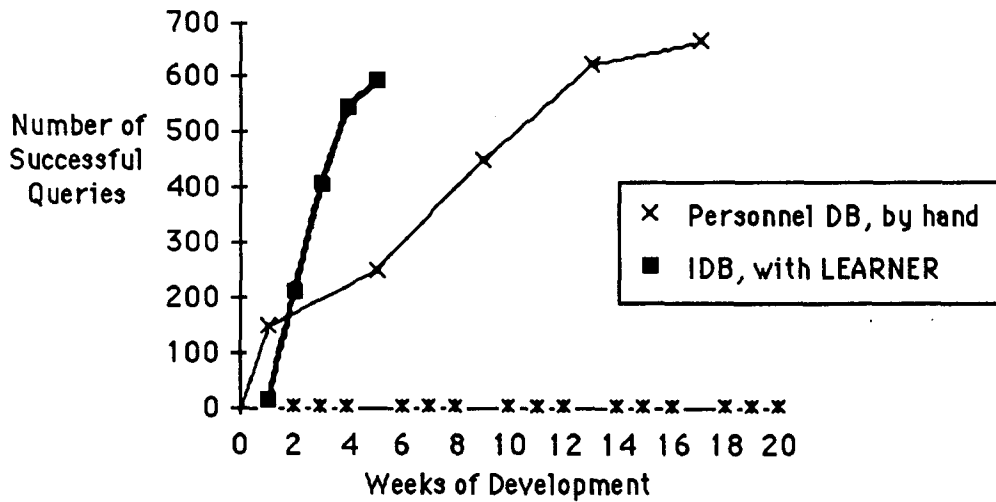


Figure 2. Speed-up of acquisition using the Learner

The results in Figure 3 and its accompanying notes show that the Learner robustly scaled up to the task, and that the time required to perform the configuration increased much less than the number of fields in the database, the vocabulary size, or any other simple metric of size. In fact, for a modest 1/3 increase in configuring effort, a configuration roughly 3 times larger was created.

	<u>Personnel</u> <u>Configuration (0)</u>	<u>1st Navy</u> <u>Configuration</u>	<u>2nd Navy</u> <u>Configuration</u>
Elapsed time	(1)	4 weeks	6 weeks
Total level of effort (2)	(3)	6+ per.wks.	8+ per. wks.
Tables in database	31	32	75
Fields in database	133	231	666
Classes in domain model	218	83	303
Attributes in domain model	316	160	680
Estimated total vocabulary (4)	3000	5500	9800
Root forms (5)	1700	3282	6354
Proper nouns read from db	1170	2656	3907
Verbs	65	6	36
Words with semantics (6)	1600	3326	6073

Figure 3. Comparing the Configuration Processes and Results

Notes to accompany Figure 3:

(0) Changes in the underlying system since this configuration was created make it impossible to measure some of the numbers in this column accurately, so the numbers dealing with vocabulary are estimates.

(1) Records were not kept at the time this configuration was created, but the configuration happened over a period of months.

(2) That this level of effort includes not just time spent using the Learner but also time required to understand the domain, and to do some testing and revision. About 60% of this time was spent using the development version of the Learner.

(3) Records were not kept at the time this configuration was done, but it involved many person-months.

(4) This is an estimate which includes inflected forms of regular words and words that were acquired directly from database fields.

(5) This includes words read from the database and all words directly represented in the vocabulary; it excludes inflected forms of morphologically regular words.

(6) This is a rough measure of the semantic complexity of the domain, since it excludes words that are abbreviations or synonyms.

## CONCLUSIONS

The Learner significantly reduces the time required to create configurations of the Parlance natural language interface for databases with hundreds of fields from months to weeks. This dramatic speed-up in knowledge acquisition scales up robustly, and works as effectively on large databases as it does on small ones.

## REFERENCES

Ceruti, M.G. and Schill, J.P., FCCBMP Data Dictionary, Version 3, Naval Ocean Systems Center, Code 423, San Diego, CA., May 25, 1988.

---

Parlance and Learner are trademarks of Bolt Beranek and Newman Inc. or its subsidiaries. VAX and Rdb are trademarks of Digital Equipment Corp. Oracle is a trademark of Oracle Corp.