DOES PRECIS HAVE FEET OF CLAY? PROBLEMS WITH THE UNIVERSALITY OF THE ROLE OPERATORS

Gillian Michell
School of Library and Information Science
University of Western Ontario
London, Ontario
N6A 5B9

ABSTRACT

One current trend in information retrieval that can facilitate resource sharing and cost reduction is the spread of common indexing languages. Of the more sophisticated languages available, clearly the best known and most widely adopted is PRECIS. One of the claims made for PRECIS by its designer is that the role operators it uses correspond to linguistic universals, making PRECIS eminently suitable for multilingual indexing. This strong claim, if justified, could make PRECIS the common indexing language of choice where translation problems have high priority, as they do in Canada. This paper shows there is little support for this claim to universality because the role operators are defined in terms of superficial characteristics of language that are far from universal.

L'IMPORTANCE DU SYSTEME PRECIS LES PROBLEMES QUE CAUSE UNE UNIVERSALITE DANS UN VOCABULAIRE D'INDEXATION EN RELATION AU ROLE DE L'OPERATEUR

RESUME

Il semble que l'on a tendance à croire que l'adoption d'un vocabulaire d'indexation commun peut faciliter le rendement d'un système de partage de ressources et aussi réduire nos frais. Un des meilleurs vocabulaires disponibles et certainement le plus populaire est PRECIS. D'après les auteurs de ce vocabulaire, une indexation propre à une multitude de langues offre à l'opérateur une immense possibilité d'accès à toutes sortes d'information. PRECIS pourrait-il être la réponse à nos problèmes de traduction au Canada? En réalité cette étude démontre un très faible appui à ce système. Nous avons raison de croire que chaque opérateur possède son propre vocabulaire ayant des caractéristiques qui ne peuvent pas être copiées.

DOES PRECIS HAVE FEET OF CLAY?

Probably the most impressive changes in information retrieval in the seventies have arisen from the development of large, machine-accessed data bases. The importance of these systems in facilitating resource sharing and cost reduction is unquestionable, but some aspects of their design are greatly in need of improvement if their potential is to be fulfilled. The aspect I am concerned with in this paper is indexing language - the language that is used to describe or catalog the documents in the system and then to query the system in order to retrieve those documents in answer to information needs.

The most typical information retrieval system offers its users and indexers a fairly crude language to work with. Most often it consists of simply a vocabulary of allowable index terms, usually called a thesaurus, and a rudimentary method of joining these terms, which is most often Boolean operators. Retrieval using a language of this type is often characterized by precision which is unacceptably low. This is scarcely surprising, since Boolean combinations of terms are no match for sophisticated information requests. The subject of interest in (1a), for example, is hardly well-represented by (1b):

(1)a. Effects of institutionalization on the linguistic development of mental retardates.

b. INSTITUTIONALIZATION and LANGUAGE DEVELOPMENT and MENTAL RETARDATION.

(Even in this example I have been generous in supposing the vocabulary of the indexing language provides for a more or less one-to-one match with that of the query.)

In response to the need for more sophisticated indexing languages, a different kind of indexing language has developed, the string indexing language. A string indexing language is one which gives the indexer rules for stringing together a number of interconnected terms in order to form a complex indexing phrase which will express specifically the subject of the document. Such a phrase might look much more like (1a) and be correspondingly more easily interpreted by the user. A few such languages are Farradane's Relational Indexing (Datta and Farradane 1974) (Craven 1977) Craven's NEPHIS , and Bhattacharyya's POPSI (Bhattacharyya 1975), but by far the best known and most widely adopted is Derek Austin's PRECIS.

PRECIS, like other string indexing languages, allow single-entry specific expression of complex document subjects, but its designer has made an additional claim for it which, if justified, makes PRECIS unique. The claim is that its structure incorporates certain characteristics of language which are fundamental to all human languages. It follows from this that PRECIS should be especially suitable as a language for multi-lingual indexing, a claim Austin and his colleagues have made in a series of articles which appeared in Libri in 1976 (Austin 1976; Sørensen and

Austin 1976a; Sørensen and Austin 1976b; Lambert 1976). In a country like Canada where translation needs often have high priority, this design feature alone puts PRECIS in a favoured position as being a good choice for adoption as a common indexing language. Its basic design could provide a lingua franca for the document collections of today and serve as an important aid in resource sharing.

The question to be answered, of course, is whether the claims made for PRECIS are justified. I will be concerned in this paper with the linguistically-based claims presented in a paper by Sørensen and Austin (1976a) which is subtitled "A Linguistic and Logical Explanation of the Syntax". As a linguist, I have been examining the evidence presented both in the PRECIS documentation and in the linguistic literature, and I have found a number of areas where the linguistic claims have little support. I will address one point, namely, whether the PRECIS role operators are universal in nature; that is, whether they can be looked at as linguistic universals.

A linguistic universal is any characteristic of human language which is found in all languages. Simple examples are nouns - a universal linguistic category or part of speech, and questions - a sentence type that every language has some means of expressing.

The kind of linguistic universal that PRECIS role operators are claimed to exemplify is somewhat more complicated. The idea is that nouns in any sentence (or in a subject description) can be classified into one or other of a small set of general functions. These functions explain what that noun does relative to the action or state described by the verb. These are probably best explained by example. In sentence (2), John could be said to have the function of being the agent - the animate instigator - of the action-verb open. This function

(2) John opened the door with the key.

or role can be called Agent. In the same sentence, key can be seen to have the role of an instrument, something inanimate that is used to carry out the action of the verb. Instrument or Instrumental is a label commonly used for this. It is more difficult to describe the function of door. The function it is usually assigned to is called Objective, which might be characterized as a role identified by the meaning of the verb and most directly affected by the verb. Note that the role Objective has nothing to do with the idea of object, as in direct or indirect object of a verb. Object is a notion that has to do with the grammatical relations of syntax, while Objective is a function that we recognize in interpreting the meaning of a sentence when it is used. Many other sentences match this framework of functions. Sentence (3) is an example. A comparable subject description is (4).

(3) Sam (AGENT) filled out the card (OBJECTIVE) with his pencil. (INSTRUMENT)

(4) Computer (INSTRUMENT) production of horoscopes (OBJECTIVE) by student-programmers (AGENT)

(Note that in (4), the subject description, the verb produce has been nominalized - made into the noun production.) It is roles or functions like these that PRECIS role operators are meant to describe.

Another point about roles that should be noted is that the role a noun takes has nothing to do with whether the noun is the subject or the object or the indirect object of the verb. In sentence (5), the noun in the Instrument role is the subject, while in sentence (6), the noun in the Objective role is subject.

(5) The key (INSTRUMENT) opened the door. (OBJECTIVE)

(6) The door (OBJECTIVE) was opened by John. (AGENT)

As these examples illustrate, a particular role cannot be regularly associated with a particular grammatical relation like subject. It can be shown that subject-object relations are relatively superficial and changeable characteristics of language--in fact, there are languages which do not have sentence subjects or objects at all. Japanese, for instance, lacks the subject relation. Subjects and objects, then, do not meet the criteria for being language universals.

The linguistic approach that talks about the roles of nouns in sentences is known as Case Grammar (Fillmore 1968, Fillmore 1971). Its name notes the relationship between the roles that nouns play and the way grammatical cases are used in languages like Latin or German or Russian to indicate the roles of nouns in sentences (among other things). But the cases or roles being talked about in linguistic theory are "deep" cases; that is, they do not depend on the surface characteristics of particular languages like Russian for their definition but instead represent underlying relationships between nouns and verbs which all speakers are aware of and which are part of our understanding of language. Thus a noun in the Agentive role may be the subject or the object of a verb, it may be in the nominative or the ablative grammatical case, or it may follow the preposition by or through or no preposition at all. It will still have an Agentive role. Grammatical facts do not affect what role the noun has.

Incorporating roles in indexing languages is not a new idea, but claiming universality for these roles and using linguistic theory as

justification for them is. The major problem in using roles in language design and description is in coming up with a well-motivated, well-defined, and workable set of roles. Thus, an attempt like Sørensen's and Austin's to do this for the PRECIS roles represents an important step in indexing language design. Unfortunately, their attempt comes to grief on a number of counts.

The one I will illustrate here is that the definition of some of the major roles ("main line operators" in Austin's phraseology) depend on superficial relations like subject-object. This means that those roles cannot be universal.

The roles in question are role operators (1) and (3), Key system and Agent. The way these two role operators are represented in the PRECIS literature is reproduced in (7).

(7) 1. Key system: object of transitive action; agent of intransitive action.
3. Agent of transitive action; Aspects; Factors.

Two kinds of superficial sentence relations are appealed to here: object, which we have already discussed, and transitivity/intransitivity in verbs.

The superficiality of transitivity is easily demonstrated by the fact that the same verb can be both transitive and intransitive. Consider sentences (8a) and (8b).

8. (a) The library opened on Tuesday.
(b) The mayor opened the library on Tuesday.

In (8a) open is intransitive - that is, it cannot have a direct object following it, while in (8a) open is transitive.

What effect does this definitional problem have on role assignment? Let us contrast the PRECIS situation with the role assignment we would get using the linguistic definitions I presented above for Agent and Objective, reproduced in (9).

(9) Agent: the animate instigator of the action
Objective: the role identified by the meaning of the verb and most directly affected by the verb.

In (8b), mayor has the Agent role, while library has the Objective role. In (8a), library still has the Objective role, and this is as it should be, since the relationship between open and library remains the same in both sentences. It is unaffected either by whether library is subject or object of the verb or by whether the verb is transitive or intransitive.

When we use the PRECIS roles and their definitions, the thinking is different. In (8b), mayor is assigned role operator (3), because it is the Agent of the transitive action open. This accords with our understanding of Agent as a deep case. Library in (8b) is assigned role operator (1), because it is the object of the transitive action open--an assignment that depends completely on the superficial relations object and transitive. In sentence (8a), however, Library does not fit the first definition of role operator (1), since it is not the object of a transitive action. But the second definition is problematic too. While open is intransitive in (8a), Austin's use of agent conflicts with ours. We could decide to allow Library to be an Agent, although its role relationship with open is clearly different from that of mayor and open in (8b). This would give Library the same role operator in both sentences but calls into question what Austin actually means by Agent. An obvious interpretation is that he really means subject. Such a definition would maintain consistency on a superficial level but destroys any claim to universality in terms of deep structure. The only alternative is to say that no role operator fits Library, which, given the extreme ordinariness of the example we are dealing with would surely be inappropriate as well.

The conclusion is pretty well forced upon us that these role operators are not defined in terms of universal characteristics of language and are therefore not universals themselves.

The points I have made here cover only a very small part of what is required to give a thorough critique of the linguistic claims made for PRECIS. A far more extensive presentation of Case Grammar and the linguistic theory of which it is a part are necessary for adequate treatment of a system as complex as PRECIS. An approximation of this is attempted in Michell (1978).

REFERENCES

Austin, Derek. 1976. PRECIS in a Multilingual Context; Part 1: PRECIS: An Overview. Libri 26:1-37.

Bhattacharyya, G. 1975. Fundamentals of Subject Indexing Languages. In Proceedings of the Third International Study Conference on Classification Research, Bombay.

Craven, Timothy. 1977. NEPHIS: A Nested-Phrase Indexing System. Journal of the American Society for Information Science 28:107-114.

Datta, S., and Farradane, Jason. 1974. A Psychological Basis for General Classification. In Conceptual Basis of the Classification of Knowledge, ed. J.A. Wojciechowski, 319-331. Munich: Vertag Dokumentation.

Fillmore, Charles. 1968. The Case for Case. In Universals in Linguistic Theory, ed. E. Bach and R.T. Harms, 1-88. New York: Holt, Rinehard and Winston.

Fillmore, Charles. 1971. Some Problems for Case Grammar. In Report of the 22nd Annual Round Table Meeting on Linguistics and Language Studies, 35-56. Georgetown: Georgetown University Press.

Lambert, Germaine. 1976. PRECIS in a Multilingual Context; Part 4: The Application of PRECIS in French. Libri 26:302-324.

Michell, Gillian. 1978. PRECIS "Role Operators" and the Linguistic Notion of Deep Case: A Critique. Manuscript.

Sørensen, Jutta and Austin, Derek. 1976a. PRECIS in a Multilingual Context; Part 2: A Linguistic and Logical Explanation of the Syntax. Libri 26:108-139.

Sørensen, Jutta, and Austin, Derek. 1976b. PRECIS in a Multilingual Context; Part 3: Multilingual Experiments, Proposed Codes, and Procedures for the Germanic Languages. Libri 26:181-215.