SYNTACTIC AND SEMANTIC RELATIONSHIPS—
or: a review of PRECIS: a manual of concept analysis and subject indexing, D. Austin

P. F. Broxis

PRECIS\(^1\) is probably one of the most important projects in subject indication undertaken during recent years and one which could well influence the future course of librarianship and information science. The background to PRECIS has already been outlined by Austin\(^3\) and it should be sufficient to say here that it was developed at the time when the British National Bibliography started to co-operate with the MARC project and decided that the combination of the chain-indexing procedure with the Dewey Decimal Classification used for the main subject sequence did not lend itself to satisfactory computerisation. However, it should be mentioned that chain-indexing does not necessarily have to be tied to a classification scheme and can be used with alphabetical subject-headings as is the practice for the British Technology Index, where a slightly modified form of chain-indexing is used. The British Technology Index was in fact computerised before the MARC project began.

PRECIS stands for Preserved Context Index System and is an attempt to bridge the syntactic and semantic aspects of indexing through natural language as used in everyday speech. It should be stressed that, despite claims to the contrary which have been made, PRECIS is a system of classification. Confusion arises because we are conditioned to thinking of a classification scheme as being a predetermined notated system, rather than one which exhibits syntactic and semantic relationships in a logical order, be that in the form of a printed classification scheme such as Dewey, or as alphabetical subject-headings. In PRECIS, subject-heading strings are manipulated so as to preserve the full connotation of the main string in each of the references made. This differs from chain-indexing in that successive terms of a heading are not truncated in making references to the next highest term in the string. An example of the way in which this is achieved in PRECIS is:

\[(0) \text{Canada} \]
\[(1) \text{paper industries} \]
\[(2) \text{management} \]

\((0), (1)\) and \((2)\) being role operators used both to determine facet order and to instruct the computer as to the necessary manipulations required to produce the required references as follows:

CANADA
Paper industries. Management

PAPER INDUSTRIES. Canada
Management

and

MANAGEMENT. Paper industries. Canada.

This technique has been termed 'shunting' and the basic mechanism can be represented as:

\[
\begin{array}{cc}
\text{Lead} & \text{Qualifier} \\
\text{Display} \\
\end{array}
\]

and is known as the standard format. Occasionally more than two lines are used to display a complete reference string.

The following table shows the extent of the role operators used in PRECIS:
### Main line operators

<table>
<thead>
<tr>
<th>Environment of observed system</th>
<th>o Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed system (Core operators)</td>
<td>1 Key system: object of transitive action; agent of intransitive action</td>
</tr>
<tr>
<td></td>
<td>2 Action/Effect</td>
</tr>
<tr>
<td></td>
<td>3 Agent of transitive action; Aspects; Factors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A Data relating to observer</th>
<th>4 Viewpoint-as-form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected instance</td>
<td>5 Sample population/Study region</td>
</tr>
<tr>
<td>Presentation of data</td>
<td>6 Target/Form</td>
</tr>
</tbody>
</table>

### Interposed operators

- **Dependent elements**
  - p Part/Property
  - q Member of quasi-generic group
  - r Aggregate
- **Concept interlinks**
  - s Role definer
  - t Author attributed association
- **Coordinate concepts**
  - g Coordinate concept

### Differencing operators

<table>
<thead>
<tr>
<th>(prefixed by §)</th>
<th>h Non-lead direct difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>i Lead direct difference</td>
</tr>
<tr>
<td></td>
<td>j Salient difference</td>
</tr>
<tr>
<td></td>
<td>k Non-lead indirect difference</td>
</tr>
<tr>
<td></td>
<td>m Lead indirect difference</td>
</tr>
<tr>
<td></td>
<td>n Non-lead parenthetical difference</td>
</tr>
<tr>
<td></td>
<td>o Lead parenthetical difference</td>
</tr>
<tr>
<td></td>
<td>d Date as a difference</td>
</tr>
</tbody>
</table>

### Connectives

<table>
<thead>
<tr>
<th>(Components of linking phrases; prefixed by $)</th>
<th>v Downward reading component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w Upward reading component</td>
</tr>
</tbody>
</table>

### Theme interlinks

<table>
<thead>
<tr>
<th>x First element in coordinate theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>y Subsequent element in coordinate theme</td>
</tr>
<tr>
<td>z Element of common theme</td>
</tr>
</tbody>
</table>

Role Operators used in PRECIS
This certainly reveals the thoroughness and care with which Austin has pursued his researches. The use of natural language for subject-headings is not entirely new; what is new, however, is the linking of terms by prepositions and various other parts of speech in such a way that references read almost as sentences:

(1) buildings
(2) damage $v by $w to
(3) frost

which gives the following references:

BUILDINGS
Damage by frost

DAMAGE. Buildings
By frost

and

FROST
Damage to buildings.

One of the omissions from the discussion in this Manual is the extent to which an indexing language should verge into the realms of natural language usage and the extent to which it should go to the degree of refinement found in PRECIS. The British Technology Index, for example, relies on the use of, first, a consistent facet order, and, second, a very restricted number of operators, or role indicators, to express the meaning of both heading and reference relationships. It has to be admitted, however, that the problems of producing a national bibliography covering the entire field of knowledge is rather different from that of an index restricted to a particular subject field. It has already been accepted that the system as used by the British Technology Index would be inadequate for the social sciences, and Coates has already suggested the likely modifications which would be necessary. Nevertheless, Austin's analysis does make a valuable contribution to our understanding of various indexing problems which have not previously been mooted. This, indeed, makes the Manual essential reading for the serious student of indexing methodology.

One of the more controversial aspects of PRECIS is Austin's treatment of phrases. Any compound term can be analysed into a focus and a difference, so that the phrase turbulent flow in pipes is treated as:

\[
\begin{array}{c|c|c}
\text{Turbulent flow} & \text{in pipes} \\
\hline
\text{Difference} & \text{Focus} & \text{Difference}
\end{array}
\]

in pipes being a prepositional phrase. In PRECIS, phrases are always kept as a unit and never factorised as:

pipes — flow — turbulent

where pipes are regarded as the subject of the action flow which is in the turbulent mode.

Differencing operators are outside the main group of operators in that they function semantically rather than syntactically. Lightweight concrete reinforced foot bridges could be coded:

(1) bridges $i foot $i concrete $m reinforced $m lightweight

and would produce entries:

Bridges
Foot bridges
Concrete bridges
Reinforced concrete bridges
Lightweight concrete bridges

with the whole term in its natural language order appearing as a display. It is recommended, however, that consideration should be given to making see also references from one part of a compound to another:

Bridges
See also
Concrete bridges.

It is argued that this will be more economic and make the index easier to consult, especially if there is a likelihood of there being more than twenty-five or thirty displays under a given heading. The disadvantage of such a permissive policy is twofold; first, it can lead to inconsistent indexing decisions unless very great care is taken in checking precedents; and, second, the user is faced with the problem of a variable indexing policy to grapple with when undertaking a search under a number of different topics.

The example on frost damage to buildings already quoted earlier serves to illustrate another feature used in PRECIS known as the predicate transformation. This transformation occurs when an action, operator 3, comes into the lead and the next highest term in the string is coded 2. Operators s and t also produce this format. It is so named because the object upon which an agent is performed is regarded as a predicate.
This results in a turntable effect, to use another of Austin's railway metaphors. This can be illustrated graphically as:

\[
\begin{array}{c}
\text{Lead} \\
\text{Qualifier} \\
\text{Display}
\end{array}
\]

which corresponds to what happened in the example.

Care has to be taken not to confuse whole—part relationships with difference relationships. Management decision making should be treated as:

\[
\begin{align*}
(2) \ & \text{management} = \text{Action} \\
(p) \ & \text{decision making} = \text{Part of action}
\end{align*}
\]

which illustrates the extent of the whole—part relationship in PRECIS; it is used not only of things and their parts but also of actions and their parts or properties. This mechanism is also used for places: Paris being a part of France. Rather more unexpected, however, is the extension of this principle to topics such as:

\[
\begin{align*}
(1) \ & \text{Great Britain} \\
(p) \ & \text{law}
\end{align*}
\]

on the grounds that there is a whole—part relationship between a social system and an abstract concept which should be treated as a subsystem. To have regarded law as an agent would be incorrect in this case:

\[
\begin{align*}
(1) \ & \text{Great Britain} \\
(3) \ & \text{law}
\end{align*}
\]

although this would be the correct analysis in indexing the topic law relating to libraries in Great Britain, which would be coded:

\[
\begin{align*}
(0) \ & \text{Great Britain} \\
(1) \ & \text{libraries} \\
(3) \ & \text{law}
\end{align*}
\]

The logic behind this strategy shows the meticulous attention given to the syntactic side of PRECIS. The corollary of this is the fine distinction made in deriving indexing strings.

Place and location in PRECIS can occur at a number of points in a string and can be introduced by the operators 0, 1, 3, or 5. The following is a brief account of the use of these operators:

0 — used in cases where the relation between locality and the following concept is one of geography

1 — used where locality is the principal or only entity in the subject

3 — used where a place name represents a social entity in the role of an agent:

\[
\begin{align*}
(1) \ & \text{Vietnam} \\
(2) \ & \text{war $d 1961—} \\
(\text{sub 2) } (2) \ & \text{Vietnam war $ 1961—} \\
(s) \ & \text{role $v of $w in} \\
(3) \ & \text{United States}
\end{align*}
\]

This is the least common mechanism for introducing a country.

5 — used when place is used as a study region with the subject per se having a wider application:

\[
\begin{align*}
(1) \ & \text{book trade} \\
(2) \ & \text{management} \\
(5) \ & \text{study regions} \\
(q) \ & \text{Germany}
\end{align*}
\]

which represents a general study of the book trade based on the author's experience of that in Germany—it is not about the management of the book trade in Germany.

While these various roles of place and location have always existed, little attempt has previously been made to identify them systematically and to distinguish between them in practical indexing.

This survey of some of the more important aspects of the syntactic side of PRECIS cannot be left without mentioning that there are a number of devices available for varying a term when it appears as a lead term and for achieving the desired typographical form of the entry. The string already used concerning the book trade will produce entries like:

**Book trade**

**Management—Study regions: Germany**

where the typographical program automatically selects the required type-face and inserts the correct punctuation between terms as required.

Once a string has been established it is allocated a SIN (subject indication number) and this is entered into the computer store. Should the identical string be required for another document, the indexer merely cites the SIN number.

When discussing the difference operators it was shown how see also references were created for terms such as concrete bridges. This is one of the ways in which the syntactic side of PRECIS
is interrelated with the semantic side of the system. Another way in which this is achieved is by what is termed the quasi-generic member, the operator q. Consider:

\[(1) \text{ pests} \]
\[(q) \text{ aphids} \]

aphids are a species of homoptera which in turn is a species of insects. These and other intermediate terms are not included in the main string. Reliance is placed on the semantic side of PRECIS, using q, to display such relationships. There is also a mechanism designed to by-pass an intermediate term should there be no other function for the term than that of a connection link, producing the direct reference:

Insects

See also

Aphids.

The semantic aspects of PRECIS are, in the main, fairly conventional. An attempt is made to relate collateral (or sibling) terms—those having one ancestor in common—as, for example, carpets and rugs, both being floor coverings; but they also share many common attributes and tend to quasi-synonymy. The danger of this kind of relationship is the problem of deciding the bases for establishing collateral terms such as these as there is a large grey area where the views of indexers are bound to differ, making consistency of the semantic structure difficult to achieve.

Two-way relationships, welds/welding and welding/welds, are also displayed in PRECIS. Whilst not doubting the value of such collocation in theory, these particular terms will always file close to each other. Again, there is the problem of control in relating such root words. One solution would be to make such references only where unrelated words file between variants of the same root, as arc and arcs or electron and electronics, which are separated by architecture and electronics respectively, due to alphabetisation.

A feature which has always been a problem in indexing is that of polyhierarchical relationships, term belonging to more than one class. Austin seems to have extended the meaning of polyhierarchical to include terms having more than one synonym (paragraph 31.20), but this is an incorrect use of the term. PRECIS attempts to make all possible links, so that pianos will be related to stringed instruments, keyboard instruments, and percussion instruments—all of which will be related to musical instruments.

In the same way as a SIN is assigned for the syntactic strings, a RIN (reference indicator number) is assigned to the strings required for the semantic structure.

The overriding consideration which the syntactical side of PRECIS, and indeed the system as a whole, brings to mind is that of its economics. No information has yet been released on this aspect of the scheme, with the result that we are without knowledge of its true costs as compared with other possible methods.

The final part of the Manual demonstrates the application of PRECIS to languages other than English, with examples in French and German. Various algorithms of the operators are also given, together with answers to the exercises set at stages throughout the text. Finally, there is an index derived from the numerous examples used throughout the text, but unfortunately this does not extend to include the semantic structure.

In the following paragraphs an attempt will be made to make some general observations. It should be appreciated that any review of a work as important as this Manual can only scrape the surface, with the comment which could be made on many points of detail having to be omitted. For the same reason detailed consideration of the computer coding and operation has been avoided.

An important consideration of PRECIS is that, in its present form, it is designed as a two-sequence system consisting of a sequence of entries, arranged by the Dewey Decimal Classification in the case of the British National Bibliography, or by any other means, as, for example, an arbitrary running number, and a separate sequence of references produced by PRECIS. No facility has so far been provided to permit entries and references to be organised into a single sequence index. There are two probable reasons for this. The first is that the length and structure of the strings produced from the coded instructions would make direct references unwieldy. The second is that the primary string produced would prove unsuitable for use as a main heading and, if so used, would result in very poor collocation of entries arising from the rules of differencing and the effect of having subject-headings presented in a display form.

Although described as one of concept analysis and subject indexing, the PRECIS Manual goes beyond this remit in that it is also concerned with the mechanics of computer assistance in manipulating the syntactical and semantic aspects of
the scheme. The typographical details are also dealt with in depth. These various aspects of the scheme are considered in relation to each other, with the result that the student of this work needs to pause from time to time in order to isolate the different strands which constitute the system. It should be emphasised that, whilst PRECIS is basically designed for computer assisted production, it can nevertheless be produced by manual methods.

Austin sometimes falls into the trap of making claims which are not unique to PRECIS. On page 5 (0.8), for example, in discussing the inconsistencies which can arise between indexers, Austin claims that, 'To a certain extent, however, PRECIS guards against inter-indexer inconsistency by requiring all indexers to test a subject systematically for the presence or otherwise of certain concepts which have known syntactical roles. A subject containing an action, for example, must be tested for the object of the action, since this often determines how the rest of the subject should be handled'. Such a procedure should be routine in any indexing service where, in addition, indexers should continually be consulting their authority files, whatever form these may take. Reliance is also placed on the order of operators for crystallising facet order. There are also elaborate rules for deciding the order of adverbs and adjectives. In practical indexing, however, the order of such terms can sometimes vary, especially when describing an experimental situation where, for example, a steel specimen may either be annealed and then quenched, or quenched and then annealed. At this level of indexing, rules which are too rigid in detail may well prove to militate against the accurate description of a subject. Procedures of this kind do not automatically lead to two indexers reaching identical decisions. A document, for example, may contain a number of action terms, all equally important; one indexer may decide to state all these actions, whereas the other may decide to state only the final action involved. What is a paramount requirement for consistent indexing is a combination of rules for formulating subject-headings, together with a policy relating to the degree of exhaustivity and specificity which the index aims to achieve.

One of the problems Austin had to reconcile in devising PRECIS, and one which could affect its application in practice, is the number of alternatives which appeared possible. The result of this has been that an appreciable amount of space has been devoted to discussing these possibilities and to justifying why only the standard routine should be adopted. The criterion for a good indexing system ought to be how well the principles can stand by themselves with as few exceptions as possible to the general rules. The passage of time will undoubtedly show how well or otherwise PRECIS attains this measure.

References

The Eighth Annual Gale Literary Datebook

The Gale Research Company, publishers of Detroit, suspended publication of their Gale annual literary datebook after the 1971 issue, but have produced an eighth volume to celebrate the American Bicentennial, 1976.

It is a 48-page folio publication consisting of twelve pages devoted to each month with the names of literary personages and their year of birth entered under the appropriate day, and most of the remainder being devoted to a reproduction of the 1748 edition of the beautiful calligraphic engraving by George Bickham of The American Colonies. It is a detailed conspectus of America of his time, being the most complete work of its kind, and combining detailed geographical information with beautifully engraved maps and graceful ornamental calligraphy.

This was a section of The British monarchy, which was Bickham's last major undertaking. Bickham (1684-1769) was the great writing-master of the eighteenth century, and the combination of his beautiful calligraphy with the well-designed calendar has made a very pleasing publication which those who are concerned with celebrating the birthdays of literary personages (both historic and contemporary) will find very useful.