# A Basic Theory of Roles as Syntactical Control Devices in Coordinate Indexes*

By JOHN C. COSTELLO, Jr.
Battelle Memorial Institute, Columbus 1, Ohio
Received September 30, 1963

Regardless of the subject matter treated, narrative written documents other than fiction are prepared to accomplish the delayed, passive, and indirect transmission of knowledge from the author to one or more other individuals who have or will have interest in the content. Written documents of this type are almost invariably prepared in a natural language, such as English, German, Russian, Finnish, Chinese, or Japanese. Only rarely is the original text initially prepared in some code, the components of which represent or approximate the meanings of word symbols for concepts. More frequently, the natural language of the original text is converted later to a code, or encoded, for convenience in transmission, storage, or other manipulation.

This paper is concerned with documents written in natural languages which contain descriptive, or idea information, that is, documents which are narrative, subjective, textual, discursive, or judgmental.[1] The paper is oriented toward nonfictional documents which deal with scientific and technical subjects and which are designed for long-range utility rather than for transient messages of short half-life, such as advisory notices that actions have been completed or will be undertaken. In other words, our concern is with documents such as technical reports, journal articles, and patents, and, specifically, with the indexing of such documents.

The natural languages authors use in recording and describing scientific and technical information must have the capabilities for modifying word-roots (variously called morphemes, bases, or stems) so that differences in meaning or context can be expressed clearly and unambiguously. Such meanings and contexts for verbs are person, number, voice, tense, and mood. For nouns, pronouns, and adjectives these contexts and meanings are number and case, the latter being a means of expressing the relationship which the nouns or pronouns bear to each other and to the action expressed by the verb in the sentence.

With respect to method and capability for expressing different meanings and contexts for roots, languages fall into four classes—inflectional, agglutinative, isolating, and analytical.[2-8]

In inflectional languages, the meanings and contexts of roots are modified by suffixes attached to them. Inflection of verbs is referred to as conjugation, and inflection of nouns, pronouns, and adjectives is referred to as declension. The following paradigm illustrates the conjugation of the Latin verb root *amā-*, to love.

Present Indicative Active

| Singular | Plural |
|---|---|
| Amō (I love) | Amāmus (we love) |
| Amās (you love) | Amātis (you love) |
| Amat (he, she, or it loves) | Amant (they love) |

The declension of the noun *puella*, girl, is illustrated in the next paradigm:

| | Singular | Plural |
|---|---|---|
| Nominative | puella | puellae |
| Genitive | puellae | puellarum |
| Dative | puellae | puellīs |
| Accusative | puellam | puellas |
| Ablative | puella | puellis |

These two paradigms illustrate the accomplishment of changes in meaning in inflectional languages by attachment of suffixes, or case-endings, to roots. The declension of pronouns and adjectives is similar to the declension of nouns, so that a "beautiful girl" as a direct object in Latin is *puellam pulchram*, the adjective agreeing in case and number with the noun it modifies.

Particular note should be made at this point of the five common cases in which a noun may be declined in Latin—nominative, genitive, dative, accusative, and ablative. Two others are used relatively infrequently, the vocative and locative. Latin is considered a highly inflected language, with its five (or is it seven?) cases for nouns. Indo-European, the precursor of most modern European languages, had eight cases,[3] German has four, Russian six, and Finnish fifteen.[7] According to one authority, "The total number of possible grammatical cases has been estimated at thirty-six, which is probably an understatement."[3]

It should be noted that the suffixes, or case-endings, of inflectional languages are not able to stand by themselves as words.

In agglutinative languages, case is indicated by affixing to the root a postposition, which functions as a suffix or case-ending but which can and often does stand by itself as a complete word. Thus, for "walked" the form might instead be "walk-did." Languages of this type include Turkish, Finnish, Hungarian, and some African and native American Indian languages. Declension by agglutination accomplishes the same objective as declension by inflection, that is, indication of the number of

* Presented at Special Interest Group on Information Retrieval, Association for Computing Machinery National Meeting, Denver, Colo., August 29, 1963.

things referred to and the relationship of the nouns to each other and to the action of their controlling verbs. The concept of number is not always as simple as our singular and plural distinction, for some languages have separate case-endings for one, two, three, and four-and-more of a class. Sanskrit and Greek utilized singular, dual, and plural forms, and Arabic still employs the device of dual number.[3]

Isolating languages make no modification of the root in any form. Roots thus are always complete words. The most important language of this class is Chinese. Chinese words are all of one syllable, and all are ideographic or pictographic characters. There are 214 basic characters. These have been combined into some 40,000 compound characters, of which 4,000–8,000 are in general use. Isolating languages have no inflectional or agglutinative endings. What inflectional and agglutinative languages accomplish for their nouns by case-endings or postpositions is accomplished in isolating languages solely by word order. "Position is everything in the life of a Chinese sentence. A word used in the subject must always come before the word used as a verb, with the object following. There are no genders, no endings, no cases, no declensions or conjugations—only individual words."[3]

"All fully developed languages need approximately the same number of permutations of the basic idea contained in a given word, and the devices employed to achieve these effects vary greatly. Latin achieves them largely by means of inflections, whereas English makes wider use of separate auxiliary words—pronouns, prepositions, and such assisting verbs as shall, have, may, etc."[3]

English, French, Spanish, and Italian, as well as certain other European languages, use prepositions extensively to express ideas that are shown by case-endings in inflectional languages and by postpositions in agglutinative languages. Such languages belong to the analytical class. Analytical languages have lost all, or almost all, of the case-endings, French, Spanish, and Italian having retained more of them than English. English still retains -s or -es for most plurals, and -'s and -s' for the possessive, but English uses a great many prepositions where the inflectional languages use case-endings.

"A regular uninflected (or analytical) language like English does its grammatical work almost entirely through its word order. It makes a vast difference whether we say 'The man kills the tiger' or 'The tiger kills the man.' In Latin, the declension of nouns and the conjugation of verbs largely remove the necessity for rigid word order—here a change in word order makes a change in emphasis, while the fundamental idea in the statement made remains unchanged."[3]

Let us summarize the classes of languages before proceeding to develop the idea of roles as case-ending or postposition equivalents:

1. Inflectional languages do their grammatical work largely through root-suffixes known as case-endings which cannot stand by themselves as complete words.
2. Agglutinative languages do their grammatical work primarily through appended root-suffixes called postpositions, many of which can and regularly do stand by themselves as complete words.
3. Isolating languages do their grammatical work exclusively by word order and make no use of inflection, agglutination, or prepositions.
4. Analytical languages do their grammatical work through word order and auxiliary words such as prepositions, pronouns, and "assisting verbs," with minor influence exerted by vestiges of inflection.

Inflectional and agglutinative languages require minimum syntax, or word-order control, since case-ending ground rules permit placement of a noun or pronoun in any relative position in a sentence without creating potential for misinterpretation. Analytical languages *rely heavily* on word-order control to prevent misinterpretation, whereas isolating languages *require* word-order control.

The subject of this paper, "A Basic Theory of Roles as Syntactical Control Devices in Coordinate Indexes" is related to what has been discussed above in the following way:

An inverted coordinate index is a means whereby a searcher for information first goes directly to a card file (or a computer tape) and selects those system records which represent the concepts on which he wants information; second, visually or mechanically coordinates, with the proper search logic and strategy, the concept records; and, third, identifies by coordination of concepts the unique numbers of documents (that is, references) which contain information on the combination of concepts coordinated.[9]

These system records, in the earlier and less sophisticated systems, as well as in the later more sophisticated systems, are dedicated to noun terms, and occasionally to adjective and adverb terms. No verb terms are used, although noun terms ending in -ing, such as GRINDING, have been frequently and erroneously regarded as verb forms rather than as gerunds and hence noun forms. Often called uniterms[10] or descriptors, these terms stand in inverted coordinate indexes by themselves, bearing no functional or contextual orientation. From this viewpoint, the uniterms of earlier forms of coordinate indexes, because of inversion, are equivalent to Chinese ideographic or pictographic characters, lifted as entities unto themselves from the analytical English language text in which they had been used, completely out of context, with no indication of relationship to other concepts.

Although the documents being included in information systems describe technical and scientific research and development programs in English, an analytical language, and although the indicative abstracts prepared to summarize the content of these documents utilize the same analytical English language, the search devices used to provide access to the documents utilize terms which bear no more inflectional or agglutinative relationships to each other than do the completely unitary words of a noninflected, isolating language like Chinese. Thus, documents dealing with tigers killing men are retrieved when searchers want documents on men killing tigers. Realistically, however, the coating of copper pipes with lead and the coating of lead pipes with copper, as well as the lining of polystyrene bottles with polypropylene and the lining of polypropylene bottles with polystyrene, could become sources of annoying and frustrating false

retrieval in coordinate indexes based on the syntactical principles of isolating languages from which word order, the syntactical control, is absent.

Accordingly, information-system practitioners have devised contextual control schemes to compensate for the absence of syntax to permit the representation of noun-terms in inverted coordinate indexes with a simple scheme of case-ending or postposition equivalents, the objective being to permit searching on concepts-in-context so that retrieval of false or nonpertinent references would be minimized. The system which gives the greatest promise of becoming generally adopted is that of the Engineers Joint Council,[11,12] which in reverse chronological order evolved from the system of roles of the American Institute of Chemical Engineers,[13,14] the du Pont system of roles,[15,16] and similar systems in operational use at the Linde Company[17] and Western Reserve University[18,21] and in experimental use earlier in the U. S. Patent Office.[19]

The Engineers Joint Council system of roles provides a type of "case" orientation on technical and scientific contexts. Their inflectional or agglutinative nature, and the definitions of roles for noun-terms, is such that they may be applied equally as effectively to the biological and social sciences as to the technologies and physical sciences.

Reference was made earlier to the fact that the Finnish language expresses noun-relationships in 15 cases. This is illustrated in Table I.[7]

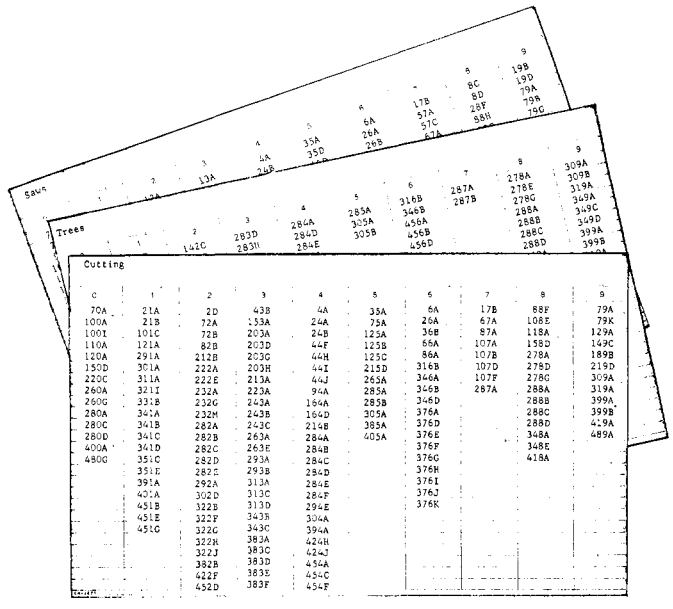


Fig. 1.—Uniterm cards for use as locators of information on "The cutting of trees using saws."

If the language in which the indexing was done were Latin, an indicative abstract statement of document content might be "Succidere serris arbores nuntiatur (narratur)." The terms to be entered in the coordinate index would be "Succidere," "Serris," "Arbores," and

Table I
Finnish Language Declension of Puu (Tree)

| Case | Singular | Plural | Meanings |
|---|---|---|---|
| Nominative | Puu | Puut | The tree(s) as subject |
| Partitive | Puuta | Puita | Some of the tree(s) |
| Genitive-Accusative | Puun | Puitten | Of the tree(s); tree(s) as direct objects |
| Inessive | Puussa | Puissa | In the tree(s) |
| Elative | Puusta | Puista | From the tree(s) |
| Illative | Puuhun | Puihin | To the tree(s) |
| Adessive | Puulla | Puilla | On the tree(s) |
| Ablative | Puulta | Puilta | Motion from the tree(s) |
| Allative | Puulle | Puille | Motion toward the tree(s) |
| Abessive | Puutta | Puitta | Without, or absence of, the tree(s) |
| Prolative | Puitse | Puitse | Motion along the tree(s) |
| Translative | Puuksi | Puiksi | Change of state of tree or trees |
| Essive | Puuna | Puina | Continued state of being of tree(s) |
| Comitative | Puine | Puine | With the tree(s) (accompaniment) |
| Instructive | Puun | Puin | By means of a tree or trees |

Next, let us look at the term records from a straight uniterm system which a searcher might use in searching for information on "the cutting of trees using saws." In Fig. 1, it can be seen that the terms coordinated are *cutting*, *trees*, and *saws*. The accomplishment of the search here is based on the selection of those isolated concept terms for words which had been listed out of context in indexing as representing topics discussed. In uniterm coordinate indexes, there is no attempt to relate noun concepts with respect to each other or to the functions they perform. Thus every concept term has only one system record, and all contexts or functions of the term are implied on the unit record for that term.

the term records onto which the document number is to be entered would appear as in Fig. 2.

An alternative indicative abstract statement *with exactly the same meaning* would be "Succisio serris arborum nuntiatur (narratur)." The terms to be entered in the coordinate index then would be "Succisio," "Serris," "Arborum," as shown in Fig. 3. Both sentences mean "The cutting of trees using saws is discussed"—in the first example, the idea is expressed by the infinitive "succidere" with the word for trees in the accusative case, "arbores." In the second example, the "process" or "operation" noun "succisio" requires that the word for trees be used in the genitive case, "arborum."

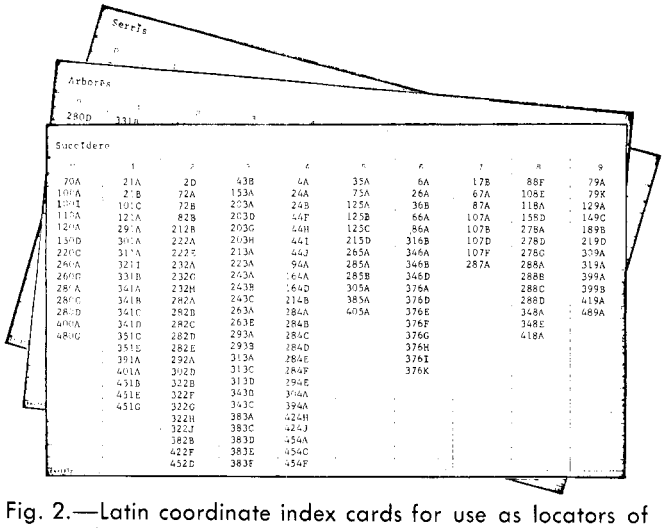

Fig. 2.—Latin coordinate index cards for use as locators of information on "The cutting of trees using saws."

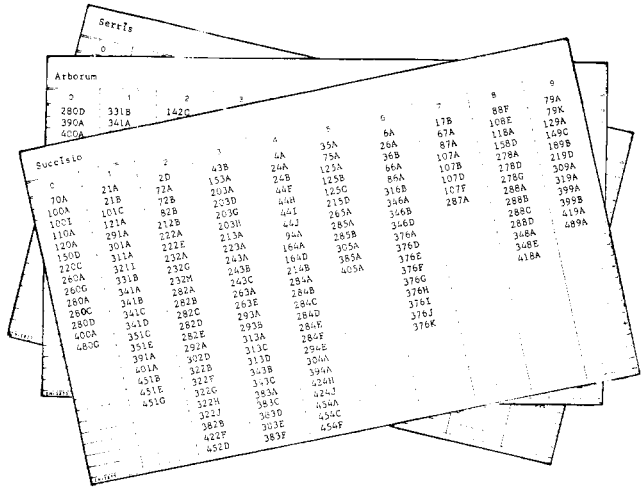

Fig. 3.—Possible alternate Latin coordinate index term cards for use as locators of information on "The cutting of trees using saws."

In these two examples, the Latin word unit concept term records were created directly from the words in the indicative abstract statements, these words possessing the capability of being separated from each other without loss of meaning or relationship because of appended case-endings. The number of system records for every noun could be five if singulars and plurals were to be combined for each case, or ten if singulars and plurals were to be kept separate. (The number could be six or twelve if the less frequently used locative case were included.) Thus, the word "saw," in a Latin coordinate index, would have the term records illustrated in Fig. 4. A searcher for information would have to select the particular term records on which the root serr-, as well as those for arbor- and succid-, bear the case-endings corresponding to the words in indicated contexts in the search question. The term records for serr-, their meanings, and the names of the six cases they represent are shown in Table II.

Earlier we saw that the cutting of trees can be expressed either by "succīdere arborēs" or "succīsio arborum," and that both constructions are correct and accurate. Accordingly, a searcher looking for information on "the cutting of trees using saws" would have to formulate two search questions. One question would require the coordination of succīdere, serra; serrīs, and arborēs, the other the coordination of succīsio, serra; serrīs, and arborum. Even with a highly inflected language, richness of expression would seem to indicate the need for a set of ground rules according to which the coordinate indexing could be done consistently. For the above examples, one suggested ground rule for indexing in Latin would be that indexers should specify or describe the action treated in the document by the present active infinitive as that which is discussed or narrated.

To illustrate further the inadequacy of syntactically uncontrolled inverted uniterm coordinate indexes, let us compare the indexing of a second stylized abstract statement with the one we have been discussing. The second stylized abstract statement is "The document

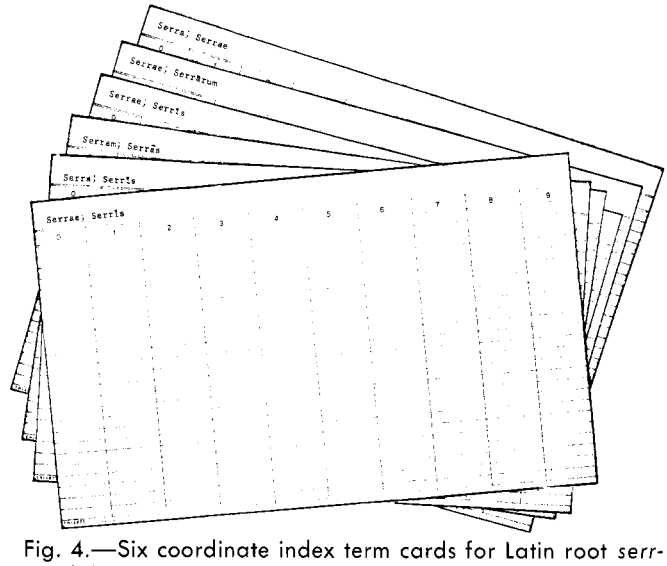

Fig. 4.—Six coordinate index term cards for Latin root *serr-* (English *saw*, N.), one for each of the six most common cases, with singular and plural forms combined.

Table II

| Term record | Meaning of context | Case |
|---|---|---|
| Serra; Serrae | A saw or saws as subjects of sentences | Nominative |
| Serrae; Serrārum | Of a saw or saws, or possessed by a saw or saws | Genitive |
| Serrae; Serrīs | A saw or saws as indirect objects | Dative |
| Serram; Serrās | A saw or saws as direct objects or subjects of infinitives | Accusative |
| Serra; Serris | A saw or saws as objects of most prepositional phrases except those indicating location | Ablative |
| Serrae; Serrīs | At, in, or on a saw or saws | Locative |

| | |
|---|---|
| 8<br>The primary topic of consideration is; the principal subject of discussion is; the subject reported is; the major topic under discussion is; there is a description of | 8 |
| 1<br>Input; raw material of construction; reactant; base metal (for alloys); components to be combined; constituents to be combined; ingredients to be combined; material to be shaped; material to be formed; ore to be refined; sub-assemblies to be assembled; energy input (only in an energy conversion); data and types of data (only when inputs to mathematical processings); a material being corroded | 1 |
| 2<br>Output; product, by-product, co-product; outcome, resultant; intermediate product; alloy produced; resulting material; resulting mixture or formulation; material manufactured; mixture manufactured; device shaped or formed; metal or substance refined; device, equipment, or apparatus made, assembled, built, fabricated, constructed, created; energy output (only in an energy conversion); data and types of data (only as mathematical processing outputs) | 2 |
| 3<br>Undesirable component; waste; scrap; rejects (manufactured devices); contaminant; impurity, pollutant, adulterant, or poison in inputs, environments, and materials passively receiving actions; undesirable material present; unnecessary material present;undesirable product, by-product,co-product | 3 |
| 4<br>Indicated, possible, intended present or later uses or applications. The use or application to which the term has been, is now, or will later be put. To be used as, in, on, for, or with; for use as, in, on, for,or with; used as, in, on, for, or with; for later use as, in, on, for,or with | 4 |
| 5<br>Environment; medium; atmosphere; solvent; carrier (material); support (in a process or operation); vehicle (material); host; absorbent, adsorbent | 5 |
| 6<br>Cause; independent or controlled variable; influencing factor; "X" as a factor affecting or influencing "Y"; the "X" in "Y" is a function of "X" | 6 |
| 7<br>Effect; dependent variable; influenced factor; "Y" as a factor affected or influenced by "X"; the "Y" in "Y is a function of X" | 7 |
| 9<br>Passively receiving an operation or process with no change in identity, composition, configuration, molecular structure, physical state, or physical form; possession such as when preceded by the prepositions of, in, or on meaning possession; location such as when preceded by the prepositions in, on, at, to, or from meaning location; used with months and years when they locate information (not bibliographic data) on a time continuum | 9 |
| 10<br>Means to accomplish the primary topic of consideration or other objective | 10 |
| 0<br>Bibliographic data, personal names of authors, corporate authors and sources, types of documents, dates of publication, names of journals and other publications, other source-identifying data, and adjectives | 0 |

Fig. 5.—Meanings of roles in the EJC system.

discusses saws to be used for the cutting of trees." In Latin indexing, the terms provided for the second statement would be "Serrae," "Succīsionī," "Arborum," quite different from the indexing of the first statement, which was "Succīdere," "Arborum," "Serrīs," while uniterm indexing of *both* statements is trees, cutting, saws. A search in a uniterm index, which would require three term records, would produce references which treat "the cutting of trees with saws" *and* references which discuss "saws to be used for the cutting of trees." In the Latin inverted coordinate index, such false retrieval would not be possible because the searcher would coordinate only those concepts *in contexts* which correspond to the indicative abstract statement he prepared as his search question. For the terms in the two statements, five term records would be required for storage in the Latin index but only three in the English uniterm index.

Coordinate indexing applications to date have been limited largely to collections which contain information on the physical sciences, the fields of engineering, and the biological sciences. The relationships of concepts in coordinate indexes may in part, but only in part, be handled satisfactorily by the cases of highly inflected or agglutinative languages and their meanings. Other relationships of concepts are less grammatical in nature, requiring specification of relationships which might be regarded as synthetic from point of view of prescriptive grammars of established natural languages.

Of the several role systems devised to make it possible to denote relationships or contexts for otherwise isolated concept terms in inverted coordinate indexes, only the one developed by the Engineers Joint Council has achieved a quasi-official status. It will be used in the rest of this paper to illustrate the nature of roles as semi-equivalents to case-endings or postpositions.

In the EJC system there are 11 roles. These are designated by numbers from 0 to 10. The roles serve to designate the grammatical, semi-grammatical, and scientific and technical functions and relationships of word-components of indicative or descriptive abstract sentences. Roughly these correspond to case-endings of nouns in the inflectional languages, or perhaps more closely to the postpositions of agglutinative languages, but like case-endings and unlike postpositions, the roles have no meaning as "words" by themselves.

The complete EJC system of roles is shown in Fig. 5.

First let us consider role 8. This postposition corresponds roughly to "nuntiātur" or "narrātur" in the Latin indicative abstract statement "succīdere serrīs arborēs nuntiātur (narrātur)" (the cutting of trees using saws is discussed). When role 8 is agglutinated (*i.e.*, appended or affixed) to a noun, it imparts to the noun the meaning "is discussed," "is reported," "is recorded," "is described," or "is considered." Thus, succīdere nuntiātur = cutting 8 = cutting is discussed. Good grammatical Finnish requires that the indicative abstract statement read "Kerrotaan että... Kaadetaan" which translates "They say that... are cut." Role 8 equates to the Finnish "Kerrotaan että" which in turn is equivalent to the Latin "nuntiātur."

Role 10 has a function equivalent to the Latin instrumental ablative,[20] the Finnish adessive-genitive-accusative, and the Russian and Indo-European instrumental. It denotes the means of instrument with which, by which, or by means of which something is accomplished. When role 10 is agglutinated to noun X, it imparts to noun X meanings such as "by means of X," "by X," "using X," "with X," "through or through the use of X." Thus serrīs = saws 10 = using or by means of saws. In Finnish "sahan havulla" means "of a saw with the help of" and therefore serrīs = sahan havulla = saws 10 = using or by means of saws.

Role 1 does not have an exact grammatical equivalent in Latin, Finnish, or any other language examined, since by definition it imparts to the noun to which it is agglutinated the meaning of "input or raw material" to a process or operation specified by a noun term in another role in the indicative abstract statement. Role 1 when agglutinated to a noun *implies specifically* that the concept represented by that term is an input or raw material in a process or operation which brings about a change in that concept's *original identity, original composition, original configuration, original molecular structure, original physical state,* or *original physical form.* In the Latin abstract statement used earlier, arborēs = trees 1 = trees as input or raw material to a process or operation in which original identity, composition, configuration, molecular structure, physical state, or physical form is changed. The Latin word "arborēs" equates approximately to "trees as direct objects," since the case-ending identifies the accusative case. However, the accusative case represents *two* contexts which it has been found to be desirable to keep separate in a system of roles for description of scientific and technical information, and hence, role 1 does not equate exactly to the accusative case. (See role 9 below.) Role 1 approximates the Finnish translative case.

The exact equivalent indexing in Latin and English of the example as originally given would be

| | | | |
|---|---|---|---|
| Succīdere nuntiātur | cutting is discussed | cutting | 8 |
| Arborēs | trees as direct object | trees | 1 |
| Serrīs | saws as means of accomplishment | saws | 10 |

Now, extending the original sample abstract statement, let us introduce the idea of a product resulting from the operation of cutting the trees, such as veneers. Now our statement might read "The cutting of trees into veneers using saws is discussed." In indexing, role 2 is agglutinated to the term veneers, since role 2 identifies the concept of product or output of an operation or process in the course of which a term in role 1 has been changed with respect to original identity, original composition, original configuration, molecular structure, physical state, or physical form. In our example, trees went into the sawmill and veneers came out; thus the indexing of this extended abstract statement is:

| | |
|---|---|
| Cutting | 8 |
| Trees | 1 |
| Veneers | 2 |
| Saws | 10 |

Assume that we wanted to specify in the abstract statement that the document indicated that the veneers were used for, or were intended for later use in, plywood, as in "The cutting of trees into veneers, for later use in plywood, using saws, is discussed." Then we would agglutinate role 4 to the term plywood, since role 4 has the contexts or meanings of

to be used
for use
used
for later use
} as, in, on, for, or with

The further extended abstract statement would be indexed as

| | |
|---|---|
| Cutting | 8 |
| Trees | 1 |
| Veneers | 2 |
| Saws | 10 |
| Plywood | 4 |

or, accomplished in another way:

(8) (1) (2)
The cutting of trees into veneers, for later use in
(4) (10)
plywood, using saws, is discussed.

Earlier we discussed the English uniterm and Latin coordinate indexing of the two abstract statements "the cutting of trees using saws is discussed" and "saws to be used for cutting trees are discussed." The following is a comparison of the indexing of these statements using straight uniterms, Latin words with case-endings, and terms with EJC roles.

| Latin | Uniterm | EJC | |
|---|---|---|---|
| Succīdere | Cutting | Cutting | 8 |
| Arborēs | Trees | Trees | 1 |
| Serris | Saws | Saws | 10 |
| Serrae | Saws | Saws | 8 |
| Succisīoni | Cutting | Cutting | 4 |
| Arborum | Trees | Trees | 1 |

Role 4 is somewhat analogous to the Latin dative of purpose.

One can begin to see that roles as arbitrarily defined appendages to terms resemble case-endings and postpositions of inflectional and agglutinative languages, and in certain instances (as role 10) actually equate to case-endings. However, in practice most roles have only *similarity* to case-endings since case-endings are grammatical and syntactical in nature, whereas roles are functionally oriented in nature and are arbitrarily defined to identify meaningful conceptual relationships and contexts in scientific and technical discourse.

Role 0 is used for adjectives and for bibliographic data, such as terms for corporate sources; terms for the personal and corporate names of authors; terms for date (month and year) of publication; terms to indicate type of document; terms to indicate the name of the source journal or magazine; and terms to indicate the volume and number of the source journal or magazine. Thus, the further indexing of the abstract statement as

| | |
|---|---|
| Cutting | 8 |
| Trees | 1 |
| Veneers | 2 |
| Saws | 10 |
| Plywood | 4 |
| November | 0 |
| 1962 | 0 |
| Smith, J. W. | 0 |
| Longitudinal | 0 |
| Journal article | 0 |
| Journal of Lumber Manufacturing | 0 |

represents additional abstract statement concepts and could be read, "The longitudinal cutting of trees into veneers for use in plywood, using saws, is discussed in a journal article by J. W. Smith in the November, 1962, issue of the *Journal of Lumber Manufacturing*." While all the term-roles used in this example were "packed" into one sentence, it could have been broken into two or three sentences, and perhaps should have been, to provide smoother reading.

Adjectives are always indexed in role 0 to minimize the number of term-role records required for this type of term. It is recognized that on rare occasion it is possible to coordinate an adjective in role 0 falsely with a term in another role which it does not actually modify in the abstract statement. The indexing of adjectives in the same role as the nouns they modify, as in the Latin, would eliminate completely this small amount of false retrieval but would result in the creation of many more additional term-role records for adjectives than their retrieval power justifies.

For purposes of further illustration, let us assume that for some reason the operation under discussion was of such a nature that it had to be carried out in a specified gaseous or liquid environment, such as either in nitrogen or water. To indicate this, role 5 would be agglutinated to the terms nitrogen and water to produce the term-roles

| | |
|---|---|
| Nitrogen | 5 |
| Water | 5 |

and then our abstract statement would read "The cutting of trees in an atmosphere of nitrogen or in a water medium to produce... etc." Role 5 corresponds approximately to the Latin locative, the Latin ablative of location, or the Finnish inessive, but *only with respect to materials in which* or *on which* an operation or process is carried out. It is not used to denote geographical location (role 9) nor is it used for the apparatus or equipment in which or by means of which (role 10) the operation or process is accomplished.

If in the process of cutting the trees, a by-product is formed which is regarded by the author of the document or the indexer as being *undesirable* or *unwanted*, such, perhaps, as fine wood dust or a gas evolving from the wood, it would be indexed in role 3 as

| | |
|---|---|
| (Name of Gas) | 3 |
| Wood Dust | 3 |

Role 3 imparts to any term the function or context of undesirable or unnecessary component present—either with imputs in role 1, outputs in role 2, media in role 5, or means of accomplishment in role 10. It is a convenient way to identify materials as impurities, contaminants, adulterants, or pollutants. Used on terms of equipment or devices, it identifies rejects of substandard units. On a term for either a material or device, it can impart the implication of waste or scrap. Role 3 has no counterpart to a case-ending or postposition in the languages examined for this study.

Role 9 is similar to role 0 in that it has more than one context or function which it can impart to a term.

*First*, role 9 is appended to terms representing concepts which are inputs to an operation or process and *which are the same before and after* the operation or

process with respect to original identity, original composition, original configuration, original molecular structure, original physical state, or original physical form. A document which discusses the weighing, packaging, storing, and shipping of veneers to be used for plywood would be indexed

| | |
|---|---|
| Weighing | 8 |
| Packaging | 8 |
| Storing | 8 |
| Shipping | 8 |
| Veneers | 9 |
| Plywood | 4 |

Similarly, the design, development, operation, maintenance, and repair of saws used for cutting trees into veneers to be used for plywood would be indexed

| | |
|---|---|
| Design | 8 |
| Development | 8 |
| Operation | 8 |
| Maintenance | 8 |
| Repair | 8 |
| Saws | 9 |
| Cutting | 4 |
| Trees | 1 |
| Veneers | 2 |
| Plywood | 4 |

In this first context, role 9 equates to the Latin accusative and the English objective for direct object but only when there is no change effected by the operation or process. (See the discussion of role 1.)

*Second*, role 9 can impart to a noun the meaning of possession. In this sense role 9 equates almost completely with the Latin genitive and the English possessive. It is the role equivalent to possessive (not contractive) -'s, -s', and the prepositions of, in, and on when used with a word to show possession. Thus, "the permeability of substrates is discussed" is indexed

| | |
|---|---|
| Permeability | 8 |
| Substrates | 9 |

and "the viscosity of petroleum is considered" is indexed

| | |
|---|---|
| Viscosity | 8 |
| Petroleum | 9 |

Substrates and petroleum take role 9 since they possess, respectively, the permeability and viscosity being discussed.

*Third*, role 9 has the meaning of location, as indicated by the prepositions of, in, on, or at. In the attaching of labels *on bottles*, the detection of radiation *in space vehicles*, the measurement of thickness of plastic coatings *on surfaces*, the detection of carbon dioxide *in mines*, and the production of steel *in Communist China in November, 1962*, all the italic phrases take role 9 in the location sense—Bottles 9, Space Vehicles 9, Surfaces 9, Mines 9, Communist China 9, November 9, 1962 9. In this sense, role 9 equates partially to the Latin locative and ablative of location.

"In" and "on" are prepositions which require special caution in indexing. Cutting wood using saws, transporting oil in pipelines, and shipping lumber on barges all indicate the use of *role 10* with saws, pipelines, and barges—these are means of accomplishment and in this context "in" and "on" equate to "using" or "by means of."

Role 6 and role 7 are the last two roles to consider, and they must be considered together. These have no counterparts in the cases of inflectional or agglutinative languages. They were devised to facilitate the indexing of two concepts when the information discussed the effect of one on the other, a type of investigation reported with great frequency in the scientific and technical literature. Role 6 imparts to a concept term the context that it is the independent or controlled variable, or influencing factor, and role 7 imparts to a term for another concept the meaning that it is the dependent or affected variable, or influenced factor. When one or more terms are indexed in role 6, one or more other terms must be indexed in role 7 to complete the cause–effect, or influencing factor-influenced factor relationship. Conversely, if one or more terms are indexed in role 7, there must be one or more terms indexed in role 6. A discussion of how two concepts simultaneously or alternately affect each other requires that both be indexed in both role 6 and role 7.

The discussion of roles and their contexts in this paper has been limited to the extent necessary to demonstrate their essential nature. More detailed and complete explanations are available in other publications[11, 12] and should be referred to before undertaking coordinate indexing using the EJC system of roles.

## SUMMARY

Documents reporting work in the physical sciences, biological sciences, engineering disciplines, and social sciences which are written in the United States are almost invariably written in English. Indicative abstracts of these documents are prepared in English for use by scientists and engineers in the United States. English is an analytical language, in which syntactical control is accomplished much more extensively by word order, prepositions, and assisting words than it is by inflection, and not at all by agglutination. Simple inverted coordinate indexes, such as those of the uniterm variety, were built on term-records dedicated to words lifted out of syntactical relationship in original documents or their abstracts. The terms in a uniterm index thus resemble individually the characters of isolating languages, such as Chinese, in that the relationship controls, context indication, and functional orientation are lost. Systems of roles have been developed to provide standardized grammatical, quasi-grammatical, and scientific and technical contexts for otherwise isolated concept terms selected to serve as locators of information. From this viewpoint, role systems are synthetic inflectional and agglutinative schemes in which the individual role numbers resemble case-endings or postpositions, and the role definitions can be considered as synthetic "case" definitions.

The ultimate purpose of any system of roles is to limit or minimize the retrieval of nonpertinent references which results from the apparent, but actually nonexistent, relation of two concept terms in indicative abstracts of original documents. The system of roles adopted by the Engineers Joint Council was used for illustrative purposes in this paper, since its soundness of approach has been widely demonstrated, it is most widely known, and it gives the greatest promise of being widely adopted by technical societies and their publications.

## ACKNOWLEDGMENT

The author wishes to express his appreciation to Mrs. Alma Ghilardi, Librarian, and Mr. Daniel K. Dreyer, Translator, both of the staff of Battelle Memorial Institute, for their assistance in gathering background material for this paper and for the Latin and Finnish expressions.

## REFERENCES

(1) J. C. Costello, *Mat. Res. Std.*, **1**, 474, 481 (1961).
(2) "The Columbia Encyclopedia," Columbia University Press, New York, N. Y., 1950.
(3) M. A. Pei, "The Story of Language," The New American Library of World Literature, New York, N. Y., 1960.
(4) H. F. Scott, and W. L. Carr, "The Development of Language," Scott, Foresman, and Co., Chicago, Ill., 1921.
(5) "Collier's Encyclopedia," P. F. Collier and Son Corp., New York, N. Y., 1960.
(6) O. Jespersen, "The Philosophy of Grammar," Henry Holt and Co., New York, N. Y., 1935.
(7) M. A. Pei, "Languages for War and Peace," S. F. Vanni, New York, N. Y., 1943.
(8) L. Bloomfield, "Language," Henry Holt and Co., New York, N. Y., 1933.
(9) J. C. Costello, *Am. Doc.*, **12**, 20 (1961).
(10) "The Uniterm System of Indexing Operating Manual," Documentation, Inc., Washington, D. C., 1955.
(11) "The Engineers Joint Council System of Roles," Battelle Memorial Institute, Columbus, Ohio, 1963.
(12) J. C. Costello, "Training Manual and Workbook for Use in Abstracting and Coordinate Indexing Training Course," Battelle Memorial Institute, Columbus, Ohio, 1963.
(13) R. E. Morse, *Chem. Eng. Progr.*, **57**, 55 (1961).
(14) B. E. Holm, *ibid.*, **57**, 73 (1961).
(15) J. C. Costello, *Am. Doc.*, **12**, 111 (1961).
(16) B. A. Montague, *ibid.*, **13**, 104 (1962).
(17) F. R. Whaley in "Information Systems in Documentation," J. H. Shera, A. Kent, and J. W. Perry, Ed., Interscience Publishers, Inc., New York, N. Y., 1957, pp. 352–383.
(18) J. W. Perry and A. Kent, "Tools for Machine Literature Searching," Interscience Publishers, Inc., New York, N. Y., 1958.
(19) D. D. Andrews, "Interrelated Logic Accumulating Scanner (ILAS)," U. S. Patent Office Research and Development Report # 6, Department of Commerce, Washington, D. C.
(20) E. L. Hettich, and A. G. C. Maitland, "Latin Fundamentals," Prentice-Hall, Inc., New York, N. Y., 1950.
(21) M. R. Hyslop, *Am. Doc.*, **14**, 292 (1963).

---

## ADI ANNUAL MEETING

The American Documentation Institute has announced that its 1964 annual meeting will be held October 4–9, 1964, at the Sheraton Hotel, Philadelphia, Pennsylvania.

## INSTITUTE FOR SCIENTIFIC INFORMATION

In its third expansion move in 3 years, the Institute for Scientific Information moved its headquarters to the Mall Building at 4th and Chestnut Streets in Philadelphia, Pennsylvania.

Publishers of *Current Contents*, *Index Chemicus*, and the *Science Citation Index*, ISI occupies an entire floor at its new location, more than double its former space.

## DOCUMENTATION INCORPORATED TO DEVELOP A PORTABLE READER–PRINTER FOR MICROCOPIES

Award of a contract in the amount of $33,000 to Documentation Incorporated, Bethesda, Maryland, to further development of a portable, inexpensive reader–printer for microcopies, was reported by the Council on Library Resources, Inc.

Reader–printers are projectors which are used for viewing microcopies, but which possess the additional capability of producing enlarged paper prints of selected microimages. Of the reader–printers currently on the market, none is portable and none costs less than approximately $600.

The proposed reader–printer, on the basis of a working model which had been demonstrated to the Council, will be approximately the size of an attache case and weigh about 20 pounds or less. The target price is $100 to $200.

## ENGINEERING THESAURUS

Engineers Joint Council announces the publication of its "Thesaurus of Engineering Terminology" for May 1, 1964, following nearly 2 years of developmental work.

The Thesaurus contains over 10,000 terms which represent selected vocabulary of all engineering disciplines. The terms are arranged alphabetically and displayed to show synonymous, hierarchical, and unspecifiable relationships.

Advanced orders for the Thesaurus at special prepublication rates may be placed with EJC. Orders placed prior to May 1, 1964 are $10.00 per copy, if prepaid; if invoicing is required on shipment, $11.50. After May 1, the prices will be $15.00 and $16.50, respectively; postage extra for foreign shipment. Write to EJC, Dept. T., 345 East 47th St., New York, N. Y. 10017.

## SURVEY OF RESEARCH LIBRARY RESOURCES IN NEW YORK CITY

Recommendations for more effective utilization of reference and research library resources have resulted from a study of library information resources and needs in New York City, it was announced by W. M. Woods, Executive Director, special Libraries Association.

The study, made possible by grants of $16,000 each from the Old Dominion Foundation, New York, and the Council on Library Resources, Inc., Washington, D. C., was conducted by Nelson Associates, Inc., Management Consultants. It was undertaken at the request of a Committee on Reference and Research Library Resources in New York City, comprised of leading librarians representing public, university, and special libraries, which was formed to obtain the grant, sponsor the study, and make recommendations for the implementation of those proposals that seem feasible.

Copies of the report are available from the New York State Library, Albany, N. Y.