The Computational Lexicon: All Lexical Content Is Predicate

Pieter A.M. Seuren
Nijmegen University
Department of General Linguistics and Dialectology
PO Box 9103
6500 HD NIJMEGEN - Holland
email: pseuren@vms.uci.kun.nl

Paper ICLA '94, Penang, Malaysia

In any computational linguistic environment (generation, analysis, translation, etc.) a lexicon L is needed for the language or languages converned. L is a necessary but not a sufficient requirement: other elements such as a pcorresonding grammar G, with which it interacts in all manner of finely tuned ways, are also required.

L is, in principle, a set of ITEMS, each item containing information of various kinds and categories. The information stored in an item falls into three main categories: phonological, grammatical and semantic information. In all three cases, the content and organization of the information will, to a large extent, depend on the theory used.

The phonological information will have to specify the phonological form in which the item appears, so that it can be realized acoustically by means of a rule-governed sound generator.

The grammatical information will be organized according to the grammatical model or theory in terms of which G is cast. In general, the item will have to specify the morphophonemic variants of the item (THREE - THIR-; GO - WENT), its word-class in surface structure, the morphological and syntactic rules or environmental restrictions induced by it, etc. It is also generally assumed that lexical verbs must be fitted out with the appropriate argument frame, in some theories complete with thematic roles.

So far so good, but I want to concentrate on one specific aspect of the semantic information that is required in each lexical item. The questions to do with the representation of lexical meanings are so numerous and so difficult that a full discussion is out of the question. Moreover, any theory of lexical semantics requires the help of neighbouring disciplines, in particular psychology (cognitive science) and its computational cousin AI, and these more often than not fail to provide the necessary support.

What I propose to discuss in this paper is the question of how the grammatical information of an item can be made to fit onto the semantic information, and vice versa.

The semantic and grammatical information must meet at a structurally well-defined interface. This is the question I now want to investigate.

The main point to be made in this context is this: In a computational lexicon it is useful to adhere to the principle that EACH LEXICAL ITEM IS A PREDICATE. This may sound weird, but quite the opposite is the case.Nouns, adjectives, lexical particles, adverbs, prepositions etc. are all best described as predicates in the lexicon. The idea was originally mooted by McCawley in the late '60s (see McCawley 1973), but never fully exploited.

Let us ask first: What is a predicate? A predicate is a classificatory label expressing the fact that a certain category of elements has a certain property. Model-theoretic semanticists say that a predicate is a function from elements to truth-values. If I stick the label APPLE on a thing I classify the thing as an apple, and that may be true or false. The label APPLE is thus a function from things to truth-values, even though, in a noun phrase, it may serve to identify a unique apple.

Some predicates take n-tuples of elements as input. For example, JOHN LOVES JANE contains the predicate LOVE, which assigns to the pair <John,Jane> the property of being a member of the set of pairs such that the first member loves the second. The semantic predicate LOVE comes to the surface as a verb.

That adjectives are predicates is an old insight in the theory of grammar and meaning. In fact, languages differ in whether they treat all or some adjectives as verbs or as proper adjectives. Even within English one finds pairs like the verb SQUINT and the adjective CROSS-EYED, or the verb LIMP and the adjective LIMP.

The elements that are classified by a predicate need not always be things (individuals). The negation (NOT), for example, is a predicate over possible facts: if a possible fact is real it does not deserve the predicate NOT, but if it is unreal it is true to say that it 'NOTs'. Note that in some languages, such as Finn~~m~~ish and related languages, the negation does indeed occur in sentences as a verb, with verb endings etc. In Finnish one does not say JOHN DID NOT LOVE JANE, but something similar to 'John "not-ted" to love Jane'.

Logical quantifiers (SOME, A(N), ALL, MOST, FEW, HALF, etc.) are higher order predicates: they say something not about pairs of individuals but about pairs of sets of individuals. A(N), for example, says that the two sets have a non-empty intersection: JOHN BOUGHT AN APPLE says that the intersection between the set of apples and the set of things that John bought is non-empty. JOHN BOUGHT ALL APPLES says that the set of apples is included in the set of things John bought. Etc.

Prepositions are predicates. THE BOOK IS ON THE TABLE is read semantically as 'the book "on-s" the table', with the two-term predicate ON. And HE FELL ON THE GROUND is analysed as 'his falling "on-ned" the ground'.Conjunctions are analysed similarly, except that the object term is an embedded S-structure, like the subject term. JOHN LEFT WHILE SHE SANG is analysed as 'John's leaving "whiled" (=partially coincided with) her singing'.

Likewise for manner adverbs: JOHN DROVE CLUMSILY says that John's driving was clumsy.

It thus appears that indeed all lexical content words can be regarded as predicates at some deep level of analysis. The advantage is, prima facie, the streamlining of any semantic calculus one wishes to develop for the processing of linguistic information. More precisely, the reduction of lexical content words to predicates forces one to adopt a method of grammatical analysis that reduces surface sentences to logical structures in the language of modern predicate calculus. It turns out, moreover, that the type of grammar required in this context is eminently suitable for the specification of syntactic structures and is thus a strong candidate in the arena of syntactic theories.

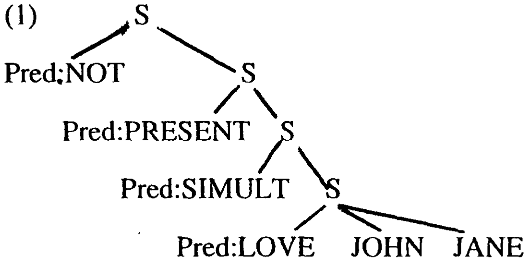
One immediate obvious consequence is thus that the grammar G must provide a precise formal specification of how the actual surface sentences correspond to their semantic analyses. A few questions present themselves. The first question is how certain predicates come to function as nouns, others as adjectives, or prepositions, etc. Let us therefore first have a look at the internal structure of a definite noun phrase like THE DOG. Let us read this as 'the x such that x is a dog'. Model-theoretically, the phrase 'x is a dog' denotes the set of all dogs. The definite operator 'the x' is a function from sets to individuals: it picks out one specific dog from the set of dogs, the dog, that is, which the ongoing discourse is about. (Clearly, situational or discourse information is required for this function to work, which makes it a difficult one for standard model-theoretic semantics. And it is used by discourse semanticicts as an argument for their more cognitively oriented approach.)

What is relevant here is the fact that DOG functions as a predicate in the definite noun phrase THE DOG. Its subject term is not itself a denoting phrase but a variable, so that the procedure of reducing all lexical content words to predicates is not circular.

More generally, we say that it is the job of the grammar G to ensure that all 'abstract' predicates are assigned the proper surface category in the language concerned.

This brings us to the second question. This regards the machinery of G. The condition that all lexical content words are predicates at the level of semantic analysis imposes a heavy constraint on G: it forces G to work with semantic analyses that consist largely of hierarchically ordered S-structures, with the (abstract) predicate on one side and one or more embedded Ss as argument terms (next to possible definite NP-terms). Thus, for example, a sentence like JOHN DOES NOT LOVE JANE will be represented at the level of semantic analysis as the right-branching tree structure:

(1)



This kind of analysis makes it possible to represent meanings of sentences in terms of LOGICAL STRUCTURE. The structures known from modern Predicate Logic, though usually written as bracketed string formulae, are in fact tree structures of the kind shown above. This in itself is an immense advantage, as logical structures are champions of computability, far superior to surface structures of natural language sentences.

It is, moreover, possible to formulate context-free rewrite rules that define the right semantic analysis structures (SAs) for the language concerned. These SAs are then fed into a transformational G-machinery which turns them in a systematic and orderly fashion into surface structures. This type of G is called SEMANTIC SYNTAX, (SeSyn) developed by the author over the past few decades. Apart from the context-free Formation Rules, which generate the SA-structures, the rules of SeSyn are transformational. They come in two kinds, the cyclic rules which operate on successive S-cycles from the bottom up, and the postcyclic rules which bring about further adjustments.

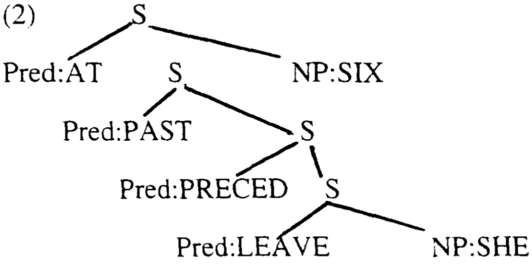
One salient feature of SeSyn is the fact that a number of transformational rules, especially those in the cyclic part of the grammar, are induced ('triggered') by the predicate of the S-cycle concerned. For example, the English predicate (verb) EXPECT allows for the cyclic rule of SUBJECT RAISING (SR). This rule looks for the subject term of the embedded object clause. It then places this subject term in the position of its own S, which is moved one position to the right and is demoted to the status of /S (=VP). This gives rise to sentences like JOHN EXPECTED BRAZIL TO WIN THE CUP FINAL.

The point is now that if all sorts of abstract operators like NOT or quantifiers or adverbial or prepositional adjuncts are treated as predicates (with or without an extra object term), they may be made to induce cyclic rules in precisely the way it was demonstrated for the verb EXPECT. This not only brings about a drastic streamlining of the grammatical machinery, it also creates the structural space for different scope assignments of adverbial, prepositional and other adjuncts. Consider the example SHE HAD LEFT AT SIX. This sentence is ambiguous between a reading where at six o'clock the situation was such that she had already left, and one where her actual leaving took place at six. Note that the sentence AT SIX SHE HAD LEFT only has the former readings and lacks the latter.

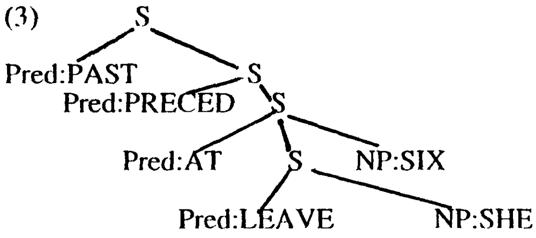
This difference is expressible in terms of a scope difference between two positions of the prepositional adjunct AT SIX: it either sits at the top of the SA-structure, or it sits further down in the SA-tree, just above the S containing the lexical predicate LEAVE. In the former position it results in the first reading given (the one exclusively held by AT SIX SHE HAD LEFT). In the latter, lower position it gives the reading where the actual leaving took place at six.

The SeSyn grammar G ensures that the former reading, with AT SIX at the top of the tree, may result in both SHE HAD LEFT AT SIX and AT SIX SHE HAD LEFT, whereas the latter reading, with AT SIX further down in the SA-tree, can only result in SHE HAD LEFT AT SIX. This makes the sentence SHE HAD LEFT AT SIX ambiguous between the two readings. The relevant SA-structures are as follows:

(2)



(3)



This is only one example of the possibilities of semantic analysis afforded by the SeSyn model of syntactic description. The remainder of the available time will be devoted to a demonstration of the machinery of SeSyn as applied to English. (The Prolog implementation was made by Henk Schotel, assistant professor in the department of Philosophy of Language at Nijmegen University).

It is important to keep in mind that this machinery requires as a necessary condition the treatment of all lexical content words as predicates.

References:

McCawley, James D. 1973: Grammar and Meaning.
Taishukan Company, Tokyo.

Seuren, Pieter A.M. (in preparation), Semantic Syntax.
Blackwell, Oxford.