

## ESPERANTO INFLECTION AND ITS INTERFACE IN HPSG

Wei Li

Department of Linguistics  
Simon Fraser University

### 1.0 INTRODUCTION

We give the outline of Esperanto morphology first. Then we show the Esperanto inflection analysis in feature structures. This serves as the basis for our research on the interface between morphological analysis and sentential parsing in the framework of Head-driven Phrase Structure Grammar (Pollard & Sag 1994).

#### 1.1. Outline of Esperanto Morphology

An Esperanto notional word consists of two parts: stem and inflection.

word → stem-inflection  
 stem → (prefix\*) - compound - (suffix\*)  
 compound → (root\*) - root  
 inflection → (participle) - class - (number) - (case)

Note: as usual, ( ) for optionality, \* for n times repetition ( $n \geq 0$ ).

For this paper, we only focus on inflection morphology. Therefore, we assume that the Esperanto lexicon contains stems and inflection morphemes. As a planned language, Esperanto is totally allomorph-free. Due to agglutination, Esperanto inflection is very productive. Based on the 17 inflection morphemes, 106 morphologically legitimate word forms can be produced for each stem: 42 verb forms, 28 noun forms, 28 adjective forms and 8 adverb forms (Li 1994).

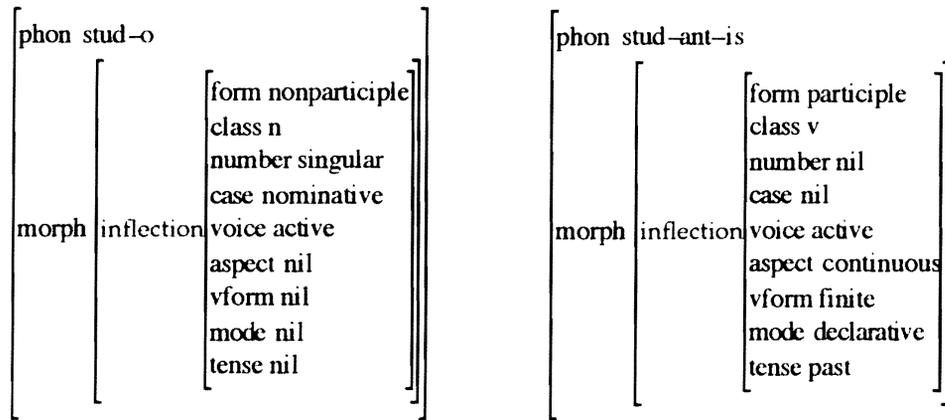
#### 1.2 Analysis of Esperanto Inflection Morphology in HPSG

Using morphology MORPH | INFLECTION features, Esperanto inflection can be described as follows (Li 1994).

- |      |       |   |
|------|-------|---|
| (1)  | -o    | [class N]   |
| (2)  | -a    | [class A]   |
| (3)  | -e    | [class ADV]   |
| (4)  | -j    | [number PLURAL]                                     |
| (5)  | -n    | [case ACCUSATIVE]                                   |
| (6)  | -i    | [vform INFINITIVE] [class V]                        |
| (7)  | -u    | [mode IMPERATIVE] [vform FINITE] [class V]          |
| (8)  | -us   | [mode SUBJUNCTIVE] [vform FINITE] [class V]         |
| (9)  | -as   | [tense PRESENT] [vform FINITE] [class V]            |
| (10) | -is   | [tense PAST] [vform FINITE] [class V]               |
| (11) | -os   | [tense FUTURE] [vform FINITE] [class V]             |
| (12) | -ant- | [aspect CONTINUOUS,voice ACTIVE] [form PARTICIPLE]  |
| (13) | -int- | [aspect PERFECT,voice ACTIVE] [form PARTICIPLE]     |
| (14) | -ont- | [aspect TOBE,voice ACTIVE] [form PARTICIPLE]        |
| (15) | -at-  | [aspect CONTINUOUS,voice PASSIVE] [form PARTICIPLE] |
| (16) | -it-  | [aspect PERFECT,voice PASSIVE] [form PARTICIPLE]    |
| (17) | -ot-  | [aspect TOBE,voice PASSIVE] [form PARTICIPLE]       |

The implementation of the automatic analysis of Esperanto inflection is not difficult. In 1986, we implemented a procedural analysis algorithm in BASIC (Li 1986). In 1994, we designed an

HPSG morphology parser. Compared with sentential parsing, morphological “parsing” is trivial, but the principles and methodology are similar. With the input of *stud-o* (study: noun) and *stud-ant-is* (was/were studying), for example, the Esperanto inflection morphology parser will output the following results in HPSG feature structures:

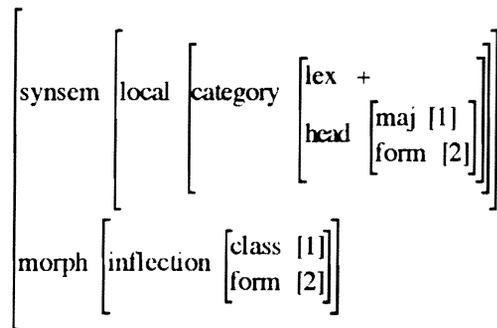


## 2.0 INTERFACE BETWEEN MORPHOLOGICAL ANALYSIS AND SENTENTIAL ANALYSIS

### 2.1 Why Interface?

The output of the morphology analysis will serve as input to the syntax and semantics. But the morphology feature structures defined above are not readily useful for sentential parsing. We need to determine the syntactic/semantic properties of the processed word and build its “synsemic” features. In other words, the system requires a proper interface of information flow.

Some information from inflection morphology gets spread to the syntax for every word. This is the most general information flow in the interface.



But that is far from enough for HPSG sentential parsing. First, the key to the syntactic analysis is the expectation features like syntactic [SUBCAT] for potential head-complement structure and [MOD] for potential head-modifier structure. Second, in addition to syntactic structure, parsing a sentence involves building its semantics based on the semantics of each individual word. All of this information must be supplied before it can be passed to the next phase for parsing.

### 2.2 Derivation and constraints

In light of the philosophy of generative grammar, all of the morphological information given in the lexicon can be seen as an *underlying representation* (UR). From UR, we can derive the *surface representation* (SR) to feed the syntax and semantics. This process is triggered by the result of the morphological analysis. The derivation from UR to SR reflects the information flow in the interface.

HPSG is a non-derivational linguistic theory. The conventional process of derivation will be reformulated as a set of feature constraint rules. Those rules are based on the UR information from the lexicon. (For the convenience of presentation, we will still use the terms *underlying* and *surface*.) Before we define the interface constraint, we need first to design the UR for Esperanto.

### 2.3 Underlying Structures in the Esperanto Lexicon

The underlying representation in the lexicon consists of two types of information.

(1) For each stem, there is an underlying syntactic category based on the meaning of the stem (Kalocsay & Waringhien 1958). The underlying categories MORPH | MAJ are V for action or change, N for things, and A for quality (of V or N). As we know, a stem of any underlying category may surface as a verb, a noun, an adjective, or an adverb due to the highly regular inflection morphology of Esperanto, e.g.

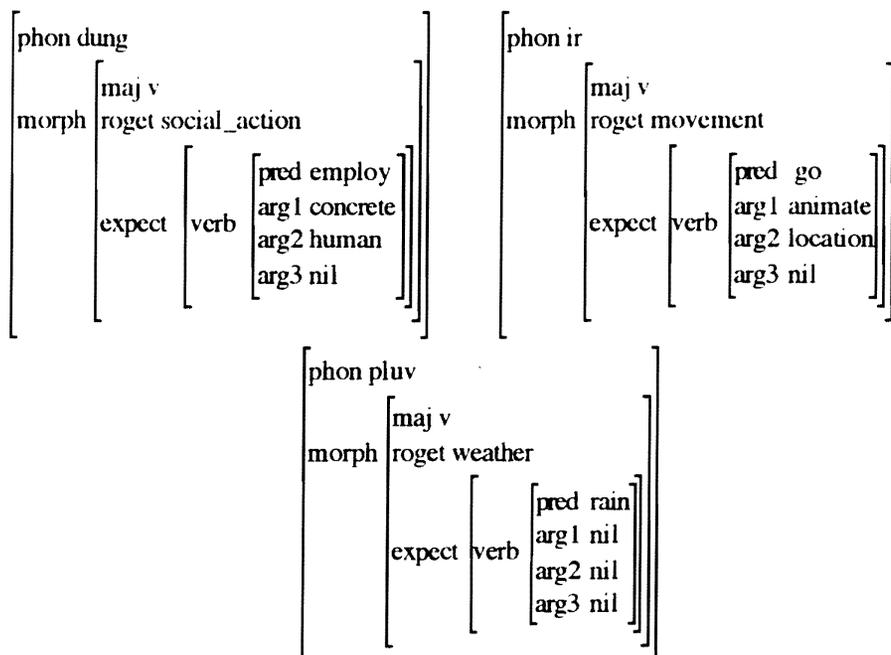
N → N: bicikl-o	(bicycle)	N → V: bicikl-i	(to bicycle)
N → A: bicikl-a	(of bicycle)	N → Adv: bicikl-e	(by bicycle)

(2) The lexicon also contains the underlying semantic expectation features MORPH | EXPECT. The [EXPECT] features are based on the *selectional restriction* of a predicate concept on the semantic classification of its arguments. The semantic classification information is contained in the MORPH | ROGET feature (named after Roget's Thesaurus). The proposal of [EXPECT] and [ROGET] are significant and have various applications in language processing (Li & McFetridge 1995). In fact, it represents some of our common sense knowledge. This type of knowledge serves two purposes for Esperanto analysis. First, the syntactic expectation features [SUBCAT] and [MOD] are based on this semantic [EXPECT]. Second, the selection restriction defined in the [EXPECT] contributes to the building of semantics and the disambiguation work in parsing.

### 2.4 Sample entries of Esperanto lexicon

All the underlying information is stored in the lexicon. We list below some sample entries for underlying verb, adjective and noun.

#### (1) Verb



We assume that the maximum valency of arguments for a predicate is 3. The interpretation of arg1, arg2, or arg3 as semantic roles is decided by the predicate. Typically, arg1 is the subject, arg2 direct object, and arg3 indirect object.

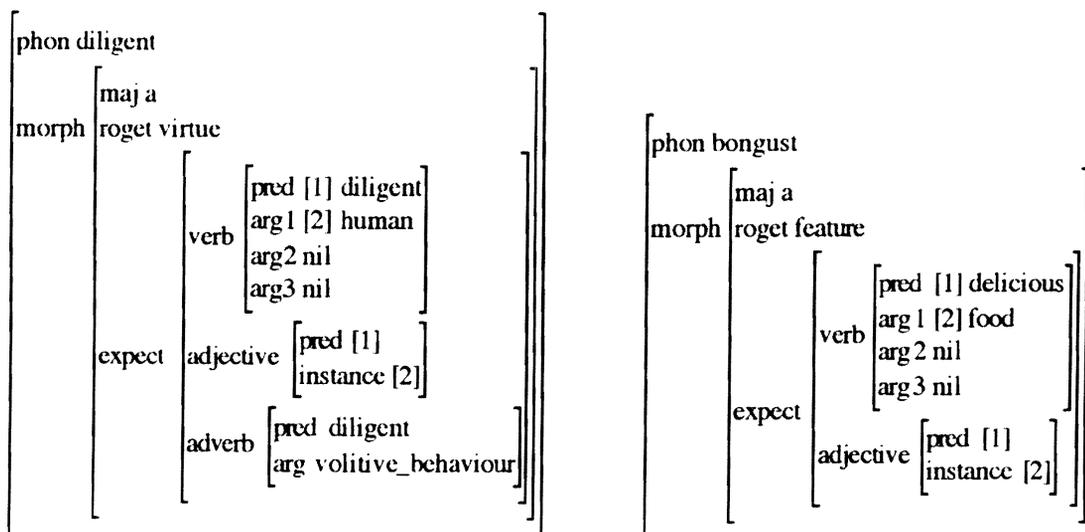
The notion *dung-* (employ) expects a concrete subject argument (arg1: employer) and a human object argument (arg2: employee). This type of information reflects our conceptual world, which is linguistically modeled and lexically encoded as is.

The notion *ir-* (go) is also a predicate taking two arguments, but the second argument is not a usual object but a destination. (At surface level, the role of destination in Esperanto may take different forms, either a prepositional *al*-phrase or an adverb in accusative case.)

The last verb *pluv-* (rain) is a 0-argument predicate. It also surfaces as subject-less sentence, not like English taking a pseudo-subject *it*:

Pluv-is eg-e.  
rain-ed great-ly: *It* rained heavily.

## (2) Adjective

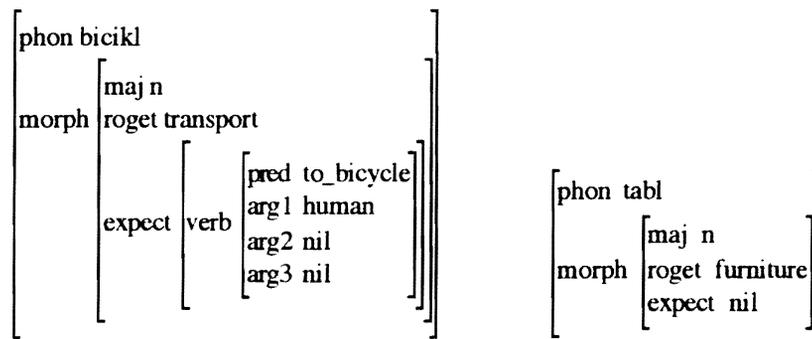


For the underlying adjective, the adjective semantics is equal to the verb semantics (co-indexed by [1] and [2]), as seen clearly in the following pair.

- (i) *La manghajh-o est-as bongust-a.*      (ii) *La manghajh-o bongust-as.*  
the food is delicious.                      the food delicious (as a verb)

*bongust-* (delicious) describes and expects food. *diligent-* (diligent) is a notion of human virtue which therefore can modify either a human being (e.g. John, that girl) or a human volitive behavior (e.g. study, work).

## (3) Noun



*bicikl-* (bicycle) has a related verb notion requiring a human subject; but *tabl-* (table) can hardly be associated with any conceivable verb notion (at least to most minds). So only *bicikl-* can surface as a verb when the subject is human:

Li bicikl-as rapid-e.  
he bicycle-s fast.

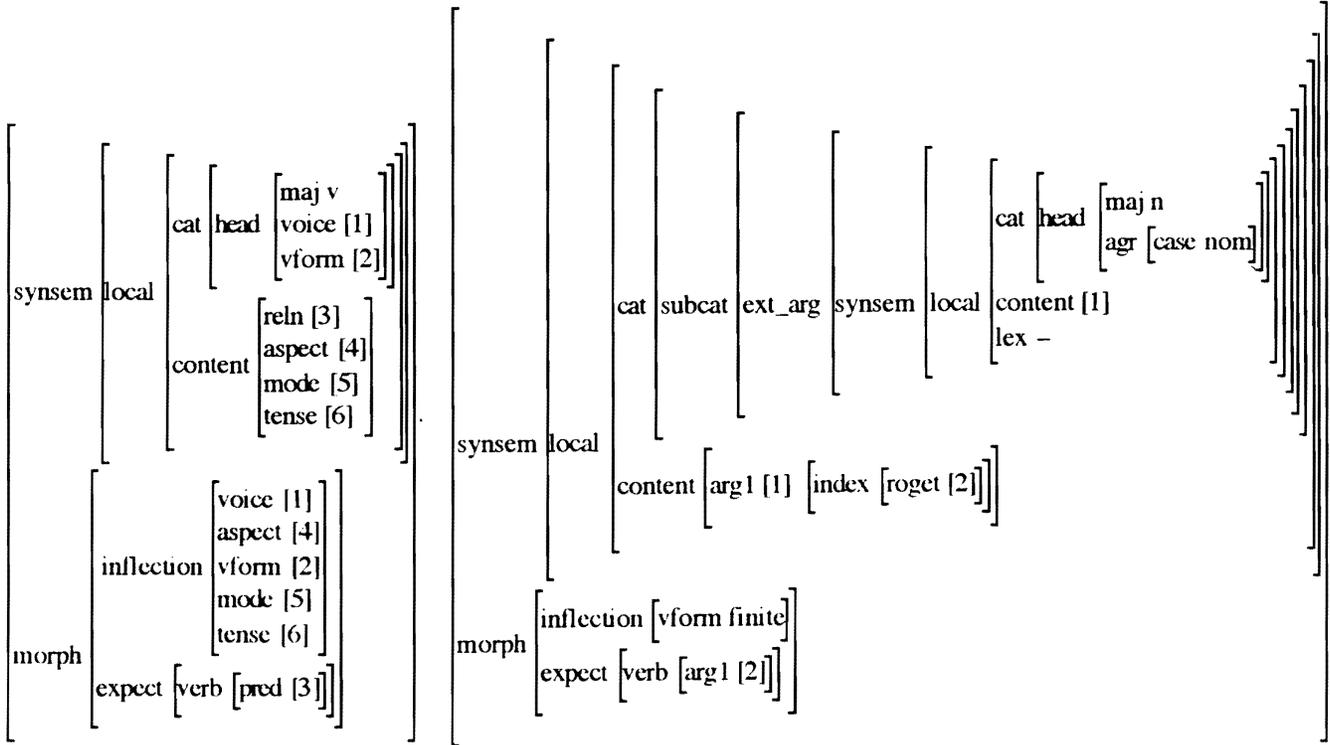
With the exception of nouns, all other categories (V, A, Adv) may have corresponding underlying expectation semantics. This is understandable, for in our present HPSG framework of semantics, noun is the only category not taken as a predicate (semantic head); it is only an argument. Verbs and adjectives expect (and take) nouns. Adverbs expect verbs or adjectives. Nouns expect nothing; they are categories which are expected. Expectation is unidirectional from the head, hence “head-driven”.

### 3.0 INTERFACE CONSTRAINT RULES

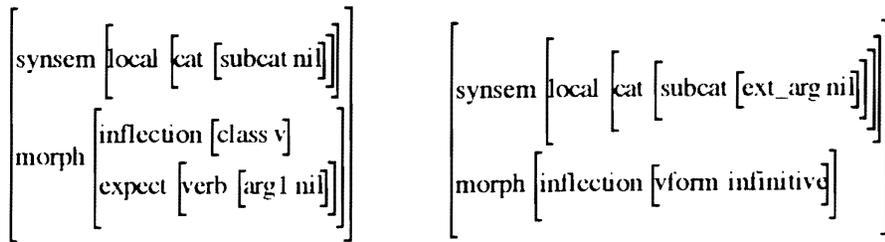
In the following, we only show some basic rules. The complete set of the interface constraint rules for Esperanto is defined in our implementation.

#### 3.1 Verb constraint

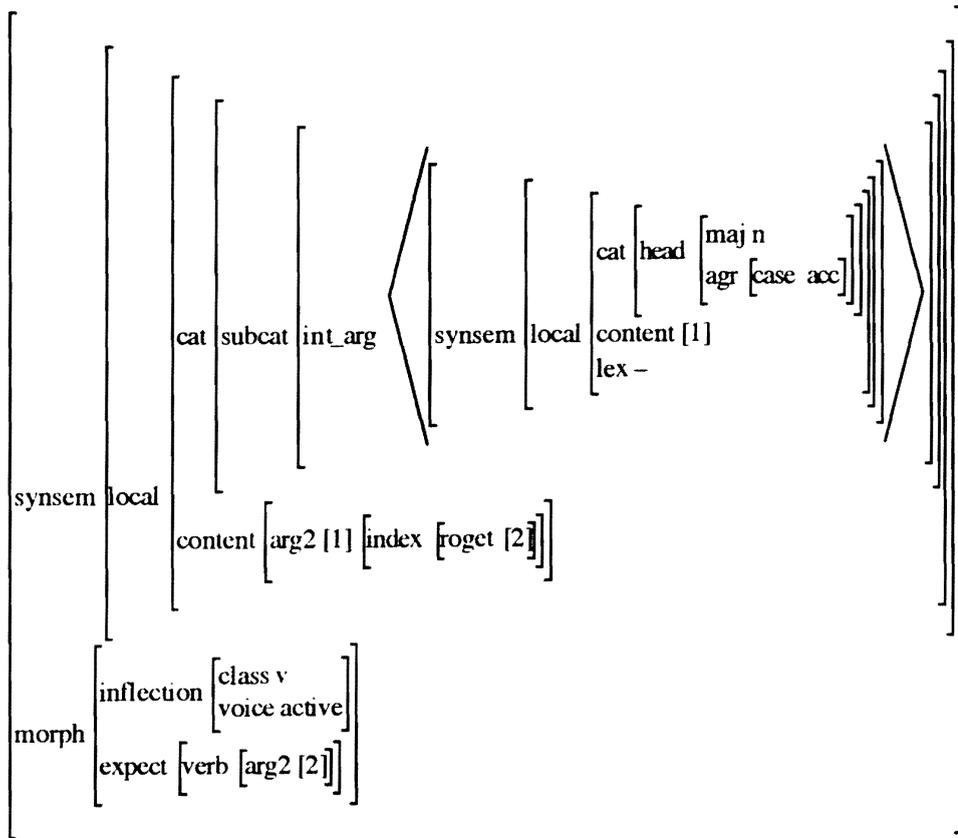
For a verb, the underlying verb semantics surface as the semantics of the word. The verb-related inflection features [VFORM] and [VOICE] are considered as syntactic head features while [TENSE], [MODE] and [ASPECT] contribute to the semantics, so we have the following feature constraint (next page, left structure). A finite verb subcategorizes for a nominative NP as its subject (next page, right structure).



But a 0-argument verb like "pluv-as" (rain-s) has no subject in Esperanto (below left). For an infinitive, the underlying subject is not syntactically required (below right).

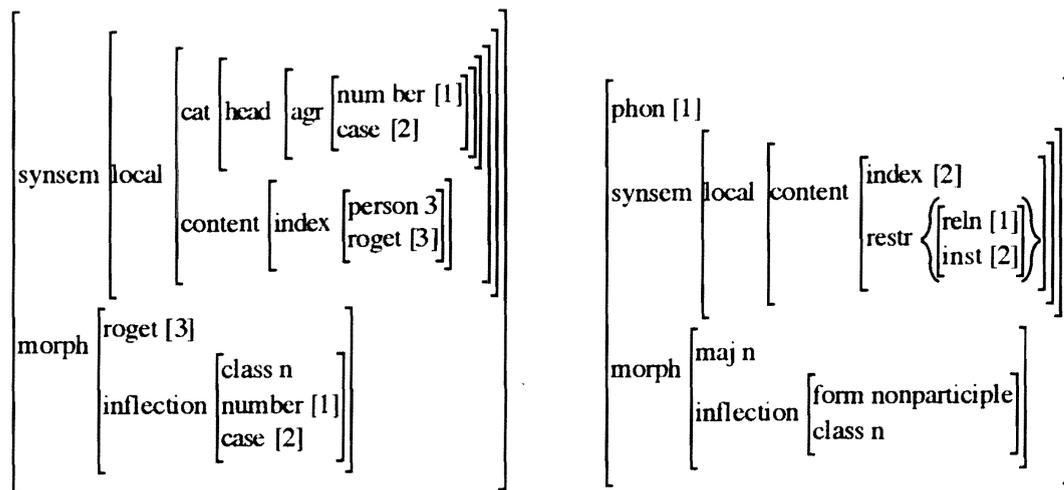


The syntactic property for an active transitive verb would be to subcategorize for an NP in accusative case as its complement (see structure on next page).



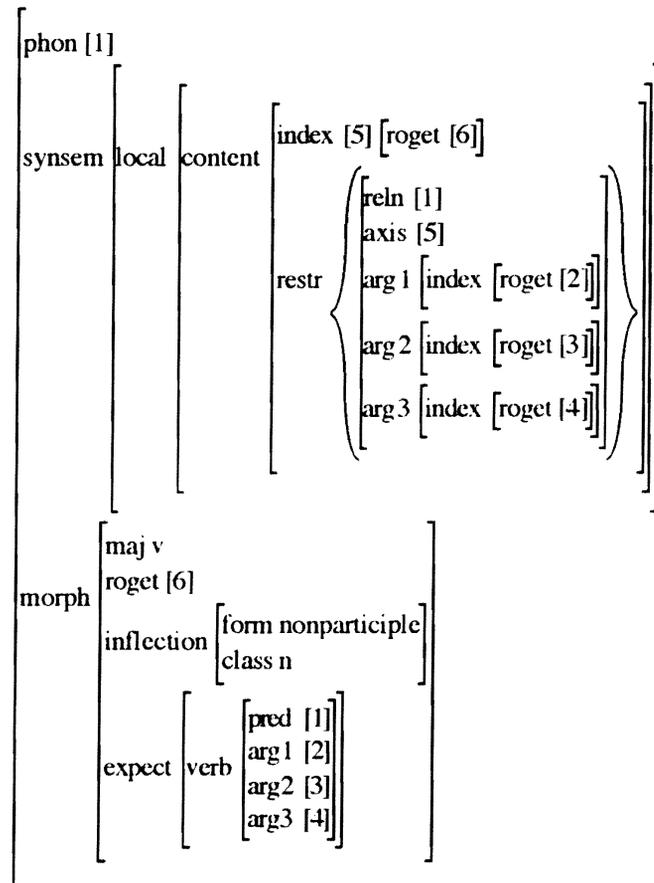
### 3.2 Noun constraint

The inflectional features [NUMBER] and [CASE] get into the syntactic agreement feature because in Esperanto, the noun agrees with the adjective in number and case. The semantic classification [ROGET] is put into the [INDEX] feature to be used for selectional restrictions (Li & McFetridge 1995) (left). The semantics for a noun standing for the object in the physical world is represented by the index plus its restriction (right).



Things are different for a noun derived from an underlying verb or adjective, for it no longer refers to an object in the world, but stands for an action or quality. As a surface noun, it plays a structural role as other nouns do, but as an underlying verb or adjective, it still retains its verbal (or

adjectival) relation as a predicate with some or all of its arguments often unrealized at surface level. To accommodate this dual function in our semantics, we propose that the restriction in the semantics for the deverbal noun (or de-adjectival noun) equal the underlying verb (or adjective) semantics. The relation between the index and the restriction is represented by a feature [AXIS].

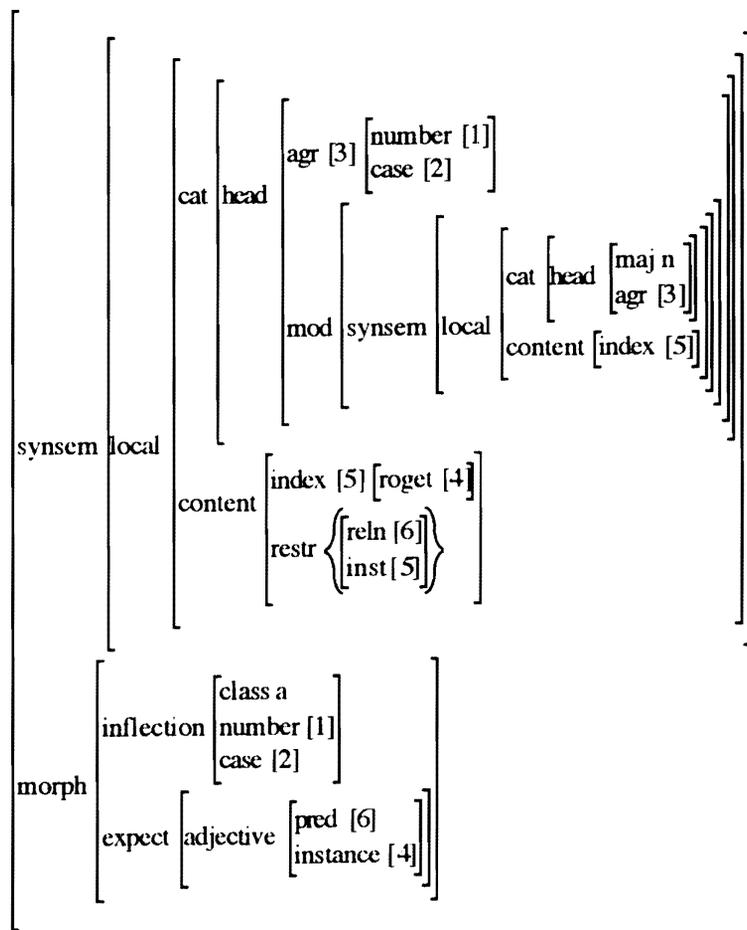


The roles of the arguments for the predicate in the restriction may be left unspecified because these are *optional complements*: they may or may not surface. One approach is to treat them as modifiers, if they appear in regular form. For example, the subject of a deverbal noun may take the form of an adjective and the object may be a non-idiosyncratic prepositional phrase. This approach can only be achieved on two conditions. First, semantically, the head in the lexicon must contain the underlying argument structure, as specified in the feature [EXPECT] in our design of lexicon. Second, syntactically the form of the arguments should not be idiosyncratic, which, fortunately, is largely true for Esperanto as a planned language. (Any idiosyncratic selection, say, requiring a PP with a specific preposition, will have to be treated in the head-complement structure defined in the subcat list of the head noun.) We usually take the modifier approach when the underlying argument is optional on surface. Nevertheless, it is not unimaginable to design an extreme linguistic model of Esperanto in which all arguments are treated as modifiers. There will be no place for head-complement structures in the syntax while the semantics can still be built in the same way. In that case, an accusative noun would select its verb and fill the object role; a nominative noun would select a finite verb as its subject; etc. The feasibility of this measure for Esperanto indicates the regularity of the planned language.

In the above discussion, we differentiate *argument* from *complement*: the former is a semantic term and the latter a syntactic term. More precisely, argument refers to the necessary role expected by the underlying predicate (in [MORPH]), while complement refers to the obligatory linguistic sign subcategorized for by the head word (in [SYNSEM]). Argument usually, but not necessarily, is realized as a complement.

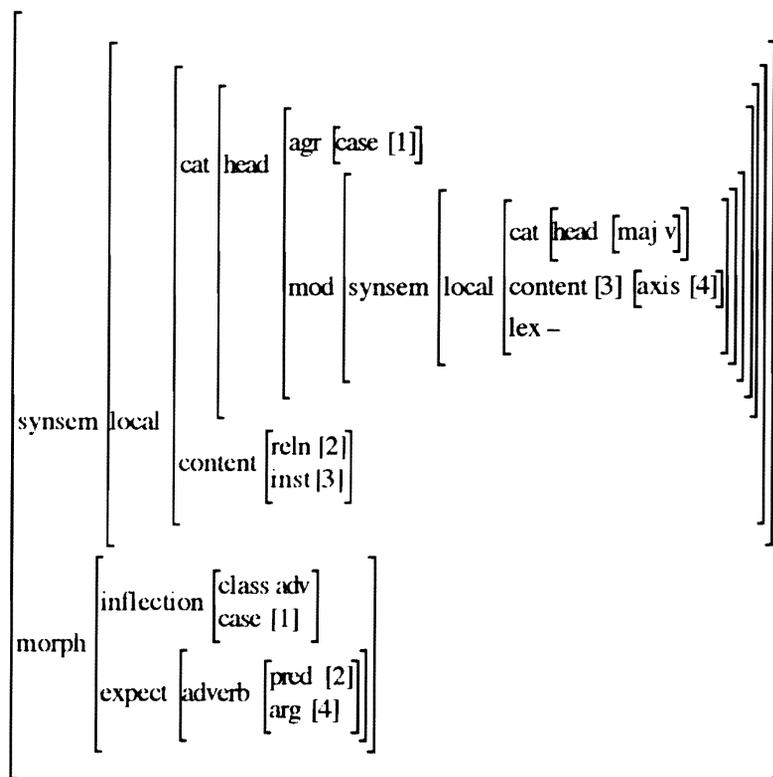
3.3 Adjective constraint

As with the noun, [NUMBER] and [CASE] are agreement features. In addition, there is a syntactic [MOD] feature which corresponds to the underlying adjectival [EXPECT].



## 3.4 Adverb constraint

Similar to adjective, adverb also has two types of semantics: head semantics for its predicative role and adjunct semantics for its modifier role. But they are token identical.



## ACKNOWLEDGMENT

The author thanks Dr. Paul McFetridge for his supervision on the Esperanto HPSG research. The experiments are conducted using his LISP implementation of HPSG parser.

## REFERENCES

- Kalocsay, K. & Waringhien, G. 1958. *Plena Gramatiko de Esperanto*, Esperanto-Propaganda Centro, Milan.
- Li, Wei. 1986. *Automatika Tradukado el la Internacia Lingvo en la Chinan kaj Anglan Lingvojn*. grkg/Humankybernetik, Band 27, Heft 4. Germany.
- Li, Wei. 1994. *Survey of Esperanto Inflection System*, NWLC, SFU, Burnaby.
- Li, Wei & McFetridge, Paul. 1995. *Handling Chinese NP predicate in HPSG*. PACLING-II Proceedings. Brisbane, Australia.
- Pollard, Carl & Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*, Center for the Study of Language and Information, Stanford University, CA.