Linguistics 567, Spring 2005, Jonathan Pool
Lab 3 (Esperanto)

Preparation

Vocabulary

Nouns:

king	reĝo
turnip	rapo
man (individual)	homo
man (male adult)	viro
staple_food	pano
water	akvo
dog	hundo
peace	paco
garden	ĝardeno
student	lernanto
tomato	tomato

Intransitive verbs:

rave	frenezi
flow	flui
die	morti
walk	marŝi
rain	pluvi

Transitive verbs:

ridicule	moki
find	trovi
enter	eniri
fear	timi
like	ami

Other verbs:

go	iri

Verb that is normally used only intransitively: "morti" (die).

Case

Esperanto has overt case.

Subjects have the nominative case. Direct objects of transitive verbs have the accusative case. Case applies to nouns, pronouns, adjectives, adverbs, and pronominal and adjectival "correlatives". Other contexts, including predicate phrases and prepositional objects, behave according to patterns not yet described here.

No verb requires an exceptional case.

Case is expressed morphologically with the ending -n. This ending is invariant.

Case interacts with semantics. The accusative case can generally alternate with prepositions. But then one might not know which preposition has been replaced by an accusative-case inflection. Typically, the accusative alternation appears when the preposition would have designated a destination, a time, a duration, or a dimension.

Case also interacts with case. It is generally considered marginally grammatical for a verb to have more than one accusative-case complement. For example, if a transitive verb is made causative, it has two implicit complements, the one that is made to act and the one that is acted on, and either if made overt can have the accusative case, but if both are overt then the speaker avoids making them both complements of the verb by reformulating the clause.

Agreement

Agreement types:

person	no
number	yes
gender	no
case	yes
definiteness	no

Agreement constituents:

subject-verb	no
object-verb	no
determiner-noun	if uninflected determiner, no; if inflected determiner, yes
adjective-noun	yes

Agreement paradigm: The adjective (or inflected determiner) and noun both take the -j ending if plural, or the -n ending if accusative, or the -jn endings if both plural and accusative.

Test Suite for Case and Adjective-Noun Agreement

```
frenezaj  reĝoj lernantojn   timas
;crazy-PL kings students-ACC fear-PRES
;'Crazy kings fear students.'

reĝoj   frenezaj lernantojn   timas
;kings crazy-PL students-ACC fear-PRES
;'Crazy kings fear students.'

reĝoj   frenezajn    lernantojn   timas
;kings crazy-PL-ACC students-ACC fear-PRES
;'Kings fear crazy students.'

reĝoj   frenezaj frenezajn    lernantojn   timas
;kings crazy-PL crazy-PL-ACC students-ACC fear-PRES
;'Crazy kings fear crazy students.'

reĝoj frenezajn    frenezaj     lernantojn   timas
;kings crazy-PL-ACC crazy-PL-ACC students-ACC fear-PRES
;'Crazy kings fear crazy students.'
```

```
;(very rare word order)

*reĝoj frenezanj    lernantojn   timas
;kings crazy-ACC-PL students-ACC fear-PRES

*reĝoj frenezajn    lernantonj    timas
;kings crazy-PL-ACC student-ACC-PL fear-PRES

reĝoj  frenezajn   timas      lernantojn
;kings  crazy-PL-ACC fear-PRES students-ACC
;'Kings fear crazy students.'
;(rare word order)

*frenezj    reĝoj lernantojn   timas
;crazy-0-PL kings students-ACC fear-PRES

*reĝj      frenezaj lernantojn   timas
;king-0-PL crazy-PL students-ACC fear-PRES
```

Test Suite for Case and Determiner-Noun Agreement

This test suite presupposes that inflected determiners are subject to the same syntactic rules as adjectives, there is at most 1 uninflected determiner per noun, and if present it must immediately precede the noun and all other premodifiers of the noun.

```
timas      tiajn       reĝojn    niajn       lernantoj ĉiuj
;fear-PRES such-PL-ACC kings-ACC our-PL-ACC students  all-PL
;'All students fear such kings of ours.'

timas      tiajn       reĝojn    niaj   lernantoj ĉiuj
;fear-PRES such-PL-ACC kings-ACC our-PL students  all-PL
;'All our students fear such kings.'

timas      la tiajn       reĝojn    niaj   lernantoj ĉiuj
;fear-PRES the such-PL-ACC kings-ACC our-PL students  all-PL
;'All our students fear the kings of that kind.'

*timas       tiajn       la reĝojn    niaj   lernantoj ĉiuj
;fear-PRES  such-PL-ACC the kings-ACC our-PL students  all-PL

lernantoj ĉies       hundojn  mokas
;students everybody's dogs-ACC ridicule-PRES
;'Students ridicule everybody's dogs.'

*lernantoj hundojn  ĉies        mokas
;students  dogs-ACC everybody's ridicule-PRES

lernantoj ĉies flavajn hundojn kadukajn mokas
;students everybody's yellow-PL-ACC dogs-ACC decrepit-PL-ACC
;ridicule-PRES
```

;'Students ridicule everybody's yellow decrepit dogs.'

*lernantoj flavajn ĉies hundojn kadukajn mokas
;students yellow-PL-ACC everybody's dogs-ACC decrepit-PL-ACC
;ridicule-PRES

Test Suite for Argument Optionality

la   hundoj frenezas
;the dogs   crazy-PRES
;'The dogs are going crazy.'

frenezas
;crazy-PRES
;'It's crazy.'

la   reĝo eniras   la ĝardenon
;the king enter-PRES the garden-ACC
;'The king is entering the garden.'

la   reĝo eniras
;the king enter-PRES
;'The king is entering.'

*la   reĝo la ĝardenon
;the   king the garden-ACC

Test Suite for Modification

eĉ guto malgranda konstante frapante traboras la monton granitan
;even drop-N un-large constantly hitting-ADV through-bore-PRES
;the mountain-ACC granite-ADJ-ACC
;'Even a droplet, dripping continually, bores through
; the granite mountain.'

*guto traboras   malgranda monton
;drop bore-through-PRES small   mountain-ACC

*traboras   monton   guto granitan
;through-bore-PRES mountain-ACC drop granite-ADJ-ACC

la frapanta guto traboras   monton
;the hitting drop through-bore-PRES mountain-ACC
;'The dripping drop bores through a mountain.'

*la frapante   guto traboras   monton
;the hitting-ADV drop through-bore-PRES mountain-ACC

Test Suite for Polar Questions

This test suite pretends that tag questions do not exist.

```
ĉu        la  paco  mortas
;whether the peace die-PRES
;Is peace dying?'

ĉu        mortas la  paco
;whether die-PRES the peace
;Is peace dying?'

ĉu        la  paco  ne  mortas
;whether the peace not die-PRES
;Isn't peace dying?'

*la  paco  ĉu        mortas
;the peace whether die-PRES

*la  ĉu        paco  mortas
;the whether peace die-PRES
```

Test Suite for Imperatives

```
mortu
;die-IMP
;'Die.'

la    hundo mortu
;the dog    die-IMP
;'The dog shall die.'

mortu la hundo
;die-IMP the dog
;'The dog shall die.'

ĉu la hundo mortu
;whether the dog die-IMP
;'Shall the dog die?'

pluvu
;rain-IMP
;'Let there be rain (It shall rain).'

ĉu pluvu
;whether rain-IMP
;'Shall it rain.'
```

Test Suite for Modals

This test suite ignores verb derivation that accomplishes the same semantic effect as do modal verbs. It also finesses the question, whether any modal class(es) of verbs should be analyzed as existing and, if so, which. Finally, it assumes

that the order of the 3 constituents (subject, modal verb, main verb) in intransitive modal sentences and the order of the 4 constituents (subject, modal verb, main verb, object) in transitive modal sentences are free (see Kalocsay & Waringhien, 1985, p. 365, for support).

```
;lernantoj = student-PL
;povas = can
;moki = ridicule-INF
;reĝojn = kings-ACC
;All sentences in this set = 'Students can ridicule kings.'

lernantoj povas moki reĝojn
lernantoj povas reĝojn moki
lernantoj moki povas reĝojn
lernantoj moki reĝojn povas
lernantoj reĝojn povas moki
lernantoj reĝojn moki povas
povas lernantoj moki reĝojn
povas lernantoj reĝojn moki
povas moki lernantoj reĝojn
povas moki reĝojn lernantoj
povas reĝojn lernantoj moki
povas reĝojn moki lernantoj
moki lernantoj povas reĝojn
moki lernantoj reĝojn povas
moki povas lernantoj reĝojn
moki povas reĝojn lernantoj
moki reĝojn lernantoj povas
moki reĝojn povas lernantoj
reĝojn lernantoj povas moki
reĝojn lernantoj moki povas
reĝojn povas lernantoj moki
reĝojn povas moki lernantoj
reĝojn moki lernantoj povas
reĝojn moki povas lernantoj

lernantoj   devas     reĝojn    moki
;student-PL must-PRES kings-ACC ridicule-INF
;'Students must ridicule kings.'

reĝojn      volas     lernantoj  moki
;kings-ACC want-PRES student-PL ridicule-INF
;'Students want to ridicule kings.'

esperas     studentoj reĝojn    moki
;hope-PRES students   kings-ACC ridicule-INF
;'Students hope to ridicule kings.'

*lernantoj   moki          reĝojn
;student-PL  ridicule-INF kings-ACC
```

```
*lernantoj    povas moki         reĝoj
;student-PL   can    ridicule-INF kings
```

Test Suite for Sentential Negation

This test suite ignores nonsentential negation and doesn't include any grammatical examples of nonsentential negation as either positive or negative examples.)

```
paco    ne  mortas
;peace not die-PRES
;'Peace isn't dying.'
```

```
ne     mortas   paco
;not die-PRES peace
;'Peace isn't dying.'
```

```
*paco  mortas   ne
;peace die-PRES not
```

```
ne pluvas
;not rain-PRES
;'It isn't raining.'
```

```
*pluvas    ne
;rain-PRES not
```

Test Suite for Coordination

This test suite ignores the "kaj-kaj" ('both-and') construction.

```
lernantoj amas       panon      kaj akvon
;students like-PRES bread-ACC and water-ACC
;'Students like bread and water.'
```

```
lernantoj amas panon kaj tomatojn ĝardenajn
;students like-PRES bread-ACC and tomatoes-ACC garden-ADJ-PL-ACC
;'Students like bread and garden tomatoes.'
;'Students like garden bread and garden tomatoes.'
;(Ambiguous)
```

```
lernanto trovis panon     kaj tomaton     ĝardenan
;student found  bread-ACC and tomato-ACC garden-ADJ-ACC
;'A student found bread and a garden tomato.
;(Thus must be parsed so "garden" does not modify "bread")
```

```
lernanto trovis panon     kaj tomaton     ĝardenajn
;student found  bread-ACC and tomato-ACC garden-ADJ-PL-ACC
;'A student found garden bread and a garden tomato.
;(Thus must be parsed so "garden" modifies both "bread" and
```

```
;"tomato")

*lernantoj amas      pano  kaj akvon
;students  like-PRES bread and water-ACC

*lernantoj amas      panon      kaj
;students  like-PRES bread-ACC and

*lernantoj amas      kaj panon
;students  like-PRES and bread-ACC

la   reĝo amas      la marŝojn        homan     kaj hundan
;the king like-PRES the walks-N-PL-ACC human-ACC and canine-ACC
;'The king likes the human and canine walks. (One of each kind)'

la   reĝo amas      la marŝojn        homajn      kaj hundan
;the king like-PRES the walks-N-PL-ACC human-PL-ACC and canine-ACC
;'The king likes the human and canine walks.'
;(Multiple human, single canine)

la   reĝo amas      la marŝojn        homan     kaj hundajn
;the king like-PRES the walks-N-PL-ACC human-ACC and canine-PL-ACC
;'The king likes the human and canine walks.'
;(Single human, multiple canine)

*la  reĝo amas      la marŝojn        homajn      kaj hunda
;the king like-PRES the walks-N-PL-ACC human-PL-ACC and canine

reĝojn        timas    kaj mokas        lernantoj
;kings-PL-ACC fear-PRES and ridicule-PRES students
;'Students fear and ridicule kings.'

*reĝojn        timas    lernantoj kaj mokas
;kings-PL-ACC  fear-PRES students  and ridicule-PRES
```

Package Selection

Esperanto has case and agreement. The package chosen is case, adjective-noun agreement, and lexical rules.

The test suite for this task consists of the "Test Suite for Case and Adjective-Noun Agreement" and the "Test Suite for Case and Determiner-Noun Agreement", set forth above in the "Preparation" section.

Pronouns

I decided to consider the personal non-reflexive and reflexive pronouns and the impersonal pronoun to be "pronouns" and to assume that none of them may take a determiner.

Determiner Specification

Common nouns can appear with or without a determiner, and pronouns can appear only without a determiner. In this respect, the revised grammar made correct predictions. However, determiners continued to be permitted to satisfy complement requirements, and this is incorrect.

Case

The enhancement of noun lexical entries with case constraints correctly eliminated all parses of intransitive sentences with accusative subjects and transitive sentences with accusative subjects or nominative direct objects. It also halved the number of parses of grammatical transitive sentences, by eliminating parses that had analyzed the subject as direct object and the direct object as subject. Incorrect parses remained whenever a determiner appeared in a transitive sentence, because, in addition to the correct parse, the determiner was also analyzed as a complement. In addition, I have discovered that nouns and pronouns and their phrases are incorrectly analyzed as complements, and there is no limit to the number of complements accepted in a verb phrase.

Adjectives

I defined the "adjective-lex" type with an unspecified value for POSTHEAD, believing that this would reflect the free adjective-noun word order in Esperanto. I was unsure of this, because I didn't clearly understand the statement about English, "We have both orders (head-adj and adj-head), but adjectives are always prehead". I have not noticed any parsing behavior that seems to contraindicate this decision.

Agreement

The enhancement of the lexical entries for inflected determiners made sentences with determiner-noun agreement parse and those with disagreement not parse. Transitive sentences with determiners continued also to be misparsed as if the determiners were verb or verb-phrase complements.

Adjective-noun agreement and determiner-adjective-noun agreement appeared to be enforced correctly by the grammar.

Lexical Rules

I defined lexical rules to begin dealing with number and case inflection and agreement in Esperanto.

I began with the observation that every uninflected noun, pronoun, determiner, adjective, or adverb stem can be analyzed as beginning its life at a particular level of a shared inflectional scale. The levels can be described as:

0. Ready for category inflection.
1. Ready for number inflection.
2. Ready for case inflection.
3. Ready for use.

Some lexemes, such as the uninflectable definite article "la" ('the'), begin life ready for use. Others, such as the personal pronouns like "mi" ('I'), begin life ready for case inflection.

Others, such as the inflectable determiner "tiu" ('that'), begin life ready for number inflection. And others, such as the noun stem "hund" ('dog'), begin life ready for category inflection.

Wherever a stem begins on the scale, it proceeds only forward (toward readiness for use). Although the phonological/orthographic form taken by each inflection is identical for all lexeme and word types (part of speech), we still need separate lexical rules for particular types, for three reasons. One is that the types differ in the features that need to have values added as inflections take place. A second is that the semantics of a word of a final category can depend on its original category. The third is that progress along the scale is not always continuous: Any lexeme that is inflected for the adverb category then skips over number inflection and is next ready for case inflection.

In order to assure that all required lexical rules are applied (even if with null affixation) before a lexeme is ready for use and no lexical rule is reapplied (see the warning about cyclicality of constant lexeme-to-lexeme rules in the Matrix file), I defined a NEEDINF feature and made it appropriate for "word-or-lexrule" so it could be a feature of both the input and the output. Its value indicates where on the inflectional scale the input and output are located (e.g., the input is ready for number inflection, and the output is ready for case inflection). I also defined the original lexemes as INFLECTED -, so they wouldn't be usable as words until this value is changed by a case-inflection rule, which is a lexeme-to-word rule.

After debugging, I found the parsing fully in accord with expectations. The trees output by the LKB parser show each noun or inflected determiner in the appropriate place on a multinode nonbranching tree, whose nodes show the respective rules applied below and above the word.

The defects identified earlier with respect to determiners analyzed as complements of verbs and node labels not culminating with "S", in the case of transitive sentences, remain to be addressed.

I have implemented this approach so far on nouns, determiners, and adjectives. I haven't yet implemented it on verbs or adverbs.

Head-Modifier Rules

After I defined head-modifier rules, both transitive and intransitive sentences elicited correct parses and many incorrect ones. For example, the sentence "la junaj hundoj trovis tiun nigran panon" ('the young dogs found that black bread') has one correct parse, but generated an edge-limit error. The sentence "junaj hundoj trovis tiun nigran panon" ('young dogs found that black bread') also has one correct parse, but also generated an edge-limit error. The sentence "hundoj trovis tiun nigran panon" ('dogs found that black bread') also has one correct parse, but generated 16 parses. Adjectives were being accepted as complements of verbs. Also, the adjective-head rule was parsing adjective-noun constructions twice, once with the nouns analyzed as words and once with them analyzed as phrases.

Constraining non-adjective head types to have empty MOD lists decreased the number of incorrect parses, making them few enough to eliminate the edge-limit errors with grammatical sentences. In the case of "hundoj trovis tiun nigran panon", the number decreased from 16 to 4. The only constraint that affected this difference was that on the determiner. When the determiner MOD constraint was absent, the determiner was being misanalyzed not only as a complement of the verb (as before), but also as the non-head daughter in a head-adj-int-phrase with the verb phrase "hundoj trovis" as the head daughter.

Batch Testing

Testing on this Lab's test suite yielded expected results. With particular exceptions, all grammatical sentences were correctly parsed, and additional incorrect parses occurred whenever it was possible to misanalyze a determiner or adjective as a verb or verb phrase's complement.

The other prior error that recurred was the mislabeling of VP and S nodes in transitive sentences.

Having further studied word order in Esperanto, I have adopted the prevailing position of the leading descriptive grammarians (who double as advisory stylists) that, in addition to the free order of the verb and its arguments, each of the primary arguments (subject and direct object) may also be freely split within the sentence, so that adjectives and the nouns they modify may be located in any order and at any distance, at least if there are no other complements of the verb except a direct object. The grammar fails to accept test sentences with separated constituents, such as "reĝoj frenezajn timas lernantojn" ('Kings fear crazy students'), with the order 'kings-crazy-fear-students'.

Another new error is in the labeling of adjectives in the parse trees. They are labeled "ADV".

The Lab 2 test suite continues to be parsed with the same success and the same previously reported errors as before.

The Lab 1 test suite was in English, so doesn't seem applicable to this grammar.

A technical error is that LKB apparently misreads the beginning of a test-suite file when it is encoded in UTF-8.