

Linguistics 567, Spring 2005, Jonathan Pool
Lab 2 (Esperanto)

Preparation

Transitivity:

Verb that is normally used only strictly transitively: "trovi" (find).

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

Common nouns: "paco" (peace), "hundo" (dog).

What word orders are permissible in transitive and intransitive sentences?

In transitive sentences, the accepted phrase orders are: VSO, VOS, SVO, OVS, SOV, OSV.

In intransitive sentences, the accepted phrase orders are: VS, SV.

Basics of case:

Two cases: nominative and accusative. Accusative applies to direct objects and in accusative alternation, i.e. as alternative to preposition. Otherwise, nominative applies. Case variation applies only to nouns, pronouns, adjectives, adverbs, non-possessive pronominal, adjectival, and adverbial "correlatives", and participles. For this purpose, none of these types includes the article or numbers.

Basics of agreement:

Agreement exists between nouns, pronouns, adjectives, non-possessive pronominal and adjectival correlatives, and participles. The sole agreement feature is number, and its possible values are singular and plural.

What particles or determiners or other elements are necessary to complete a noun phrase?

When a noun phrase is headed by a noun, none. There are optional elements, but no required ones.

Instances of NP consisting of just a noun (no determiner or other element):

"paco" (peace)

"akvo" (water)

"hundo" (a dog)

"hundoj" (dogs)

Instances of NP consisting of uninflected determiner + noun:

"la pacoj" ([the] peace)

"la akvoj" (the water)

"la hundoj" (the dog)

"la hundoj" (the dogs)

"ies hundo" (somebody's dog)

"ambaŭ hundoj" (both dogs)

Instances of NP consisting of inflected determiner + noun:

"mia hundo" (my dog)
"hundo mia" (my dog, a dog of mine)
"miaj hundoj" (my dogs)
"hundoj miaj" (my dogs, dogs of mine)
"tiu hundo" (that dog)
"tiuj hundoj" (those dogs)
"hundo tiu" (that dog)
"hundoj tiuj" (those dogs)
"tia hundo" (such a dog)
"nenia hundo" (no kind of dog)
"kia hundo" (what kind of dog)
"ia hundo" (some kind of dog)
"ĉia hundo" (every kind of dog)

Overt determiners:

Demonstrative: "tiu" (that)
Definite: "la" (the)
Indefinite: none unmarked; "iu" (some, one, a certain)

Agreement paradigm for determiners:

Article ("la"): no inflection.

Possessive correlatives ("ies" [somebody's], "ties" [that one's], "nenies" [nobody's], "kies" [whose], "ĉies" [everybody's]): no inflection.

All other determiners: to singular nominative form ("mia", "tiu", "kia", etc.), affix "j" (plural), "n" (accusative), "jn" (plural accusative).

Sources:

Ulrich Koch, "The Morphology of Esperanto", 1992 (<http://www.uni-koblenz.de/~koch/espmo.ps.gz>).

K. Kalocsay and G. Waringhien, *Plena Analiza Gramatiko de Esperanto*, 5th edn. (Rotterdam: Universala Esperanto-Asocio, 1985).

Christopher Gledhill, *The Grammar of Esperanto: A Corpus-Based Description*, 2nd edn. (Munich: Lincom Europa, 2000).

John Wells, *Lingvistikaj Aspektoj de Esperanto* (Rotterdam: Universala Esperanto-Asocio, 1978).

David K. Jordan, *Being Colloquial in Esperanto: A Reference Guide*, rev. edn. (El Cerrito, Calif.: Esperanto League for North America, 1999).

Part 1

The basic word order of Esperanto is free (VSO, VOS, SVO, OVS, SOV, OSV). Examples:

hundo trovis pacon.
dog found peace-ACC
'A dog found peace.'

hundo pacon trovis.
dog peace-ACC found
'A dog found peace.'

trovis hundo pacon.
found dog peace-ACC
'A dog found peace.'

trovis pacon hundo.
found peace-ACC dog
'A dog found peace.'

pacon hundo trovis.
peace-ACC dog found
'A dog found peace.'

pacon trovis hundo.
peace-ACC found dog
'A dog found peace.'

The order of nouns and determiners follows two rules:

1. Uninflected determiners precede the noun. Examples:

la hundo
the dog
'the dog'

ies hundo
somebody's dog
'somebody's dog'

ambaŭ hundoj
both dogs
'both dogs'

2. Inflected determiners are free to either precede or follow the noun.

mia hundo
my dog
'my dog'

hundo mia
dog my
'my dog'

tia hundo
such dog
'such a dog'

hundo tia
dog such
'such a dog'

neniuj hundoj
no-PL dogs
'no dogs'

hundoj neniuj
dogs no-PL
'no dogs'

Simplifications of the data: Descriptive grammars of Esperanto generally avoid categorizing constructions as outright ungrammatical, instead describing them as rare, archaic, novel, poetic, inelegant, etc. I am adopting an idealized dichotomy, in which I imprecisely weight commonness, elite approbation, and regularity in classifying constructions as grammatical.

The instructions for this part do not indicate that any description of the analysis of the lexical types and phrase structures is to be included in the report. Thus, I am not including such a description. However, I note that the results have been puzzling in two respects:

Sentences with intransitive verbs have been parsed as expected, but sentences with transitive verbs have not. In these sentences, the verb phrases with complements are labeled "V" rather than "VP" and combine with the subject NPs into a root node labeled "V" rather than "S". Changing the order of the "V", "VP", and "S" labels in the labels.tdl file has some impact on this, but I have not investigated whether such a change would constitute a solution with no adverse side-effects. Further, in all orders, transitive-verb sentences yield two completed parses, because the grammar doesn't yet recognize case inflection, so it has no way to know which of the two NPs is the subject and which is the object.

Nouns can be specified by an unlimited number of consecutive determiners. For example, "la la la la hundo" ('the the the the dog') is accepted as a verb argument. Before dealing with this, I believe it is best to decide whether, and if so when, multiple determiners specifying a single noun are grammatical. The decision may depend on which words are treated as determiners (see below).

Part 2

I have classified determiners as inflected and uninflected. I have codified the prevailing usage by classifying uninflected determiners as mandatorily pre-nominal and inflected ones as free to be pre- or post-nominal. This practice also prevents ambiguity that would arise if an uninflected determiner could follow a noun, because when two noun phrases are adjacent then a determiner between them could have been parsed as attached to either one.

I have postponed the analysis of how determiners and adjectives are ordered with respect to each other, until we are dealing with adjectives.

I have tentatively classified possessive pronouns as determiners, but it may be simpler and equally predictive to classify them as ordinary adjectives, so I intend to review this issue when we are dealing with adjectives. More aggressively, it may even be reasonable to classify all inflected determiners as adjectives.

I have noticed that the existing inventory of PRED values for determiners appears to imply a much more restrictive definition of "determiner" than is sometimes (e.g., in SWB) adopted, since it is unclear to me which of these PRED value would be appropriate for words in English such as "seven", "some", "both", and "his".

Pending clarification of the extension of the "determiner" category, I have included into my grammar as the only determiner the definite article "la", to which I gave the PRED value "def_q_rel". One of the uses of "la" is to indicate definiteness (old information, focal item). However, in many contexts (e.g., abstract nouns, substance nouns, time-period nouns, proper names) it is sometimes present and sometimes absent, and the semantic effects of its inclusion are sometimes dubious, subtle, and varying. An example of inter-language variation that would affect translation is that the English possessive pronoun would usually be rendered by the Esperanto definite article attached to a noun describing a part of the human body in a clause where the person whose body it is is another argument. E.g., "I cut my hair" would usually be translated as "Mi tondis la harojn" ('I sheared the hairs') rather than "Mi tondis miajn harojn" ('I sheared my hairs').

Given that the only lexical entry for a determiner so far is for an uninflected one, I have so far included in the grammar only a single head-specifier phrase type, which inherits from head-final. For inflected determiners, if they are to be treated as specifiers rather than adjectives, it will be necessary to provide a head-initial head-specifier phrase type, and presumably to create two determiner subtypes so their types can be specified in their lexical entries and each specifier can be eligible to participate in the appropriate head-specifier phrase type.

Part 3

The subtype of basic-bare-np-phrase that I initially created for determinerless nouns produced unexpected results. It caused sentences with determiners, previously parsed correctly (except for the labels and the case ambiguity mentioned above), to have many more parses, including parses with unlabeled nouns and parses with verb-attached determiners. The two initial parses of a transitive-verb sentence with both nouns having determiners became eighteen parses.

Inspection revealed that the basic-bare-np-phrase type was defined so as not to require that its head be "noun". I added a constraint to the subtype, but this failed to have any effect. I investigated further and found that the phrase types with complements permit determiners to be complements, and they were being parsed as such.

Before investigating the imposition of additional constraints to eliminate improper parses, I thought I should understand better the definition of "determiner" and consider further the possible determiner subtypes. In fact, there are rare attested uses of "la" ('the') as a stand-alone verb complement in Esperanto, and frequent uses of other possible determiners (such as possessive pronouns and demonstratives) that way. I'm not adopting the view that rare uses must be recognized as grammatical, but I need a strategy for dealing with this issue.

None of the descriptive grammars of Esperanto appears to assert that any noun always requires a determiner. Proper nouns are normally not accompanied by determiners, even when they are in other languages, e.g., "Nilo" ('the Nile'), but usually take determiners in appropriate contexts, such as "la praa Barato" ('the India of antiquity').

Part 4

Parsing "la hundo mortis" ('the dog died') and then generating sentences from the semantic representation yields four sentences:

la hundo mortis
la hundon mortis

mortis la hundo
mortis la hundon

Those containing "hundon" (dog-ACC) are ungrammatical. The grammatical sentences have the same truth-value meaning.

Parsing "hundo trovis pacon" ('a dog found peace') and then generating sentences from the semantic representation yields twenty-four sentences:

hundo paco trovis
hundo pacon trovis
hundo trovis paco
hundo trovis pacon
hundon paco trovis
hundon pacon trovis
hundon trovis paco
hundon trovis pacon
paco hundo trovis
paco hundon trovis
paco trovis hundo
paco trovis hundon
pacon hundo trovis
pacon hundon trovis
pacon trovis hundo
pacon trovis hundon
trovis hundo paco
trovis hundo pacon
trovis hundon paco
trovis hundon pacon
trovis paco hundo
trovis paco hundon
trovis pacon hundo
trovis pacon hundon

The six sentences containing nominative "hundo" ('a dog') and accusative "pacon" ('peace') are correct and identical in truth-value meaning. The six that contain the opposite cases are grammatical, but have the wrong meaning ('peace found a dog'). And the twelve that contain either two nominative nouns or two accusative nouns are ungrammatical.

The grammar needs to distinguish nominative and accusative case on nouns in order to prevent errors of this kind.

Concluding comment

I forgot to keep track of my time in order to report the amount of time required for this assignment. However, on the basis of memory I can report that Part 1 took approximately forever, and the other parts took comfortable amounts of time.

The duration of Part 1 was due to my confusion about (1) what was required and (2) how to do it. I spent perhaps 6 hours trying to understand and satisfy the one sentence, "these new types should specify identities between the elements of ARG-ST and the valence lists, and place appropriate constraints on the HEAD and VAL values of those elements." It turned out

that it was important to understand whether the value of a feature was a sign subtype or a synsem-min subtype. I began wrongly assuming the former when the latter was the case, and assuming I could constrain HEAD values by requiring a lexical type as an element of an ARG-ST list (with the idea that I would replace lexical with phrasal types when I began dealing with phrasal types). It took me hours to discover and become familiar with the tools for exploration of feature structures and diagnosis of unification errors and to acquire some intuitions with which to make good initial guesses about why failures were occurring. I was also slowed by the cultural shock of switching from being a scientist to being an engineer, i.e. somebody who is using a model (the Matrix) but isn't expected to understand why and how it works. If I had examples from a few languages to learn patterns from, the work would probably have proceeded much faster, but I probably would have used imitation instead of cogitation and developed less robust intuitions. However, an orientation on the philosophy of the Matrix (not a page-by-page but a top-down conceptual presentation) might have helped with both speed and understanding.