

Introduction à la Linguistique Théorique

Feuille d'Exercices - II

[P. Schlenker - schlenke [arobase] humnet.ucla.edu]

Exercise 1 Syntax [Level I - straightforward application of the theories discussed in class; only Question 7 requires some thought]

In this exercise you are asked to consider 'toy' grammars for three languages, English, German and Italian.

Note: These grammars differ only with respect to (a) the setting of the head parameter for various constructions, and (b) the vocabulary (however the grammar of German contains fewer rules than the others, because there are additional complexities in German which we disregard here).

Part I. English

IP → NP I'

I' → I VP

NP → PN, D N

VP → V_i, V_t NP, V_sCP

CP → C IP

I → will, might, can, should, does, did

PN → Snoopy, Lucy, Robocop, Grover, I, you, he, she,

N → President, Troll, Dean, friend, mother, prize

D → the, some, a, every, my, his, her, my, your

V_i → sleep, run, jump, win, lose, fall

V_t → meet, date, hit, criticize, appoint

V_s → think, say, believe, claim, fear

C → that

Question 1. For each of the following sentences, circle 'yes' if it is generated by this grammar, and circle 'no' otherwise.

- | | | | |
|---|-----|---|----|
| a. the Troll will say that your mother might win | YES | / | NO |
| b. the Dean might think the President will appoint him | YES | / | NO |
| c. his mother will fear that the President might appoint Snoopy | YES | / | NO |
| d. the Troll will say that he should win the prize | YES | / | NO |

Question 2. Consider the following sentences, which are both generated by the above grammar:

- (i) the Troll will say that he might criticize Robocop.
- (ii) the Troll will say that his mother might criticize Robocop

Both sentences could be uttered by a native English speaker, but they do not give rise to the same interpretations.

- (i) **can** mean: the Troll will say that the Troll might criticize Robocop
- (i) **cannot** mean: the Troll will say that Robocop might criticize Robocop
- (ii) **can** mean: the Troll will say that the Troll's mother might criticize Robocop
- (ii) **can also** mean: the Troll will say that Robocop's mother might criticize Robocop

Given the theories discussed in class, how should these observations be explained?

Answer:

Part II. German

IP → **NP I'**

I' → **VP I**

NP → **PN, D N**

VP → **V_i, NP V_t**

I → wird (=will), kann (=can), muss (=must)

PN → Snoopy, Lucy, Robocop, Grover

N → Präsident (=President), Troll, Student (=student), Professor (=professor)

D → der (=the), ein (=a), sein (his)

V_i → schlafen (=sleep), tanzen (=dance), laufen (=run), gewinnen (=win)

V_t → treffen (=meet), kritisieren (=criticize), hassen (=hate)

Note: for simplicity CPs have been omitted here

Question 3. According to this grammar, how would you say in German:

- a. Snoopy can sleep
- b. The President will meet Grover
- c. His student will criticize Snoopy
- d. Snoopy will criticize Robocop

Note: The present grammar abstracts away from certain complexities of German. The German sentences that you will come up with are grammatical in certain syntactic environments only (e.g. after *I claim that* ____). You can disregard this fact in the present exercise.

Question 4

This grammar also generates the sentence *Er Snoopy kritisieren wird*, which means: He will criticize Snoopy. When embedded in the right syntactic environment (e.g. after *I claim that* ____, in German: *Ich behaupte, dass er Snoopy kritisieren wird*), this yields a fine sentence of German, for instance if *er* (=he) refers to Sam. *However the sentence cannot mean that Snoopy will criticize Snoopy*. By contrast, in the German version of c. above, *His student will criticize Snoopy*, *his* can refer to Snoopy. How can you explain this contrast? (Be very specific)

Answer:

Part III. Italian

IP → **NP I'**

I' → **I VP**

NP → **PN, D N**

VP → **V_i, V_t NP, V_sCP**

CP → **C IP**

I → può (=can), deve (=must)

PN → Ø, Snoopy, Lucy, Robocop, Grover

N → Presidente (=President), Troll, professore (=professor), bambino (child)

D → il (=the), un (=a),

V_i → dormire (=sleep), ballare (=dance), correre (=run), vincere (=win), odiare (=hate)

V_t → incontrare (=meet) criticare (=criticize)

V_s → pensare (=think), dire (=say),

C → che (=that)

Important Note: Ø stands for a null pronoun, which is not pronounced. This null pronoun can correspond to English *I, you, he, she, or it*.

Question 5. According to this grammar, how would you say in Italian:

- Snoopy must sleep
- Robocop must say that the Troll can win
- The Troll can say that he must win

d. Lucy must think that she can win

Question 6 This grammar does not generate the following sentence: *può dormire Snoopy*. This is an incorrect result, as this is in fact an acceptable sentence in Italian. Can you explain why it is acceptable?

Answer:

Question 7 This grammar also generates: \emptyset *deve pensare che Lucy può vincere* (literally: \emptyset must think that Lucy can win).

While the sentence 'Deve pensare che Lucy può vincere' is acceptable to Italian speakers to mean for instance (given the right context) that 'Ann must think that Lucy must win', it cannot mean that Lucy must think that she (=Lucy) must win. Suggest an explanation for this fact.

Answer:

Exercise 2 English and German (Level II)

Part A. English

Suppose we make the following assumption about English questions:

Assumption: An English question is not an IP, but an S defined by the following rule, where XP can be any moved interrogative phrase (the notation (XP) CP indicates that XP is optional):

$S \rightarrow (XP) CP$

Question 1. Explain how this theory combined with the postulation of one or two movement operations can generate the a. sentences below from trees that correspond to the b. sentences.

Hint: Assume that in *yes/no* questions, there is no XP constituent, but in *which* questions, there is one.

- (1) a. Will you go to the party?
b. you will go to the party
- (2) a. Which students will you invite?
b. you will invite which students

Answer:

Part B. German

Suppose now that German is head-final with respect to VPs and IPs, and that no other parameter distinguishes German from English.

Question 2. On this assumption, which of the following is predicted to be the German equivalent of: *Peter will criticize Annette*? (Circle the most accurate answer)

Vocabulary: *will* = *wird*
criticize = *kritisieren*

- (3)
- a. Peter wird kritisieren Annette
 - b. Peter wird Annette kritisieren
 - c. Peter kritisieren wird Annette
 - d. Peter Annette kritisieren wird

Question 3. As it turns out, the facts of German are somewhat different. The predictions are borne out in embedded clauses, but not in matrix clauses (for the purposes of this exercise you'll assume that the sentences prefixed with ?, which are in fact rather acceptable in the spoken language, are ungrammatical):

- (4) Embedded clauses
- a. *Ich glaube, daß Peter wird kritisieren Annette
 - b. ?Ich glaube, daß Peter wird Annette kritisieren
 - c. *Ich glaube, daß Peter kritisieren wird Annette
 - d. Ich glaube, daß Peter Annette kritisieren wird
- (5) Matrix clauses
- a. ?Peter wird kritisieren Annette
 - b. Peter wird Annette kritisieren
 - c. *Peter kritisieren wird Annette
 - d. *Peter Annette kritisieren wird
 - e. Kritisieren wird Peter Annette
 - f. Annette wird Peter kritisieren

Some linguists have suggested that the facts in (5) can be derived from the following assumption:

Assumption: A matrix clause in German is not an IP, but an S defined by the following rule, where XP can be any moved phrase (note that XP is obligatory).

$S \rightarrow XP CP$

Develop this line of analysis, and explain in detail how the facts in (5) can be derived (you should emphasize the similarities to and differences from the English rule for forming *which*-questions).

Answer:

Exercise 3 Adding Modifiers (Level II)

Consider again the grammar we discussed in class, both for English and for French:

IP → **NP I'**

I' → **I VP**

NP → **PN, D N**

VP → **V_i, V_t NP, V_s CP**

CP → **C IP**

Question 1.

In class we discussed but did not provide a *formal* treatment of structural ambiguities, as seen in the sentences in (1):

- (1) a. La fille a regardé le type sans lunettes
- b. The girl will hit the student with the book

The expressions *sans lunettes* and *with the book* are called 'Prepositional Phrases', or 'PPs' for short. Adjust the grammar given above so that you can generate each reading of the sentences in (1a-b). (Make sure that your grammar works, and has the required lexical rules. It is very likely that your grammar will overgenerate somewhat, i.e. generate sentences that are in fact ungrammatical. You should point this out, and explain what the source of the difficulty is).

dN d ← dP
 dI C ← CP
 V_i V_t V_s ← V_i V_t V_s NP, V_s CP
 VP ← V_i V_t V_s PP
 NP ← PN, N PP
 NP ← D N
 I ← I VP
 IP ← NP I
Hint: The following rules may be useful:

Answer:

Question 2. Can your analysis explain why (2) tends to be unambiguous, while (3) is ambiguous again?

(2) The student with the book, the girl will hit

(3) Hit the student with the book, the girl will.

Answer:

Question 3. Based on your grammar, provide a formal analysis of the following contrasts:

- (4) a. Je vais analyser ton problème avec Jean
1. Ok: You have a problem. I will analyze it with Jean.
 2. Ok: You have a problem with Jean. I will analyze it.
- b. Je vais analyser son_i problème avec Jean_i
1. Ok: Jean has a problem. I will analyze it with him.
 2. *: Jean has a problem with himself. I will analyze it.

Answer:

Exercise 4 (Level I)

Download 'Computational Constraints on Syntactic Processing in a Nonhuman Primate', by W. Tecumseh Fitch and Marc D. Hauser, which appeared in the journal *Science*, Volume 303, 16 January 2004, pp. 377-380. You may download the paper at: <http://www.linguistics.ucla.edu/people/schlenker/Hauser.pdf>

Question 1. The authors write: 'The FSG was $(AB)^n$, in which a random 'A' syllable was always followed by a single random 'B' syllable, and such pairs were repeated n times' (for any integer $n \geq 1$). Suppose the syllables of category A are:

A syllables: ba di yo

Suppose further that the syllables of category B are:

B syllables: mo nu ka

(a) Which of the following strings of syllables does **not** belong to the language $(AB)^n$? (circle one answer)

a. ba mo

b. di nu yo ka

c. yo nu di ka ba mo

d. di mo ba di yo ka

e. ba mo ba mo yo ka di nu

f. di nu yo ka di nu yo ka

g. di nu yo ka di nu yo mo

h. ba mo di nu yo ka ba nu

(b) Write a Finite State Grammar that generates exactly the language $(AB)^n$

Question 2. The authors then write: 'The corresponding PSG, termed A^nB^n , generated strings with matched numbers of A and B syllables. In this grammar, n sequential 'A' syllables must be followed by precisely n 'B' syllables' (for $n \geq 1$). As in the preceding question, suppose the syllables of category A are:

A syllables: ba di yo

Suppose further that the syllables of category B are:

B syllables: mo nu ka

(a) Which of the following Phrase Structure Grammars generates exactly this language? (Assume that the root node is S. Circle one answer)

(i) **Grammar 1.**

$S \rightarrow AB$

$A \rightarrow ba, di, yo$

$B \rightarrow mo, nu, ka$

(ii) Grammar 2.

$S \rightarrow ABS, AB$

$A \rightarrow ba, di, yo$

$B \rightarrow mo, nu, ka$

(iii) Grammar 3.

$S \rightarrow ASB, AB$

$A \rightarrow ba, di, yo$

$B \rightarrow mo, nu, ka$

(iv) Grammar 4.

$S \rightarrow ASB$

$A \rightarrow ba, di, yo$

$B \rightarrow mo, nu, ka$

(v) Grammar 5.

$S \rightarrow ASSB$

$A \rightarrow ba, di, yo$

$B \rightarrow mo, nu, ka$

(vi) Grammar 6.

$S \rightarrow ASSB, AB$

$A \rightarrow ba, di, yo$

$B \rightarrow mo, nu, ka$

(b) Give a Phrase Structure Tree for the string: ba di yo mo nu ka, which is part of the language (in drawing the tree, you will use the rules that you will have selected in Part (a))

Question 3

Does this result suggest that tamarins should or should not in principle be able to learn human languages? Motivate your answer.

Answer:

Exercise 5 Center embedding [Level III]

Question 1. Sentences in (1b-c) [which are instances of 'center embedding'] are typically considered somewhat deviant (the more question marks, the more deviant the sentence is...). By contrast, the sentences in (2b-c) seem to be more readily acceptable.

- (1) a. A bird died
 b. ?A bird (that) a cat ate died
 c. ???A bird (that) a cat (that) a dog bit ate died
- (2) a. A cat ran away
 b. A cat that ate a bird died
 c. A dog that bit a cat that ate a bird died.

a. Do you find similar contrasts in French between (3) and (4)?

- (3) a. Un oiseau est mort.
 b. Un oiseau qu'un chat a mangé est mort
 c. Un oiseau qu'un chat qu'un chien a mordu a mangé est mort
- (4) a. Un chat a disparu
 b. Un chat qui a mangé un oiseau a disparu
 c. Un chien qui a mordu un chat qui a mangé un oiseau a disparu

b. For the English examples that are deviant (and possibly for their French counterparts), argue for or against the following hypotheses:

Hypothesis 1: The most reasonable idealization is that these are grammatical sentences, which should be generated by the linguist's grammar.

Hypothesis 2: The most reasonable idealization is that these are ungrammatical sentences, which should not be generated by the linguist's grammar.

Answer:

Question 2. Suppose that the hearer obeys the following strategy to process a sentence:

Strategy: As soon as a linguistic head can be predicted, create a node for it and store it in your memory as long as the corresponding word has not been processed.

Example: Every sentence contains at least a verb. Thus as soon as the first word of a sentence is processed, say *John*, a placeholder for a verb is stored in the speaker's memory. It is held in memory as long as the verb has not been processed.

Let us apply this Strategy to a simple example:

(4) The student left

After 'the' 2 elements are stored, since 2 elements are known to be obligatory: a noun and a verb.

After 'student' is processed, only 1 element is stored, since a noun has now been encountered; only a placeholder for a verb is now stored in the hearer's memory.

We will assume that the memory cost at each word is the number of placeholders that must be stored in the hearer's memory.

a. Does this proto-theory yield an account of the contrast between (1) and (2)? Apply your analysis to the following contrast by computing the *maximal* memory cost reached at any point in each sentence :

- (6) a. ?A bird that a cat ate died
- b. A cat that ate a bird died

Answer:

b. On the assumption that the underlying assumptions about memory cost are well-motivated, how does this analysis bear on Question 1?

Answer:

Note: A much more sophisticated version of this analysis is offered in Gibson, E. (1998) 'Linguistic complexity: locality of syntactic dependencies', *Cognition* 68 (1998) 1–76 [ask the instructor for a copy].

Question 3. Consider the following sentences:

- (5) 1. A wolf ate
- 2. A wolf a wolf ate ate
- 3. A wolf a wolf a wolf ate ate ate
- ...
- n. (a wolf)ⁿ⁻¹ (ate)ⁿ⁻¹
- ...

The sentences listed in (5) are the set of strings of the form $(a \text{ wolf})^n (ate)^n$ for $n \geq 1$ (a finite number of occurrences of *a wolf* followed by the same number of occurrences of *ate*).

In this question, we wish to show that, quite generally, no language of the form $\{a^n b^n : n \geq 1\}$ can be generated by a Finite State Machine. Give proof of this fact.

Hint: Suppose that this language is recognized by a Finite State with m states, say 25. Feed the machine a very long grammatical sentence, one with more than 25 words. You may then reason as follows:

(i) Since the sentence has more words than there are states in the machine, and since each word is recognized by a state, two words must be recognized by the same state.

(ii) This means that there is a 'loop' in the recognition process. But where there was a single loop, you may now have instead 2, 3, 4, ... n ... loops, yielding a number of new sentences that are predicted to be grammatical.

Your task is to show that this leads to a contradiction if the language to be generated is $\{a^n b^n : n \geq 1\}$.

Answer:

Question 4. One could try to infer from Question 3 that English cannot be generated by a Finite State Machine, because a subpart of it cannot be generated by any Finite State Machine.

Show that this reasoning is incorrect by displaying a language which:

(i) can be generated by a Finite State Machine, but

(ii) contains as a subpart a language that cannot be generated by any Finite State Machine.

Hint 1: This is an easy question.

Hint 2: Some very trivial languages can be generated by a Finite State Machine.

Answer:

Question 5. Suppose that language L is generated by the Finite State Machine M and that language L' is generated by Finite State Machine M' . Show that the intersection of L and L' is generated by some Finite State Machine.

Hint: You will have to consider a machine M^* whose states are:

$\{ \langle s, s' \rangle : s \text{ is a state of } M \text{ and } s' \text{ is a state of } M' \}$.

Your task will be to give an explicit definition, i.e. to give a statement of the following form:

word w read in state $\langle s_1, s'_1 \rangle$ leads to state $\langle s_2, s'_2 \rangle$ in M^* if and only if: _____

Answer:

Question 6.

a. Provide a language L generated by a Finite State Machine and such that:

the intersection of L and English is just: $\{(a \text{ wolf})^n (ate)^n : n \geq 1\}$

Answer:

b. Prove that English cannot be generated by any Finite State Machine.

Answer:

Question 7. Does the argument extend to French? (you may simply (a) mention some important sentences, and (b) give the sketch of an argument)

Answer:

Question 8. A productive process to create words in English is called compounding, and is illustrated in (6):

- (6)
- a. tennis -> table tennis
 - b. scientist -> rocket scientist
 - c. missile -> anti-missile missile

Interestingly, the process illustrated in (6c) can be iterated:

- (7)
- 1. anti-missile missile
 - 2. anti-(anti-missile missile) missile
 - 3. anti-(anti-(anti-missile missile) missile) missile
 - ...
 - n. (anti)ⁿ(missile)ⁿ⁺¹

Finite State Machines have sometimes been applied to generate English words. If compounding is indeed taken to generate words, what do you take the prospects of this enterprise to be?

Answer: