Informatics 1 Cognitive Science – Tutorial 1 Solutions

Frank Keller, Carina Silberer, Frank Mollica

Week 2

The goal of this tutorial is to familiarise yourself with terms and ideas about vocabulary and grammar and to start thinking about language.

1 Vocabulary and Grammar

Linguistics (the scientific study of human language) is an important part of cognitive science, and understanding language can help us understand more general phenomena in cognition. The following tasks will address the aspects of language you heard about in class and will give you the opportunity to do some practical exercises to get a better understanding of them.

1.1 Words and Rules

Many cognitive scientists believe that human linguistic knowledge consists of two different kinds of 'mental tissue':

- a **lexicon**, containing **words**

- a **grammar**, containing **rules**

Here is an example lexicon:

John	is a	proper noun
Mary	is a	proper noun
hold	is a	base verb
catch	is a	base verb
held	is a	past tense verb
caught	is a	past tense verb

And here is an example grammar:

a sentence	can consist of a noun phrase followed by a verb phrase
	S → NP VP
a verb phrase	can consist of a past tense verb followed by a noun phrase
	VP → V$_{past}$ NP
a noun phrase	can consist of a proper noun on its own
	NP → PN

Exercise 1

1. Draw out the lexicon as trees, like the ones in lectures.

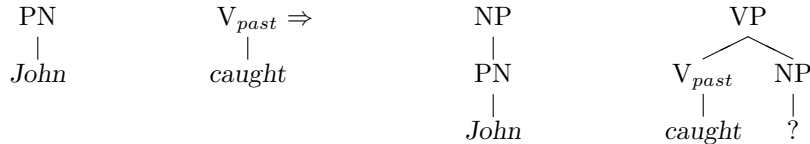| PN	PN	V	V	V$_{past}$	V$_{past}$
|	|	|	|	|	|
John	*Mary*	*hold*	*catch*	*held*	*caught*
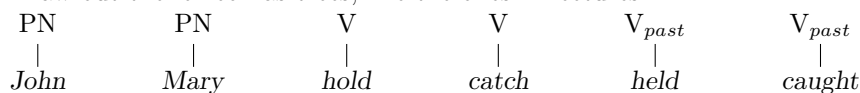
2. Likewise, draw out the grammar rules as trees.
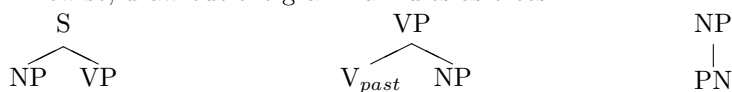
3. Use the grammar and lexicon to show that *John held Mary* is a grammatical
   sentence.

```
               S
             /   \
           NP     VP
           |     /   \
           PN  V_past  NP
           |    |      |
          John held    PN
                       |
                      Mary
```
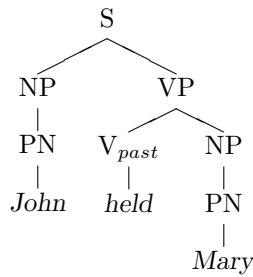
4. Show that **John caught* is **not** a grammatical sentence.

```
   PN           V_past ⇒        NP                  VP
   |            |               |                 /    \
  John         caught           PN            V_past    NP
                                |               |       |
                               John           caught    ?
```

5. Show that **Mary catch John* is **not** a grammatical sentence.

```
   NP      V      NP ⇒ There is no rule in the grammar covering base verbs (V).
   |       |      |
   PN    catch    PN
   |             |
  Mary          John
```
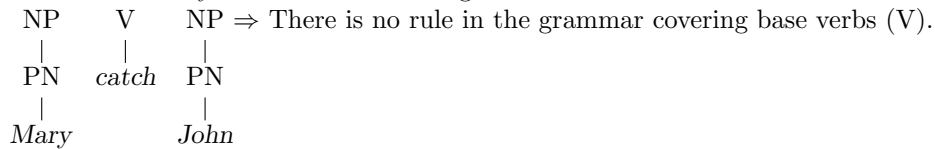
6. How many sentences are grammatical, according to this grammar and lexicon?
   Mary caught John. John caught Mary. Mary caught Mary. John caught John.
   Mary held John. John held Mary. Mary held Mary. John held John.
   ⇒ 8 sentences

1.2 Productivity & Reuse

Words and Rules theory embodies two broader principles that we discussed in class:
Productivity and Reuse. Lexical entries are memorized chunks that we consistently
reuse. They are productively combined using rules to create phrases and sentences.
In English, it's fairly straightforward to distinguish lexical entries as they are often
separated by spaces when written and there are not too many important prefixes and
suffixes.

For this exercise, let's consider Swahili, a language spoken in East Africa. In Swahili,
a single typed "word" can translate to a full sentence. However, these sentences are still
the result of productivity and reuse. Take a look at the following data:

atanipenda *'s/he will like me'*	atanipiga *'s/he will beat me'*
atakupenda *'she will like you'*	atakupiga *'she will beat you'*
atampenda *'she will like him/her'*	atampiga *'she will beat him/her'*
atatupenda *'she will like us'*	ananipiga *'she is beating me'*
atawapenda *'she will like them'*	anakupiga *'she is beating you'*
nitakupenda *'I will like you'*	anampiga *'she is beating him/her'*
nitampenda *'I will like him/her'*	amekupiga *'she has beaten you'*
nitawapenda *'I will like them'*	amenipiga *'she has beaten me'*
utanipenda *'you will like me'*	amempiga *'she has beaten him/her'*
utampenda *'you will like him/her'*	alinipiga *'she beat me'*
tutampenda *'we will like him/her'*	alikupiga *'she beat you'*
watampenda *'they will like him/her'*	alimpiga *'she beat him/her'*
wametulipa *'they have paid us'*	atakusumbua *'she will annoy you'*
tulikulipa *'we paid you'*	unamsumbua *'you are annoying him/her'*

Exercise 2

1. Identify the parts of the words that are frequently reused–i.e., the lexicon.
   See table below.

2. For each lexical entry, identify its "part of speech" (Verb/Subject/Object pronoun/Tense).
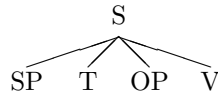   Verb: lipa, penda, sumbua, piga
   SubjectPronoun (SP): ni, tu, u, wa, a
   ObjectPronoun (OP): ku, m, wa, tu, ni
   Tense (T): me, na, ta, li

3. The data can all be explained by a single rule. Draw out that rule as a tree.

```
         S
      ___/|\___
     /   /  \   \
   SP   T   OP   V
```

4. Use the grammar and lexicon to provide the Swahili words for *You are annoying me* and *I am paying him*.
   You are annoying me: u-na-ni-sumbua
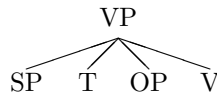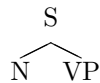   I am paying him: ni-na-m-lipa

5. How many sentences are grammatical, according to this grammar and lexicon?
   $5SP \cdot 4T \cdot 5OP \cdot 4V = 400$ sentences.

6. In Swahili, *the lion has chased me* translates to *simba alinifukuza*. Expand the grammar and lexicon to account for this new data.
   Simba is a noun.
   -fukuza is a verb.

```
     S                              VP
    /\                          ___/|\___
   N  VP                       /   /  \   \
                             SP   T   OP   V
```

ni- *I*	**tu-** *we*	**-me-** past tense
-lipa *pay*	**-penda** *like*	**-na** present tense
a- *she*	**-sumbua** *annoy*	**-na-** future tense
-ni- *me*	**-m-** *himher*	**-li-** present perfect tense
-piga *beat*	**wa-** *they*	**u-** *you* (subject)
-tu- *us*	**-wa-** *them*	**-ku-** *you* (object)

Let's take a look at another language to get some more leverage on productivity and reuse. While it often seems that we have an English word for every animal in the world, there was a point when we first encountered an animal and did not have a word for it. In English, we typically either borrow or create an arbitrary and conventionalized word (e.g., axolotl). In scientific communities, there are conventionalized rules for how to name a new animal.

Take a look at this sentence in Mohawk, an indigenous language to Northeastern US/Canada (Figure 1) which shows how Mohawk speakers refer to porcupines.

Nek	tsi	akitshé:nen	aneǹ:taks	è:rhar	raotíkhwaʔ
just	that	my-pet	it-evergreen-eat-s	dog	their-food

wahí:nonteʔ	tsi	wahatehià:ron.
past-I/him feed	that	past-he-grow

I raised my pet porcupine on dog food.

Figure 1: The top like in Latin alphabet script. The bottom line reflects a word-by-word gloss. The translation is given at the end. Taken from Mithun (1984).

Exercise 3

1. Based on Figure 1, does Mithun analyse *porcupine* as a morphological object or a psychological one?
   Morphological object because it's a linguistic description, not a processing claim.

2. The Mohawk word for *porcupine* provides insight about the productive process involved in naming the animal. Do you think Mohawk speakers store the full word or productively construct the word as needed? Why?

   There is no data on the frequency with which Mohawk speakers deal with porcupines so there is no answer but it's a useful discussion. Here are some relevant points:

   Productively construct:

   - Porcupines may be really low frequency. Why store a word that you'll probably use once in a lifetime.
   - If porcupines are rarely in the environment being productive allows you to adapt your reference. For example, would a porcupine still be named after eating evergreens when there are no evergreens around because of seasons?

   Reuse:

   - It takes effort to productively name something. If porcupines are frequent, why would you want to constantly repeat the computation?
   - If we store porcupine as a lexical entry, we also fix the conventionalized expression. This way we can talk about porcupines even in winter when there are no evergreens and when porcupines aren't physically present to provide context.