

# Informatics 1 CG – Tutorial 5

Mirella Lapata, Carina Silberer

Week 6

In the lectures, we covered the vector space model and the associationist view of word meaning. Here, we will try to build a vector space model by hand in order to understand the semantic space construction algorithm. We also discussed syntactic parsing, the process of turning sequences of words into a syntactic representation. Parsing is the first step towards understanding sentences. In this context, we talked about global and local ambiguities that the syntactic processor can encounter. The goal of this tutorial is to understand the two types of ambiguities with respect to sentence processing.

## 1 Vector Space Models

Below you are given a document. Create a word-by-word co-occurrence matrix for the target words

*almond, nut, tree, and world*

within a window of +/-2 **content** words (i.e., only count nouns, verbs, adjectives and adverbs) using the words

*healthy, market, eat, process, farmer, crop, grow, plant, and value*

as context. Don't distinguish between different word forms (e.g., count *driving* as an occurrence of the base form *drive*).

Why are almonds so expensive?

Almond sales are booming because of their well-advertised health benefits, but could drought in California lead to a world shortage of the nut?

It has been likened to a modern-day gold rush. The growth in the popularity of nuts as a healthy snack has seen a boom in business for California's almond farmers. Eighty-two per cent of the world's almonds come from America's Golden State, where it is the leading agricultural export.

"Nut crops and almonds particularly have risen in value as the world has realised the nutritional value of eating almonds and how good they are for our bodies," explains David Phippen, a life-long almond grower in the prime agricultural area of central California.

Phippen's solar-powered farm is a partnership between five families. It is one of the first in the world to use robots, designed using Nasa technology, to sort good almonds from bad. It is a highly mechanised process with sophisticated irrigation systems in the orchards but, above all, almond production is dependent on the climate.

The region - which is about a 90-minute drive from Silicon Valley - is one of the few places in the world where almond trees will grow. It has the perfect combination of a cold - but not too cold - winter, which allows the trees to lie dormant, followed by a mild spring that encourages them to wake up and bloom. Crucially, the trees need about 500-700 hours of dormancy followed by a frost-free period when they burst into life - usually around Valentine's Day.

Farmers like Phippen have responded to the global surge in business - especially from the UK and emerging markets like China and India - by expanding their orchards. "The value of each kernel has gone up dramatically and growers are looking for the best return on their investment so they're still planting almond trees at an alarming rate," he says. "If you decided to plant an orchard right now, you would wait two years for available root stock to actually plant."

source: <http://www.bbc.co.uk/news/magazine-26118225>

## 2 Understanding Sentences: Parsing

You are given the sentence

You made her duck.

and a little grammar of English:

Grammatical rules	Lexical rules	
S → NP VP	Det → a   the   her	(determiners)
NP → Det N	N → man   park   duck   telescope	(nouns)
NP → Det N PP	Pro → you	(pronoun)
NP → Pro	V → saw   walked   made	(verbs)
VP → V NP PP	Prep → in   with   for	(prepositions)
VP → V NP		
VP → V		
PP → Prep NP		

### Exercises

1. Produce a syntax tree for the sentence on the basis of the given grammar.
2. There is another, ditransitive, interpretation of the sentence, in which *her* and *duck* are interpreted as two objects of *made*. The given grammar does not account for that however.
  - (a) Extend the grammar by adding rules that allow for this interpretation. Draw the corresponding syntax tree.
  - (b) For which type of ambiguity is this an example?

### 3 Understanding Sentences – Garden Path Sentences

In the lectures we have discussed some unusual constructions called garden-path sentences. These are grammatically correct sentences that are constructed in such a way that the most likely initial parse is incorrect. During the parsing of the sentence, the reader encounters an unexpected word given the parse structure they have constructed so far. At that point, the reader will have to backtrack and reanalyse either part of the parse or the entire parse tree.

#### Exercises

1. Can you create your own garden-path sentences (at least four)? Mark the breaking point and present the initial (most likely) and final (correct) interpretations.
2. Can you construct a grammar that will parse your garden path sentence? [Note that you will have to include all the part-of-speech tags for the ambiguous word(s).]