Introduction to Natural Language Syntax and Parsing: L95
Lecture 2

Ann Copestake (standing in for Simone Teufel)

Department of Computer Science and Technology
University of Cambridge

October 2019

POS tag: N, V, A, Det, P, Num, ?

Eliud Kipchoge has become the first athlete to run a marathon in under two hours, beating the mark by 20 seconds.

The Kenyan, 34, covered the 26.2 miles (42.2km) in one hour 59 minutes 40 seconds in the Ineos 1:59 Challenge in Vienna, Austria on Saturday.

Polysemy

Time flies like an arrow.
Fruit flies like a banana.
Kim gave her dog biscuits.

POS tag sequence is often the same for ambiguous sentences:

The bank is 200 metres away.
I saw a man with a telescope.

Note:

I saw wood.

VBD vs VBP in PTB tagging scheme.

Polysemy

Time flies like an arrow.
Fruit flies like a banana.
Kim gave her dog biscuits.

POS tag sequence is often the same for ambiguous sentences:

The bank is 200 metres away.
I saw a man with a telescope.

Note:

I saw wood.

VBD vs VBP in PTB tagging scheme.

Polysemy

Time flies like an arrow.
Fruit flies like a banana.
Kim gave her dog biscuits.

POS tag sequence is often the same for ambiguous sentences:

The bank is 200 metres away.
I saw a man with a telescope.

Note:

I saw wood.

VBD vs VBP in PTB tagging scheme.

Polysemy

Time flies like an arrow.
Fruit flies like a banana.
Kim gave her dog biscuits.

POS tag sequence is often the same for ambiguous sentences:

The bank is 200 metres away.
I saw a man with a telescope.

Note:

I saw wood.

VBD vs VBP in PTB tagging scheme.

Polysemy

Time flies like an arrow.
Fruit flies like a banana.
Kim gave her dog biscuits.

POS tag sequence is often the same for ambiguous sentences:

The bank is 200 metres away.
I saw a man with a telescope.

Note:

I saw wood.

VBD vs VBP in PTB tagging scheme.

Idioms

Mostly idioms have normal syntax:

We have hit a brick wall.
The cat is out of the bag.

But there are exceptions:

They are well off.

PTB tagging guidelines say that *off* is RP (particle), but impossible to give good tags on a word-by-word basis. *well off*, *better off* etc behave like adjectives if considered as single units:

The better off inhabitants of the village protested against the tax rise.

Particle vs preposition

The man up the ladder fell.
Kim ran up the stairs.
Kim ran up a large bill.
Kim slipped up.
Kim washed up the dishes.
Kim washes the dishes up.

NB: PTB guidelines say that *to* is always tagged TO.

Particle vs preposition

The man up the ladder fell.
Kim ran up the stairs.
Kim ran up a large bill.
Kim slipped up.
Kim washed up the dishes.
Kim washes the dishes up.

NB: PTB guidelines say that *to* is always tagged TO.

Particle vs preposition

The man up the ladder fell.
Kim ran up the stairs.
Kim ran up a large bill.
Kim slipped up.
Kim washed up the dishes.
Kim washes the dishes up.

NB: PTB guidelines say that *to* is always tagged TO.

Particle vs preposition

The man up the ladder fell.
Kim ran up the stairs.
Kim ran up a large bill.
Kim slipped up.
Kim washed up the dishes.
Kim washes the dishes up.

NB: PTB guidelines say that *to* is always tagged TO.

Particle vs preposition

The man up the ladder fell.
Kim ran up the stairs.
Kim ran up a large bill.
Kim slipped up.
Kim washed up the dishes.
Kim washes the dishes up.

NB: PTB guidelines say that *to* is always tagged TO.

Particle vs preposition

The man up the ladder fell.
Kim ran up the stairs.
Kim ran up a large bill.
Kim slipped up.
Kim washed up the dishes.
Kim washes the dishes up.

NB: PTB guidelines say that *to* is always tagged TO.

Particle vs preposition

The man up the ladder fell.
Kim ran up the stairs.
Kim ran up a large bill.
Kim slipped up.
Kim washed up the dishes.
Kim washes the dishes up.

NB: PTB guidelines say that *to* is always tagged TO.

Tokenization

- ► Usually for English, words are separated by spaces.
- ► Standard PTB tokenization: split off possessive 's, put spaces round punctuation in general.
- ► But formulae etc:
  buta-1,3-diene
- ► Need to think about this for the exercises!

To do

- ► Read Section 5.1–5.6 in 'introduction to linguistics' and attempt exercises 1–3 in section 5.9 (note: this material is dense).
- ► Next lecture is on Thursday
- ► Logic worksheet
- ► Assignment 1: deadline October 21 (remember to write notes on problematic cases).