CS224N: Introduction to Syntax

We want to know how meaning is mapped onto what language structures. Commonly in English in ways like this:

> [Thing The dog] is [Place in the garden]
> [Thing The dog] is [Property fierce]
> [Action [Thing The dog] is chasing [Thing the cat]]
> [State [Thing The dog] was sitting [Place in the garden] [Time yesterday]]
> [Action [Thing We] ran [Path out into the water]]
> [Action [Thing The dog] barked [Property/Manner loudly]]
> [Action [Thing The dog] barked [Property/Amount nonstop for five hours]]

There are considerable difference in other languages.

Warlpiri exx.

Alopiji nga-rnu kapiyali-rli
ellops eat-PAST gavial-ERG
'the gavial ate the ellops'

Alopiji wirijarlu pali-ja ngapa-ngka
ellops enormous die-PAST water-LOC
'the big ellops died in the water'

Categories, grammatical relations and semantic roles

1a. Word categories: 'Traditional parts of speech'

Noun	Names of things: *boy, cat, truth*	Verb	Action or state: *become, hit*
Pronoun	Used for noun: *I, you, we*	Adverb	Modifies V, Adj, Adv: *sadly, very*
Adjective	Modifies noun: *happy, clever*	Conjunction	Joins things: *and, but, while*
Preposition	Relation of N: *to, from, into*	Interjection	An outcry: *ouch, oh, alas, psst*

Modern linguistics is based on these, but there are some differences, as linguists attempt to identify parts of speech consistently on the basis of form, by looking at grammatical properties of distribution and selection. There are certain changes from traditional grammatical classifications, e.g.:

- a class of determiners is recognized (which includes the traditional *article*)
- the class of prepositions is expanded (gathering some of what were adverbs and conjunctions)
- for computational work, there are also a lot of practical details, like numbers, punctuation

1b. Phrasal categories:

Sentences have parts, some of which appear to have subparts. These groupings of words that go toether we will call constituents (how do we know they go together? – see further below).

> I hit the man with a cleaver: *I hit [the man with a cleaver] / I hit [the man] with a cleaver*
> You could not go to her party: *You [could not] go to her party / You could [not go] to her party*

For constituents, we usually name them as phrases based on the word that *heads* the constituent:

the man from Iron Bark	is a Noun Phrase (NP) because the head *man* is a noun
extremely clever	is an Adjective Phrase (AP) because the head *clever* is an adjective
down the river	is a Prepositional Phrase (PP) because the head *down* is a preposition
killed the rabbit	is a Verb Phrase (VP) because the head *killed* is a verb

Note that a word is a constituent (if a little one). Sometimes words also act as phrases. In:

Joe grew potatoes

Joe and *potatoes* are both nouns and noun phrases. Compare:

The man from Iron Bark grew beautiful russet potatoes.

We say *Joe* counts as a noun phrase because it appears in a place that a larger noun phrase could have been.

2. Grammatical function/relation:

In: Joe's young cousin carried the huntsman out of the house

Joe's young cousin	is the SUBJECT of the verb *carried*	[roughly, doer that precedes verb]
the huntsman	is the OBJECT of the verb *carried*	[roughly, undergoer that follows verb]
Joe	is the POSSESSOR of the noun *cousin*	[the owner of a noun]
young	is a MODIFIER of the noun *cousin*	[expresses an attribute of a head]

3. Semantic role:

In the same sentence,

Joe's young cousin	is the Agent/Actor (the one who does the action)
the huntsman	is the Theme/Undergoer (the thing that undergoes a change of state or motion)
out of the house	is a Locative (place where something happens)

4. Pragmatic functions:

Consider the following discourse. In it we can observe various pragmatic functions.

There was a river containing an ellops and a gavial. The gavial liked basking on the beach. It had a large appetite. As for the ellops, it liked swimming. Backstroke was the way it swam.

Topic-Comment: *it* refers to the expected topic (about which a comment is made);
            as for is used as a switch topic construction
Focus-Presupposition: *Backstroke* is focussed with the rest of the sentence a presupposition.
Presentational focus: *There was* introduces *a river* with presentational focus.

We want to distinguish these notions (levels) clearly, but note that there is a great deal of redundancy between them. For instance, the SUBJECT is normally the Agent, and similarly the SUBJECT is normally the Topic. Linguists spend a lot of time trying to identify and explain these correlations.

Phrasal Categories – Evidence for Constituency

Why phrasal categories? They allow us to give a better description of language structure.

There is morphological evidence (phrasal affixes like 's) and semantic evidence (*The president could not complete the review*), but overwhelmingly we use distributional evidence.

A given string of elements is a constituent just in case it has one or more of the following properties:

(a) Distribution:
(i) External: it behaves distributionally as a single structural unit – i.e., it occurs as a single unit in a variety of sentence positions, e.g., noun phrases in:

after the verb	*before the verb*	*following a preposition*
I saw the statue	The statue fell down	It is under the statue

Often this includes placement in pragmatically marked positions, such as by preposing and postposing.
John talked to the children about drugs
John talked about drugs to the children
*John talked drugs to the children about
?To the children, John talked about drugs.
(ii) Internal: It seems to have a regular internal structure. We can test this by doing substitution and expansion.
the red book, this red book, many red books, *red this book
(b) No intrusion: It does not permit intrusion of parenthetical elements internally (e.g., sentence adverbs like *surely* or phrases indicative of speaker attitude, like *I think*). Intrusion is usually only at the boundaries of major phrasal constituents
*The, I think, man went home.
The man, I think, went home.
(c) Coordination: It can be coordinated with another string (without needing huge intonational marking). Normally but not always coordination is restricted to the same category.
The man and the horse went home
The happy and wise student always excels.
*I saw the happy and that clever student.
(d) Pro-forms: It can be replaced by a pro-form (forms like *it, what, there, (do) so*)
I saw *the man*. *Who* did you see? Did Mary see *him*?
I sat *on the box*. Robin sat *there* too.
(e) Sentence fragments: Constituents can be used as sentence fragments:
What did you eat? A loaf of bread
What are your weekend plans? Going to the library.
(f) Phrases can sometimes undergo ellipsis (i.e., be left out):
John didn't win the monkey, but his brother might.

Example Adjective Phrases:

(He is) so very proud of his daughter
(She is) quite sure that Mary will win
(It doesn't seem) that advantageous for us

Mapping of semantics onto syntax

Words have details of their individual meaning. That's their *lexical semantics*. A verb can be thought of semantically as a predicate which takes certain arguments. *Give* takes a *giver,* a *thing given,* and a *recipient* while realizing a *transfer* event. All of these can be realised:

        Jo gave Kate a donkey.
        Jo gave a donkey to Kate.

But they don't have to be (and then have assumed referents – money to charity)

        Jo gave to the Salvation Army.
        Jo gave blankets.
        Jo gave (generously).

Even if they are all realised, there is a choice of realisation patterns (as shown above). For most verbs of transfer, both these frames are available. But there are subtle semantic things going on. You don't get both frames for more idiomatic uses of *give*, or ones that imply affectedness:

Jo gave Kate a hairy eyeball	*Jo gave a hairy eyeball to Kate
That lecturer gives me the sh**s	*That lecturer gives the sh**s to me.
Kate gave her best friend the measles	??Kate gave the measles to her best friend.

A (very) simple English grammar:

S → NP VP

VP → V (NP (NP)) PP*

NP → { Pron
      { PN PN*
      { ({Det/PossP}) AP* N (PP)

PossP → NP Poss

PP → P NP

AP → (AdvP) A

AdvP → (AdvP) Adv

ate	V
John	PN
you	Pron
the	Det
that	Det
's	Poss
extremely	Adv
quite	Adv
armadillo	N
corner	N
green	A
in	P

Implicit in this has been a "two level" theory of structure. We have word classes (N, V, A, P) and we have (maximal) phrasal classes (NP, PP, AP, Adv_mP, S?). But we don't have anything in between.