Is There a General Structure for Grammars?

SiG workshop, CVPR 2009, June 21

David Mumford

Brown University

Outline

- A quick look at how diverse the grammars of natural languages are.

- Is there anything to hang onto? Re-usable parts, hierarchical assembly seem the core.

- Shouldn't grammars apply to input signals, output actions/plans (e.g. speech) and internal thoughts?

- Planning and causal arrows, examples

- What is a plausible common framework?

- Thoughts, labels and Roget's Thesaurus, examples

- Final remarks

The standard parse tree one uses for English and related Indo-European languages does *not* apply to a fair proportion of the languages of the world.

Note that it presupposes a) contiguous sets of words forming grammatical parts of the sentence, b) lexical morphemes – words (albeit with some attributes) for each part of the meaning.
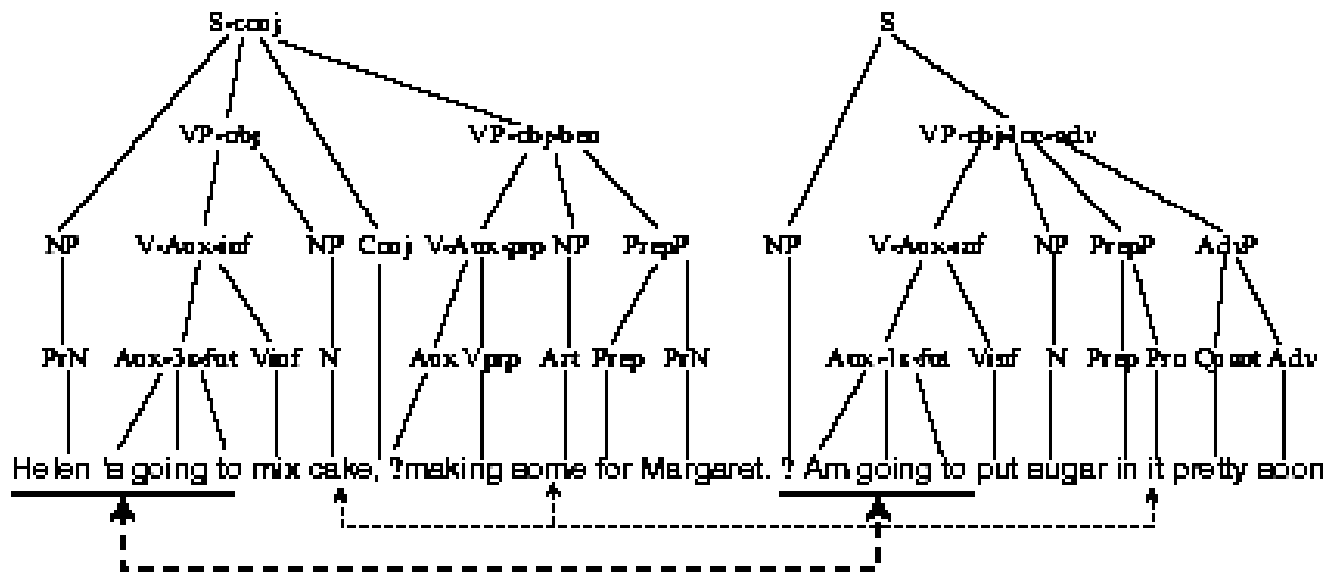
Can have <u>free word order</u> in highly inflected languages

LATIN (Virgil):

ultima Cumaei   venit   iam carminis aetas
last(Nom) Cumae(Gen) come(3spast) now song(Gen) age(Nom)
'the last age of the Cumaean song has now arrived'

RUSSIAN:Tanja ubila Mašu, 'Tanya killed Masha'
can be said correctly in all six orders!

DYIRBAL (from Southeastern Australia):
When asked to repeat a sentence, native speakers rearrange the words, unaware of any change.

Polysynthetic languages: many <u>grammatical affixes</u> are added to one <u>lexical morpheme</u> creating huge words, even one word sentences.

SIBERIAN YUPIK: *Angya-ghlla-ng-yug-tuq*
Boat-[AUGMENTATIVE]-[ACQUISITIVE]-[DESIDERATIVE]-[3SING]
'He wants to acquire a big boat'   (B.Comrie)

BININJ GUN-WOK (Australia): *Barri-kanj-ngu-nihmi-re*
'they go along eating meat',   [they-meat-eat-ing-go]
BUT maximum of <u>one level of embedding</u>

 [They stood [watching us [fight]]] is a doubly recursive parsed English sentence. It becomes a 3 word sentence:
they-stood they-were-watching-us we-were-fighting-each-other
with *NO* recursion.   (N.Evans)

Now language grammars look much more like grammars of images

1. The working hypothesis is that there is a parsed 'mentalese' (e.g. Jackendoff's *conceptual structure*) which people translate into linear phonological strings in hugely variable ways. This linear structure may or may not be used to group related constituents. In images, the linear structure is gone.

2. Either words or affixes may carry meaning. Likewise, in images, there is no basic word-like level of granularity.

3. Recursion – used by Chomsky to argue for the need of a 'language gene' – may or may not be extensively used. Images usually have minimal recursion.
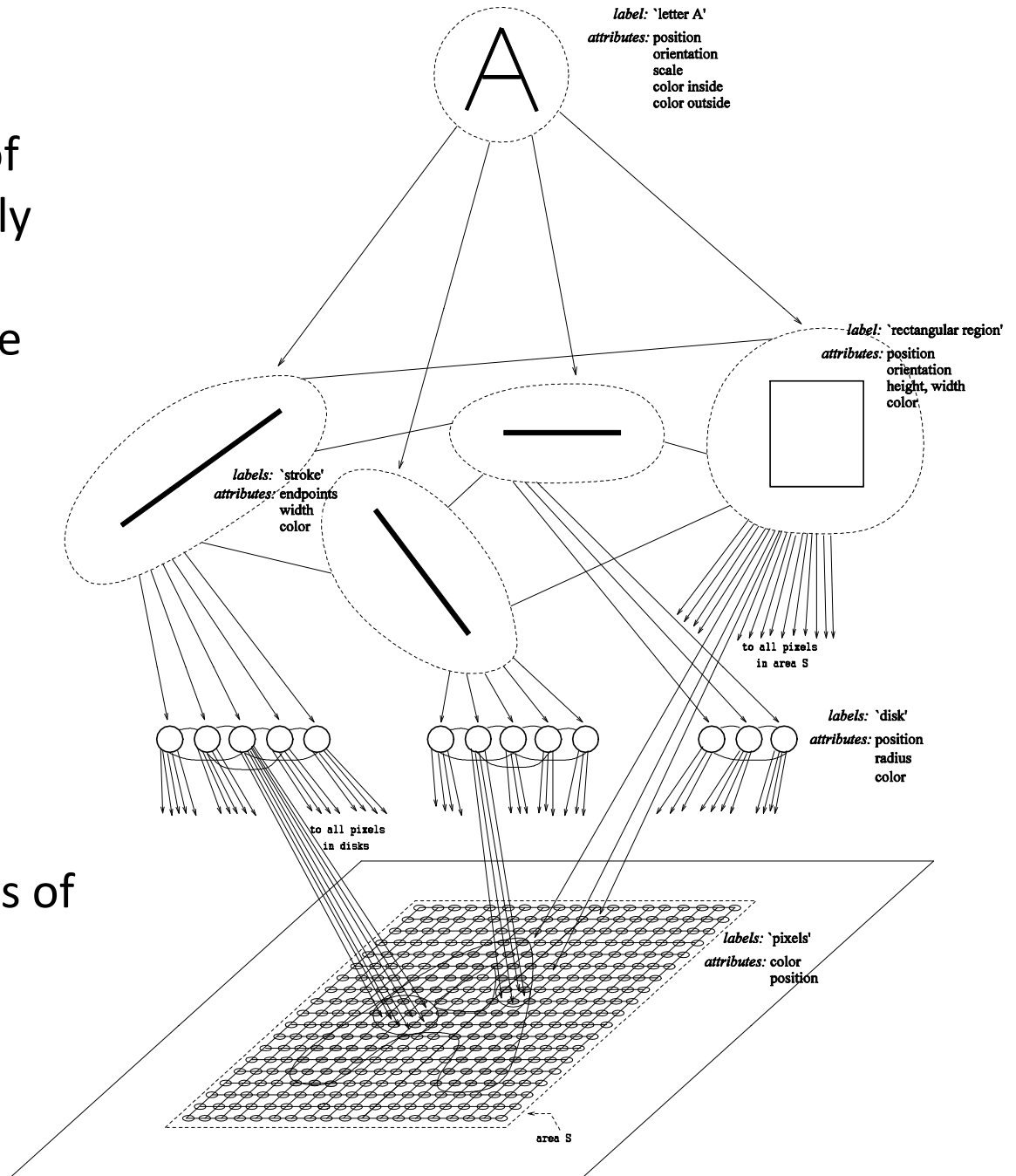
A parse graph (in both language or vision) has

a) a *node* for each subset of the signal that is naturally grouped together

b) A *vertical edge* when one subset contains or a *horizontal edge* if it is adjacent to another.

c) *Productions*: rules for expanding a node.

Ex: the letter 'A' with four hierarchical levels

a) Pixels

b) Disks tracing medial axis of strokes

c) Complete strokes and background

d) Complete letter

label: `letter A'
attributes: position
orientation
scale
color inside
color outside

label: `rectangular region'
attributes: position
orientation
height, width
color

labels: `stroke'
attributes: endpoints
width
color

to all pixels
in area S

labels: `disk'
attributes: position
radius
color

to all pixels
in disks

labels: `pixels'
attributes: color
position

area S

A particular characteristic of image grammars is the ubiquity of occlusion causing background objects to be partially hidden.

The grammar must reconstruct the hidden *amodal* contours. The dotted vertical line represents the inclusion of the hidden layer behind the head in the sky
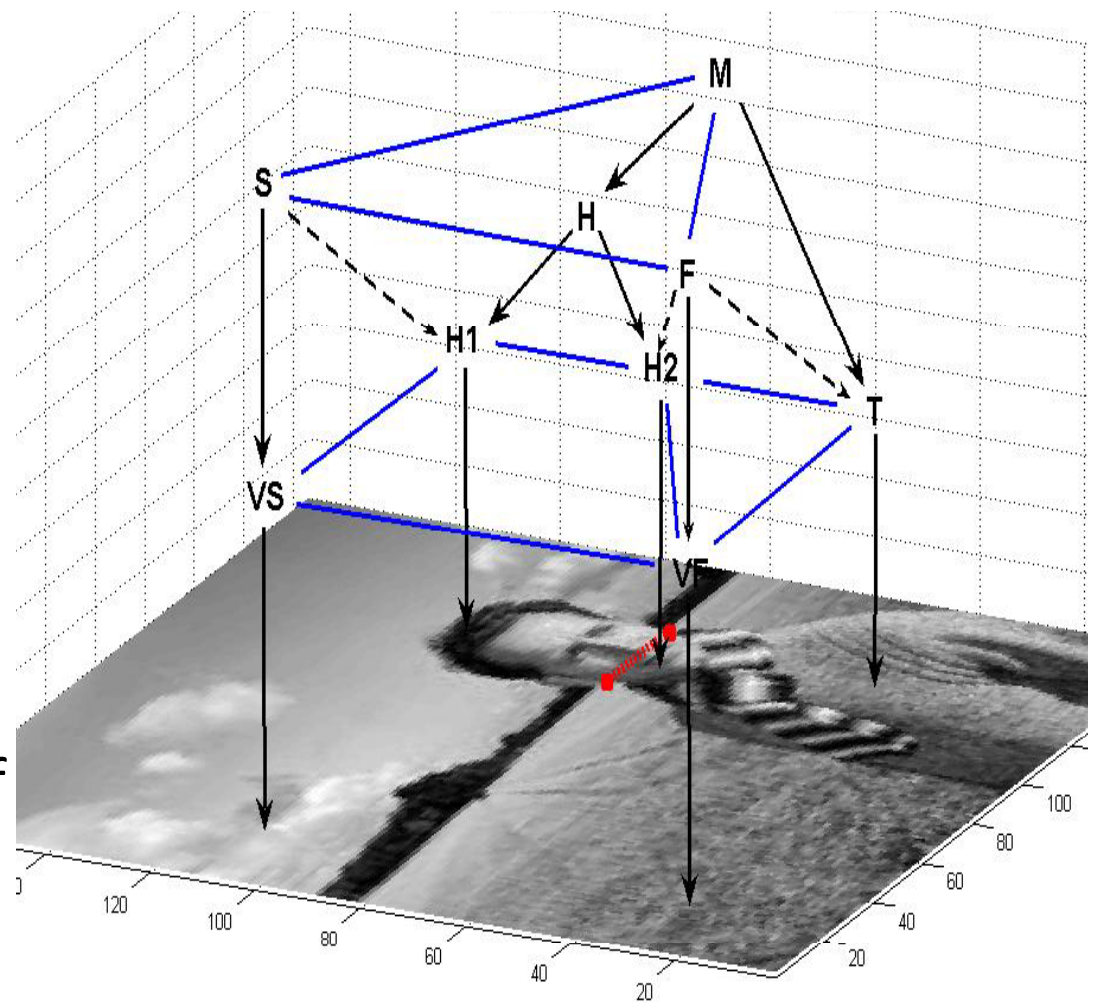
Here M=man, T=torso, H=head, VF=visible field, F=field, VS = visible sky, S=full sky incl. part behind head, red line = amodal contour separating occluded sky, field. (Language also has words missing from spoken form which must be inferred.)

What are the key properties of *all* grammars?

1. <u>Re-usable parts</u>: parts which reappear in different situations, modified *attributes* but with essentials the same

2. <u>Hierarchy</u>: these re-usable parts often can be grouped into bigger re-usable parts which also reappear, with the same subparts. The bigger piece has *slots* for its components, whose attributes that must be consistent.

BUT these are essential components of all cognition!!

# Grammars everywhere: Input/Output/In between

- Seeing an image, hearing speech are special cases of signals which can be parsed into a hierarchy of re-usable parts.

- Speaking language is a special case of a sequence of voluntary actions, mentally planned, made up of re-usable sub-actions, which may or may not be carried out.

- Mulling over stuff, imagining what's happening somewhere, exploring plans are pure mentalese made up of re-usable fragments of thoughts.

- What sort of common structure do all these share?

A third class of parse graphs: Spiegelhalter-Lauritzen +Pearl's Causal Diagrams

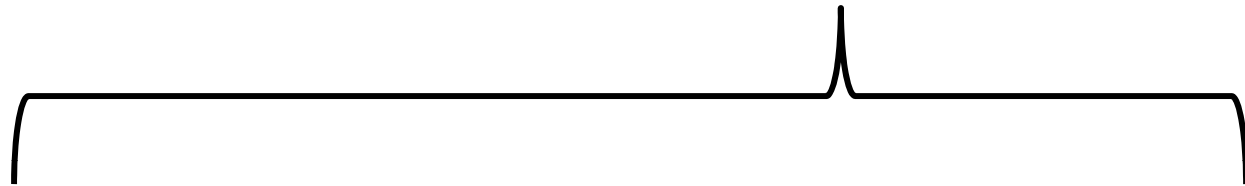1. Nodes are events or facts about situations
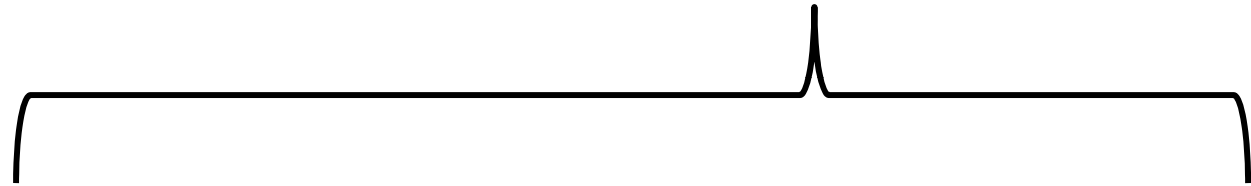2. Edges are oriented causal links
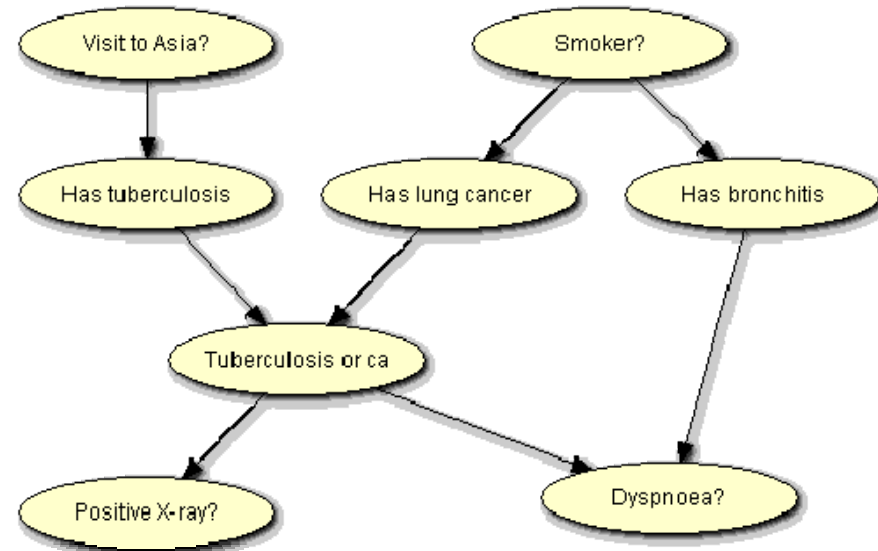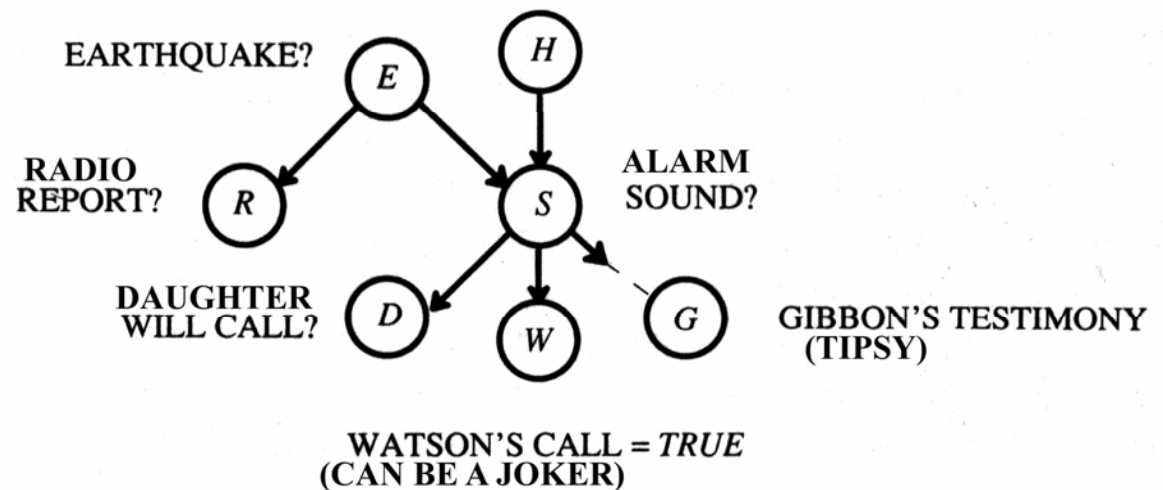
Let's merge vision parse graphs with Pearl's Causal Diagrams

Events of the day: big party ⟶ drunk ⟶ accident ⟶ arrest

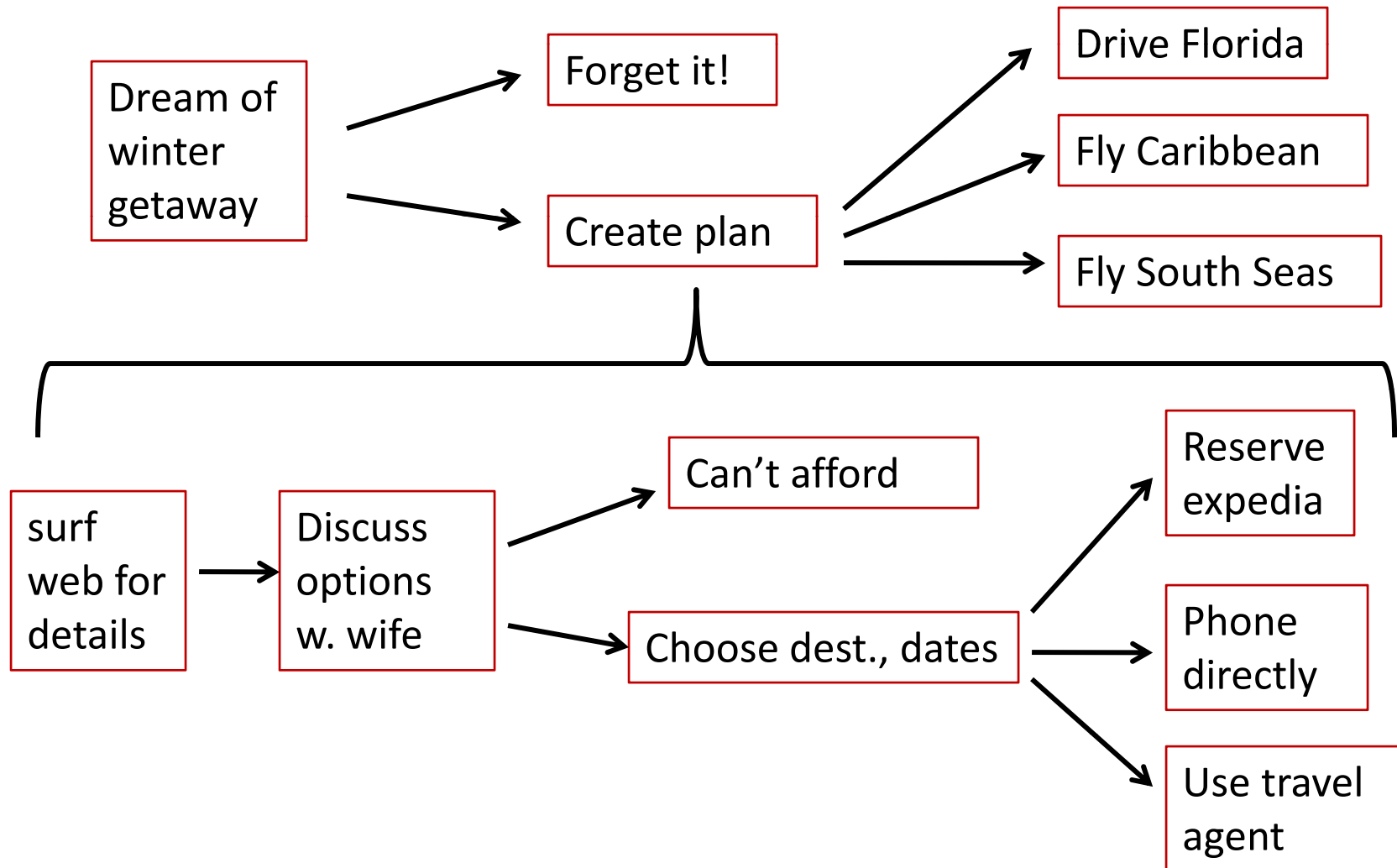**Driving acts:** turn left ⟶ see oncoming car ⟶ swerve ⟶ hit tree

Nervous System: V1-resp. ⟶ basal ganglia resp. ⟶ Arm muscles

1. Horizontal edges are causal
2. Vertical edges group events or facts about situations into bigger grouping$_s$

Planning involves 'causal/or' links plus hierarchical choices

Dream of winter getaway → Forget it!

Dream of winter getaway → Create plan

Create plan → Drive Florida

Create plan → Fly Caribbean

Create plan → Fly South Seas

surf web for details → Discuss options w. wife

Discuss options w. wife → Can't afford

Discuss options w. wife → Choose dest., dates

Choose dest., dates → Reserve expedia

Choose dest., dates → Phone directly

Choose dest., dates → Use travel agent

General idea for parsing planning and causal inference

1. Nodes are events or actions (esp. your own actions) or facts about past, present, future real or hypothetical situations

2. Horizontal edges are temporally asymmetric causal links

3. Vertical edges are groupings into larger sets of connected actions and sequences of their effects.

Note: this can include the act of speaking, forming phonemes, words, grammatical phrases – the only necessarily *linear* part of language

In all grammars, nodes have labels and attributes.

- Each node is an *instance* of its label and each label can occur multiple times in a parse graph.

- Attributes encode consistency requirements among the children and grandchildren of a node.

- In images, labels include 'edgelet', 'T-junction', 'texture', 'rectangle', 'letter', 'face', 'sky'; attributes can be 'color', 'endpt coord', 'orientation', 'lighting'

- In sentences, labels include 'noun', 'verb phrase', 'rel.clause'; attributes can be 'person', 'tense'.

- In planning, labels include 'sound', 'game', 'vol. action'; attributes can be 'time interval', 'location', 'in world *x*'.

Are there parse graphs for all thoughts?

First, what is a *thought*? My guess:

## **A constellation of objects, events and actions in one or more worlds**

In perceptions and actions, the only world is the 'here-now'. But in thoughts (and plans), you consider other locations in space and past and future time and very crucially the internal world of another agent as well as fictional worlds (e.g. '*War and Peace*'). We treat this as an *attribute*.

The constellation has natural subsets, groupings which re-occur in many other situations and these can be taken as *nodes*.

Some groupings are part of bigger ones, giving rise to vertical edges. Each part occupies a specific *slot* in the whole. Some groupings have relative positions in space or time: if they are 'adjacent' or 'causal', this can be drawn by horizontal edges. Example:

"John turned out not to be as tall as he thought he was"
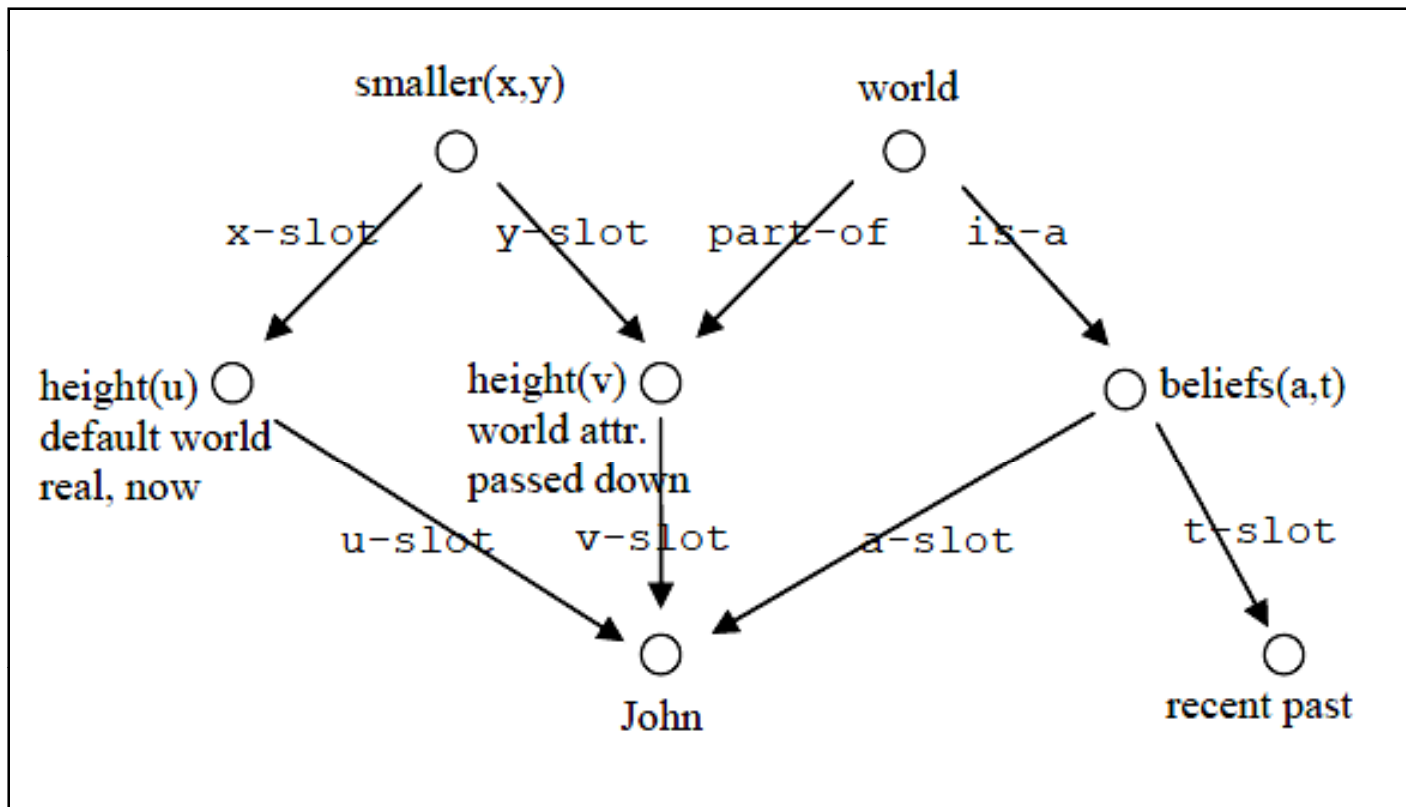
But we're going to need an awful lot of labels for thought: Roget's Thesaurus has one of the best classification systems for <u>all words</u>, hence for labels of nodes

I. Abstract relations (Existence, state, relations, quantity, order)

IIa. Time, Change and Causality
IIb. Space, Form and Motion

IIIa. Matter (animal-vegetable-mineral)
IIIb. Sensation (the six senses, esp. sound and light)

IVa. Intellect (ideas, mind)
IVb. Communication and language (incl. teaching, arts, speech)

Va. Individual Volition (will, tools, activities, health)
Vb. Intersocial Volition (hire, promise, property, money, laws)

VIa. Affections, Emotions (incl. marriage)
VIb. Values and Morals (incl. religion)

Final words

- What is needed are some micro-worlds in which to work out parse graphs which include perceptions and actions with both space-time components and hierarchy.

- Chomsky talked of a tightly constrained 'Universal Grammar' for language alone. Instead, I think we use a much broader and more universal UG with the shape I have sketched.

- Much of this UG is present in animal perception and actions, so the syntax of language now appears phylogenetically very old, not a human mutation.