Plans and the (Predicate Argument) Structure of Behavior

Mark Steedman

19th April 2015

School of informatics

Outline

- Introduction
- I: Planning
- II: From Planning to Semantics
- III: The Problem of Content
- IV: Hanging Language onto (Human) Planning
- Conclusions

Introduction

- There is a long tradition associating language and other serial cognitive behavior with an underlying motor planning mechanism (Piaget 1936, Lashley 1951, Miller *et al.* 1960).
- The evidence is evolutionary, neurophysiological, and developmental.
- It raises the possibility that language is much more closely related to embodied cognition than current linguistic theories of grammar suggest.

Introduction

- I'm going to argue that practically every aspect of language reflects this connection transparently, and that both cognitive and linguistic theories should be adjusted accordingly.

- The talk discusses this connection in terms of planning as it is viewed in Robotics and AI, with some attention to applicable machine learning techniques (Steedman 2002a,b).

- Work In Progress under ERC Advanced Fellowship 249520 GRAMPLUS and EU grant Xperience

Introduction

- The paper will sketch a path between representations at the level of the grounded sensory manifold and perceptron learning to the mid-level of plans and explanation-based learning, and on up to the level of language grammar and parsing model learning.

- At the levels of planning and linguistic representation, two simple but very general combinatory rule types, **Composition** (the operator **B**) and **Type-Raising** (the operator **T**) will appear repeatedly.

$$\mathbf{B}fg \equiv \lambda x.f(gx) \qquad\qquad \mathbf{T}a \equiv \lambda f.fa$$

- Human planning requires an additional element, in the form of plan variables, which also provides the basis for distinctively human language..

I: Plans and the Structure of Behavior

- Apes really can solve the “monkeys and bananas” problem, using tools like old crates to gain altitude in order to reach objects out of reach.

Figure 1: Köhler 1925

Figure 2: Köhler 1925

What does it Take to Plan?

- Such planning involves
  - Retrieving appropriate actions from memory (such as piling boxes on top of one another, and climbing on them),
  - Sequencing them in a way that has a reasonable chance of bringing about a desired state or goal (such as having the bananas).

☡ It is qualitatively different from Skinnerian shaping of purely reactive behavior in animals like pigeons—cf.
`http://www.youtube.com/watch?v=mDntbGRPeEU`

What does it Take to Plan?

- Köhler showed that, in apes at least, such search seems to be
  - *object-oriented*—that is, reactive to the presence of the tool, and
  - *forward-chaining*, working forward breadth-first from the tool to the goal, rather than backward-chaining (working from goal to tool).

- The first observation implies that actions are accessed via perception of the objects that mediate them—in other words that actions are represented in memory *associatively*, as properties of objects—in Gibson's 1966 terms, as *affordances* of objects.

- The second observation suggests that in a cruel and nondeterministic world it is better to identify reasonably highly valued states that you have a reasonable chance of getting to than to optimize complete plans.

What does it Take to Plan?

- The problem of planning can therefore be viewed as the problem of Search for a sequence of actions or affordances in a "Kripke model":
- A Kripke model is a tree or more accurately a lattice, in which nodes are states, and arcs are actions.
- A plan is then a sequence of actions that culminates in a state that satisfies the goal of the plan.
- Search for plans is intrinsically recursive, and requires a Push-Down Automaton (PDA) to keep track of alternative paths to some limited depth.
- It is interesting that a PDA is also necessary to process recursive languages.
- But a PDA clearly isn't enough for human language, which animals lack.

Representing Actions

- We can think of actions as STRIPS operators or as finite-state transducers (FSTs) over (sparse) state-space vectors

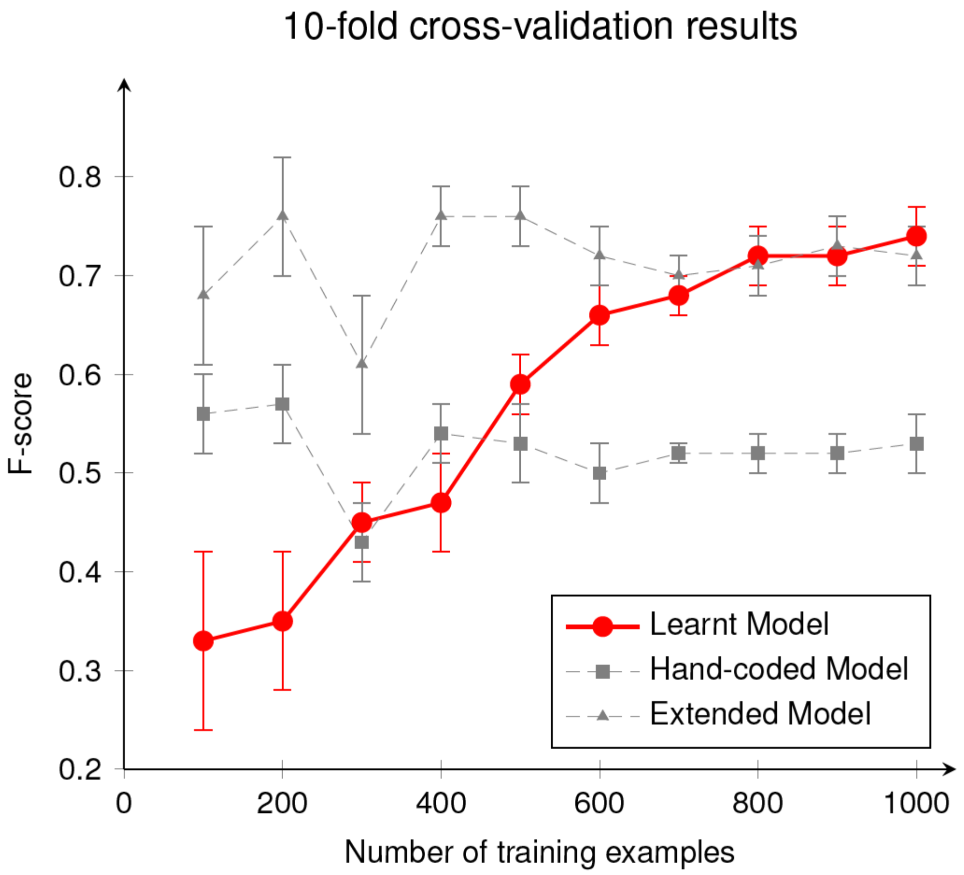
- FSTs are closed under composition, and can be represented as simple neural computational devices such as Perceptrons, or the Associative Network or Willshaw Net (Willshaw 1981 cf. Marr 1969).

☡ We still need a stack memory to run the search for plans.

Reducing Complexity

- We need the Kernel Generalization of Perceptrons to learn STRIPS rules (and their more modern descendants) as FSTs (Mourão *et al.* 2009, 2010).
- This calls for a highly structured state representation (Hume, 1738; Kant, 1781, *passim)*, of a kind that can only be developed by more than 500M years of chordate evolution, using resources on a scale that is completely beyond machine learning.
- Like everyone else, we have to define a state-description language by hand.
- Complexity is $O(n^2)$, so we still need to keep the state vector small.
- We do this via a via a "deictic" or location-based attention mechanism cf. Agre and Chapman (1987) and Pasula *et al.* (2007)

Mourão 2012: Predicting STRIPS Update



II: From Planning to Semantics

- How do we get from seriation and affordance (which we share with other animals) to language (which is uniquely human)?

- *Seriation* of actions to form a plan is Composition of FSTs or functions of type $state \rightarrow state$

- The *Affordance* of a state is a function from all those actions that are possible in that state into their respective result states.

- States are defined by the objects they include, so this is like exchanging objects for Type-Raised functions that map states into other states resulting from actions on those objects.

Actions as Functions

- Thus, the affordance of a (state including a) box to an ape is a function from actions like *the box falling*, *their climbing-on the box* and *their putting the box on another box* into resulting states whose utility the ape can evaluate.
- The functions are of the following (Curried) types, where e is the type of a state satifying preconditions including the presence of an entity, and t is a consequent state:
  - $\mathit{fall}_{e \rightarrow t}$,
  - $\mathit{climb\text{-}on}_{e \rightarrow (e \rightarrow t)}$
  - $\mathit{put\text{-}on}_{e \rightarrow (e \rightarrow (e \rightarrow t))}$

Objects as Affordances

- Thus the ape's concept of a box is an object-oriented set of Type-Raised functions of type
  - $box1_{(e \to t) \to t}$
  - $box2_{(e \to (e \to t)) \to (e \to t)}$
  - $box3_{(e \to (e \to (e \to t))) \to (e \to (e \to t))}$
- —that is, functions from the current situation to the results of the actions it affords.
- Planning is then object-oriented seriation of affordances
- So the only place for human planning to differ from animal planning in a way that supports language is in the event representation itself.

"Grounding" Actions and Affordances

- The fact that actions and objects have (fairly) simple types doesn't mean that the actions themselves are simple.
- A box *falling* is not a volitional action, and has perceptual preconditions like a *looming* flow-field. The event is a complex conjunction of entailments of a box falling, such as a *hurting* event, and the consequent state concerns issues other than the mere lowering of the box's position.
- The ramified nature of this dynamic event knowledge is the reason that languages can vary in the way they carve the conceptual representation at the joints to define their (much terser) lexical semantics.
- E.g. English *run across the road* vs. French *traverser la rue à la course*.
- To understand the connection between planning and semantics, we need to better understand the grounded event representation.

III: The Problem of Content

- Linguists and the Artificial Intelligencia have notably failed to devise a semantics that captures this cross-linguistic variety.

(1) Thomason, 1974: $\forall x[bug'x \Rightarrow \exists y[plants(y) \land kill'y\ x]]$
McCawley, 1968: $[_S CAUSE\ BUGS[_S BECOME[_S NOT[_S ALIVE\ PLANTS]]]]$
Dowty, 1979: $[CAUSE[DO\ BUGS\ \varnothing]][BECOME\neg[ALIVE\ PLANTS]]]$
Talmy, 2000: $Bugs\ ARE\text{-}the\text{-}AUTHOR''\text{-}OF\,[plants\ RESULT\text{-}TO\text{-}die]$
Van Valin, 2005: $[do'(bugs',\varnothing)]CAUSE[BECOME[dead'(plants)]]$
Goddard, 2010: *BUGS* do something to *PLANTS*; because of this, something happens to *PLANTS* at the same time; because of this, something happens to *PLANTS*' body; because of this, after this *PLANTS* are not living anymore.

- Can we identify the primitive concepts automatically, as hidden variables?

Two Approaches

- Clustering by Collocation (Landauer and Dumais, 1997; Baroni and Zamparelli, 2010; Grefenstette and Sadrzadeh, 2011; Padó and Lapata, 2007; Mitchell and Lapata, 2008; Mikolov *et al.*, 2013).
  - Composition via Linear Algebraic Operations
  - Good for underspecification and disambiguation

- Clustering by Denotation (Lin and Pantel, 2001; Hovy *et al.*, 2001), using sentences involving identifiable Named Entities (Lewis and Steedman, 2013a; Reddy *et al.*, 2014)
  - Composition via traditional Logical Operators
  - Good for inference.

Clustered Entailment Semantics

- We must distinguish paraphrase from entailment.

- X_{person} *elected to* Y_{office} entails X_{person} *ran for* Y_{office} but not *vice versa*.

- The paraphrase relation depends on global properties of the named entity relation graph.

- Lewis (2015); Lewis and Steedman (2014b) apply the entailment graphs of Berant *et al.* (2012) to generate more articulated entailment structures.

Local Entailment Probabilities

- The typed named-entity technique is applied to (errorfully) estimate local probabilities of entailments:

a. $p(conquer\,x\,y \Rightarrow invade\,x\,y) = 0.9$

b. $p(invade\,x\,y \Rightarrow attack\,x\,y) = 0.8$

c. $p(conquer\,x\,y \Rightarrow attack\,x\,y) = 0.4$

d. $p(bomb\,x\,y \Rightarrow attack\,x\,y) = 0.7$

(etc.)

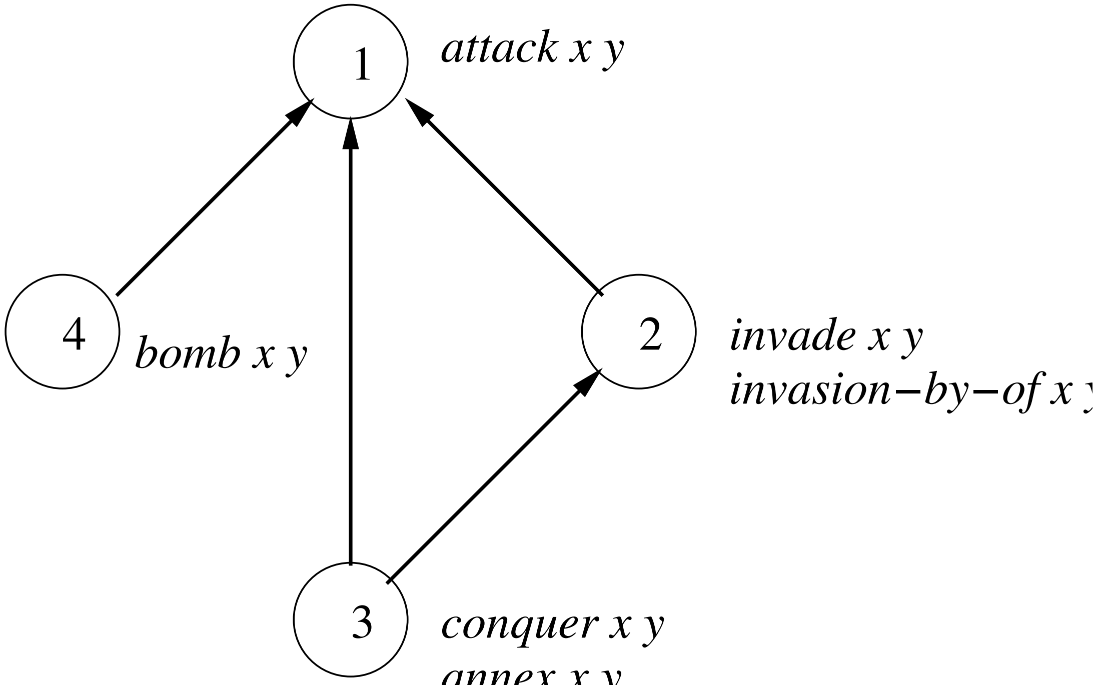
Global Entailments

- The local entailment probabilities are used to construct an entailment graph using integer linear programming with ± weights around $p = 0.5$ with the global constraint that the graph must be closed under transitivity.

- Thus, (c) will be included despite low observed frequency, while other low frequency spurious local entailments will be excluded..

- Cliques within the entailment graphs are collapsed to a single paraphase cluster relation identifier.

Entailment graph



- A simple entailment graph for relations between countries.

Lexicon

- The lexicon obtained from the entailment graph

  attack := $(S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1 xye$
  bomb := $(S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1 xye \wedge rel_4 xye$
  invade := $(S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1 xye \wedge rel_2 xye$
  conquer := $(S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1 xye \wedge rel_2 xye \wedge rel_3 xye$
  annex := $(S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1 xye \wedge rel_2 xye \wedge rel_3 xye$

- These logical forms support correct inference under negation, such as that *conquered* entails *attacked* and *didn't invade* entails *didn't conquer*
- To answer a question "Did X invade Y" we look for sentences which subsume the conjunctive logical form $rel_2 \wedge rel_1$, or satisfy its negation $\neg rel_2 \vee \neg rel_1$.
- Note that if we know that *invasion-of* is a paraphrase of $invade = rel_2$, we also know *invasion-of* entails $attack = rel_1$.

Lexicon

- Primitives like rel_3 correspond to "hidden" semantic primitives that distinguish these concepts.

- If we do the machine-reading cross-linguistically (Lewis and Steedman, 2013b), we will see that some of them correspond to universal elements masked in English (see earlier remarks about *run across the road*)).

- Others will be more arcane.

Results (Lewis and Steedman, 2014b)

System	**Accuracy (all)**	**AUC (all)**
Majority Class	56.8%	0.46
Non Compositional	57.4%	0.48
CCG Baseline	57.8%	0.46
CCG ChineseWhispers	58.0%	0.50
VectorMultiplicative	61.3%	0.51
VectorAdditive	63.5%	0.57
CCG Entailment Graphs	64.0%	0.58
CCG Entailment Graphs+ Implicative Verb Lexicon	**65.0%**	**0.59**

Philosophical Reflections

- Our hidden relations resemble "meaning postulates", such as the one that says that in every model where X is a *bachelor*, X is also *unmarried* and *male*
- Carnap (1952) introduced meaning postulates in support of Inductive Logic, including a model of Probability, basically to keep the model small and consistent.
- This suggests that our semantic representation expresses an a pragmatic empiricist view of "analytic meaning", of the kind advocated by Quine (1951).
- It can also be viewed as a statistical and text-based approach to treating "meaning as use" (Wittgenstein, 1953).

IV: Hanging Language onto Planning

- We saw that (partially) searching the plan graph is an intrinsically recursive process.

- So we need *at least* a push-down automaton (PDA) to keep track of it.

- Is a PDA expressive enough?

- It depends on the class of plans

- If the set of plan- types is unbounded, than a a PDA is not enough.

- (For the same reason that a PDA is not enough for a PS grammar with unboundedly many non-terminals.)

Language and Cooperative Planning

- Collaborative Plans are functions over arbitrary numbers of other agents:

(2) a. Find someone to help to mind the baby
b. Find someone to promise to help to mind the baby
c. Find someone to ask to promise to help to mind the baby. *(etc.)*

◇ Searching a graph with unboundedly many node-types needs an Embedded PDA (EPDA), in which the stack of the PDA can include stack-valued elements.

- Collaborative planning with other minds provides not only the only known motivation for language (Tomasello, 1999), but also the characteristic automaton that supports its use.
- So we should look at the grammar of sentences such as (2).

Combinatory Categorial Grammar

- CCG (Steedman, 2000; Bozşahin, 2012) eschews language-specific syntactic rules like (3) for English.

(3) ~~$S \rightarrow NP\ VP$~~
~~$VP \rightarrow TV\ NP$~~
~~$TV \rightarrow \{proved, found, met, \ldots\}$~~

- Instead, all language-specific syntactic information is *lexicalized*, via lexical entries like (4) for the English transitive verb, where met' is an abbreviation for some conjunction of clustered entailments of the kind discussed earlier.:

(4) $\mathrm{met} := (S\backslash NP)/NP : met'$

- In CCG, syntactic projection from the lexicon is mediated by type-raising **T** and composition **B**.

The Lexicon

- The syntactic "category" identifies the transitive verb as a function, and specifies the type and directionality of its arguments and the type of its result. For Turkish:

(5) $\text{rastladı} := (S\backslash NP)\backslash NP : \mathit{met}'$

◇ This is a good example of the different ways languages carve meaning at the joints. *rastladı* means something like "came across", and is distinct from reciprocal "meet" *tanıştı* which is the same word in English.

- A cross-linguistic clustered entailment semantics, obtained from multilingual machine-reading, would split these meanings into two distinct clusters, rather than one met'

Type Raising as Case

- We will assume that type-raising in the form of case is a universal primitive of grammar, as it is for planning in the form of affordance.

- All noun-phrases (NP) like "Harry" are (polymorphically) type-raised.

- In Japanese and Latin this is the job of case morphemes like nominative *-ga* and *-us*. (Same for Turkish, except nominative is null.)

- In English NPs are ambiguous as to case, and must be disambiguated by the parsing model (a.k.a. "structural case").

Syntactic Derivation

- (6)

$$\begin{array}{clccl}
\text{Harry} & & \text{met} & \text{Sally} & \\
\cline{1-1}\cline{3-3}\cline{4-4}
S/(S\backslash NP) & >\mathbf{T} & (S\backslash NP)/NP & (S\backslash NP)\backslash((S\backslash NP)/NP) & <\mathbf{T} \\
\cline{3-4}
 & & \multicolumn{2}{c}{S\backslash NP} & < \\
\cline{1-4}
\multicolumn{4}{c}{S} & >
\end{array}$$

- (7)

$$\begin{array}{clccl}
\text{Harry} & & \text{met} & \text{Sally} & \\
\cline{1-1}\cline{3-3}\cline{4-4}
S/(S\backslash NP) & >\mathbf{T} & (S\backslash NP)/NP & (S\backslash NP)\backslash((S\backslash NP)/NP) & <\mathbf{T} \\
\cline{1-3}
\multicolumn{3}{c}{S/NP} & >\mathbf{B} & \\
\cline{1-4}
\multicolumn{4}{c}{S} & >
\end{array}$$

“Surface Compositional” Semantics

- (8)

$$\begin{array}{lll}
\text{Harry} & \text{met} & \text{Sally}\\ \cline{1-1}\cline{2-2}\cline{3-3}
S/(S\backslash NP)\ {>}\mathbf{T} & (S\backslash NP)/NP & (S\backslash NP)\backslash((S\backslash NP)/NP)\ {<}\mathbf{T}\\
:\lambda p.p\ harry' & :met' & :\lambda p.p\ sally'\\ \cline{2-3}
 & \multicolumn{2}{c}{S\backslash NP : met'\,sally' \quad <}\\ \cline{1-3}
\multicolumn{3}{c}{S : met'\,sally'\,harry' \quad >}
\end{array}$$

- (9)

$$\begin{array}{lll}
\text{Harry} & \text{met} & \text{Sally}\\ \cline{1-1}\cline{2-2}\cline{3-3}
S/(S\backslash NP)\ {>}\mathbf{T} & (S\backslash NP)/NP & (S\backslash NP)\backslash((S\backslash NP)/NP)\ {<}\mathbf{T}\\
:\lambda p.p\ harry' & :met' & :\lambda p.p\ sally'\\ \cline{1-2}
\multicolumn{2}{l}{S\backslash NP : \lambda x.met'\,x\,harry' \quad {>}\mathbf{B}} & \\ \cline{1-3}
\multicolumn{3}{c}{S : met'\,sally'\,harry' \quad >}
\end{array}$$

Relativization

- (10) (The woman)

$$\begin{array}{ccc}
\text{that} & \text{Harry} & \text{met} \\
\cline{1-1}\cline{2-2}\cline{3-3}
 & {\scriptstyle >\mathbf{T}} & \\
(N\backslash N)/(S/NP) & S/(S\backslash NP) & (S\backslash NP)/NP \\
\lambda p\lambda n\lambda x.px\wedge nx & \lambda p.p\,harry' & :met' \\
\cline{2-3}
 & \multicolumn{2}{r}{{\scriptstyle >\mathbf{B}}} \\
 & \multicolumn{2}{c}{S/NP} \\
 & \multicolumn{2}{c}{\lambda x.met'\,x\,harry'} \\
\cline{1-3}
\multicolumn{3}{r}{{\scriptstyle >}} \\
\multicolumn{3}{c}{N\backslash N : \lambda n\lambda x.met'\,x\,harry'\wedge nx}
\end{array}$$

(11) (The woman)

$$\begin{array}{ccccc}
\text{that} & \text{Harry} & \text{says} & \text{he} & \text{met} \\
\cline{1-1}\cline{2-2}\cline{3-3}\cline{4-4}\cline{5-5}
 & {\scriptstyle >\mathbf{T}} & & {\scriptstyle >\mathbf{T}} & \\
(N\backslash N)/(S/NP) & S/(S\backslash NP) & (S\backslash NP)/S & S/(S\backslash NP) & (S\backslash NP)/NP \\
\cline{2-3}\cline{4-5}
 & \multicolumn{2}{r}{{\scriptstyle >\mathbf{B}}} & \multicolumn{2}{r}{{\scriptstyle >\mathbf{B}}} \\
 & \multicolumn{2}{c}{S/S} & \multicolumn{2}{c}{S/S} \\
\cline{2-5}
 & \multicolumn{4}{r}{{\scriptstyle >\mathbf{B}}} \\
 & \multicolumn{4}{c}{S/NP} \\
\cline{1-5}
\multicolumn{5}{r}{{\scriptstyle >}} \\
\multicolumn{5}{c}{N\backslash N}
\end{array}$$

Coordination

- (12)

$$
\begin{array}{cccccc}
\text{give} & \text{Harry} & \text{a book} & \text{and} & \text{Sally} & \text{a record} \\ \cline{1-6}
 & {}^{<\mathbf{T}} & {}^{<\mathbf{T}} & & {}^{<\mathbf{T}} & {}^{<\mathbf{T}} \\
DTV & TV\backslash DTV & VP\backslash TV & (X\backslash X)/X & TV\backslash DTV & VP\backslash TV \\ \cline{2-3}\cline{5-6}
 & \multicolumn{2}{r}{<\mathbf{B}} & & \multicolumn{2}{r}{<\mathbf{B}} \\
 & \multicolumn{2}{c}{VP\backslash DTV} & & \multicolumn{2}{c}{VP\backslash DTV} \\ \cline{4-6}
 & & & \multicolumn{3}{r}{>} \\
 & & & \multicolumn{3}{c}{(VP\backslash DTV)\backslash(VP\backslash DTV)} \\ \cline{2-6}
 & \multicolumn{5}{r}{<} \\
 & \multicolumn{5}{c}{VP\backslash DTV} \\ \cline{1-6}
\multicolumn{6}{r}{<} \\
\multicolumn{6}{c}{VP}
\end{array}
$$

CCG reduces the linguists' MOVE and COPY/DELETE to adjacent MERGE

CCG is "Near Context-Free"

- The composition rules in CCG are generalized to $\mathbf{B}^2$, and "crossed composition" $\mathbf{B}_\times$
- The combination of type-raising and generalized composition yields a permuting and rebracketing calculus closely tuned to the needs of natural grammar.
- CCG and TAG are provably weakly equivalent to Linear Indexed Grammar (LIG) Vijay-Shanker and Weir (1994).
- Hence they are not merely "Mildly Context Sensitive" (Joshi 1988), but rather "Near Context Free," or "Type $1.\dot{9}$" in the Extended Chomsky Hierarchy.

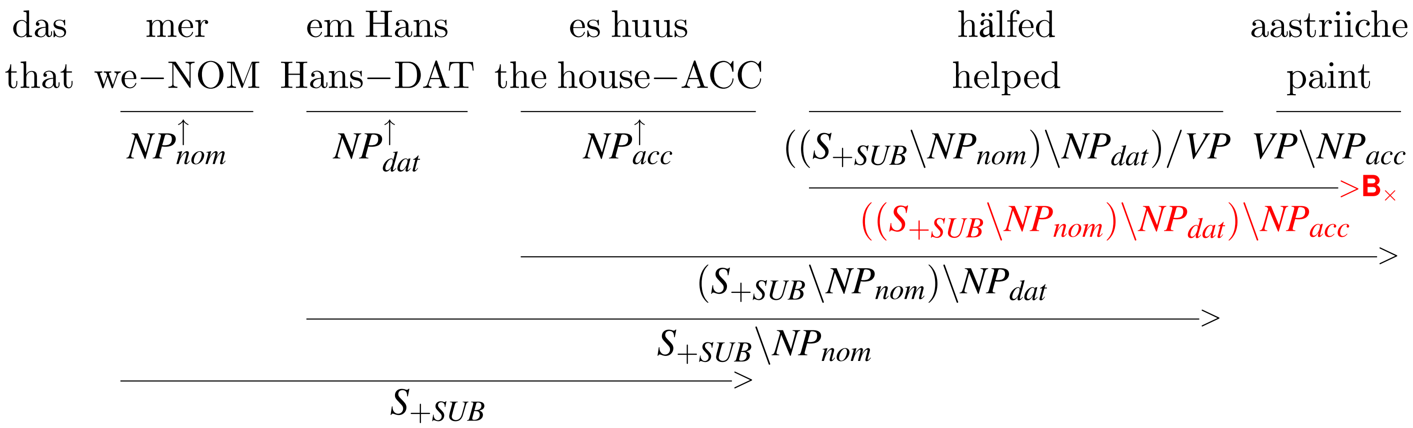
The Extended Chomsky Hierarchy

Language Type	Automaton	Rule-types	Exemplar
Type 0: RE	Universal Turing Machine	$\alpha \rightarrow \beta$	
Type 1: CS	Linear Bound Automaton (LBA)	$\phi A \psi \rightarrow \phi \alpha \psi$	
I	Nested Stack Automaton(NSA)	$A_{[(i),...]} \rightarrow \phi B_{[(i),...]} \psi C_{[(i),...]} \xi$	a^{2^n}
LCFRS/MCF	ith-order EPDA	$A_{[[(i),...]...]} \rightarrow \phi B_{[[(i),...]...]} \psi$	$\mathscr{P}(a^n b^n c^n)$
LI	Embedded PDA (EPDA)	$A_{[(i),...]} \rightarrow \phi B_{[(i),...]} \psi$	$a^n b^n c^n$
Type 2: CF	Push-Down Automaton (PDA)	$A \rightarrow \alpha$	$a^n b^n$
Type 3: FS	Finite-state Automaton (FSA)	$A \rightarrow \begin{cases} \text{a B} \\ \text{a} \end{cases}$	a^n

- All higher language classes properly contain all lower except LCFRS and I, which properly intersect.

Zürich German is Strongly Near Context-Free

(13)



das mer em Hans es huus hälfed aastriiche

that we–NOM Hans–DAT the house–ACC helped paint

$NP^{\uparrow}_{nom}$ $NP^{\uparrow}_{dat}$ $NP^{\uparrow}_{acc}$ $((S_{+SUB}\backslash NP_{nom})\backslash NP_{dat})/VP$ $VP\backslash NP_{acc}$

$\xrightarrow{} >\mathbf{B}_{\times}$

$((S_{+SUB}\backslash NP_{nom})\backslash NP_{dat})\backslash NP_{acc}$

$\xrightarrow{} >$

$(S_{+SUB}\backslash NP_{nom})\backslash NP_{dat}$

$\xrightarrow{} >$

$S_{+SUB}\backslash NP_{nom}$

$\xrightarrow{} >$

S_{+SUB}

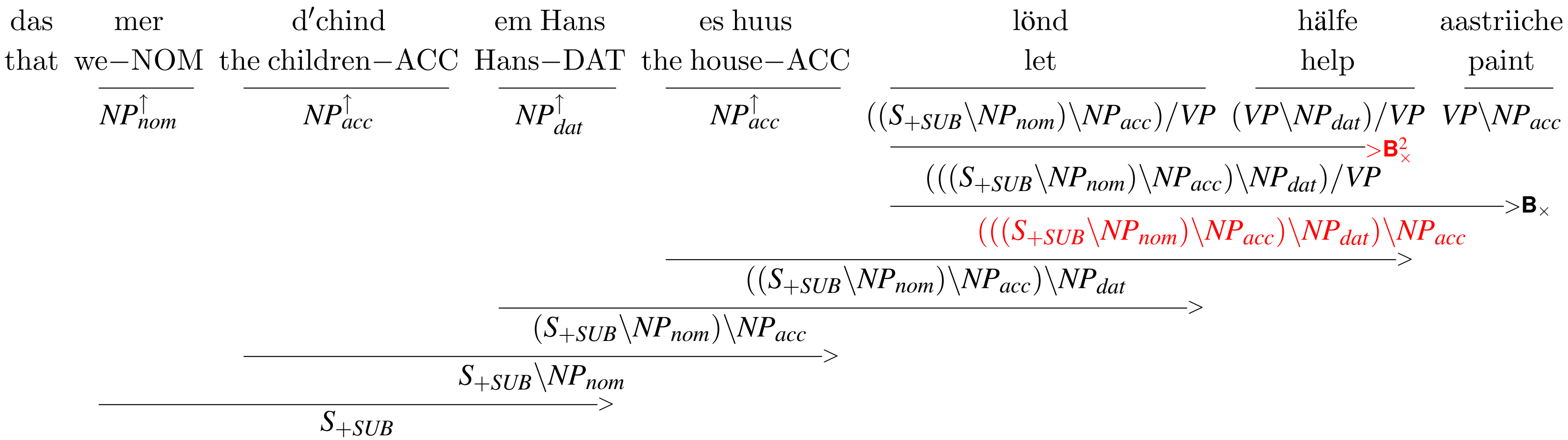
“that we helped Hans paint the house”

- The following is correctly also allowed (Shieber, 1985):

(14) Das mer em Hans hälfed es huus aastriiche.

Zürich German is Strongly Near Context-Free

(15)



das	mer	d′chind	em Hans	es huus	lönd	hälfe	aastriiche
that	we−NOM	the children−ACC	Hans−DAT	the house−ACC	let	help	paint
	$NP^{\uparrow}_{nom}$	$NP^{\uparrow}_{acc}$	$NP^{\uparrow}_{dat}$	$NP^{\uparrow}_{acc}$	$((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})/VP$	$(VP\backslash NP_{dat})/VP$	$VP\backslash NP_{acc}$

$>\mathbf{B}^2_{\times}$: $(((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})/VP$

$>\mathbf{B}_{\times}$: $(((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat})\backslash NP_{acc}$

$>$: $((S_{+SUB}\backslash NP_{nom})\backslash NP_{acc})\backslash NP_{dat}$

$>$: $(S_{+SUB}\backslash NP_{nom})\backslash NP_{acc}$

$>$: $S_{+SUB}\backslash NP_{nom}$

$>$: S_{+SUB}

“that we let the children help Hans paint the house”

- Again, other word orders are correctly allowed.
- Constituents like “es huus lönd hälfe aastriiche” are homologous to collaborative plans like earlier “Find someone to let someone help someone mind the baby”.

Conclusion (I)

- The lexicon is the only locus of language specific information in the grammar.
- The universal projective syntactic component of natural language grammar is based on the combinators **B**, **T**.
- In evolutionary terms, these combinators were provided ready-made, by a sensory motor planning mechanism most of which we share with a number of animals.
- The problem of parsing is automata-theoretically equivalent to the problem of planning for cooperation with other minds.
- Both of the latter abilities seem unique to humans.

Conclusion (II)

- The following progression over the last 200M years of vertebrate evolution may have resulted in an essentially instantaneous recent emergence of language:

  1. Pure reactive planning with non-recursive KR (finite-state);
  2. (Forward-chaining, breadth-first) deliberative planning with non-recursive KR requiring composition, type-raising, and a (simulated) PDA for search;

- A PDA also supports recursive concepts in KR. But a PDA alone isn't enough to support human planning and humean language, which other animals lack.

Conclusion (III)

- We must postulate the following further developments:

3. Human planning is characterized by the use of plan variables corresponding to unknown provided by external agencies such as phone-books, Google search, or other human beings.
   - Planning with the particular recursive concepts that are necessary human collaboration for purposes like neotenic child-reading generates plans with unboundedly many plan variables (agents) (Hrdy, 2009; Steedman, 2014).
4. Such planning requires a (simulated) embedded PDA
5. The EPDA immediately supports near-context-free Natural Language Grammar, as attested by English, Turkish, and Zürich German
   - This can happen without any further evolutionary work other than a little specialization of the vocal tract.

Appendix: Practical Applications of CCG

- It was widely expected in the ’80s that the degree of derivational ambiguity CCG allows would make it completely impractical for parsing.
- However, any grammar that covers these data has the same problem.
- The universal recognition in the ’90s of the need for statistical modeling in NLP was a great leveler.
- With such models, CCG can be parsed as fast and as accurately as anything else—
- —with the advantage of a surface compositional semantics including discontinuity and “non-projectivity”.

Practical Applications of CCG

- Many applications exploit the “surface compositional” semantics of CCG—for example:
  - Hockenmaier (2003); Clark and Curran (2004); Çakıcı and Steedman (2009); Lewis and Steedman (2014a) provide publicly available efficient parsers trained on WSJ.
  - Birch *et al.* (2007); Hassan *et al.* (2009); Mehay and Brew (2012) use CCG for statistical machine translation
  - Prevost (1995); White (2006) apply it to sentence realization
  - Briscoe (2000); Kwiatkowski *et al.* (2012); Krishnamurthy and Mitchell (2012) apply it to semantic parsing and language acquisition
  - Bos and Markert (2005); Lewis and Steedman (2013a,b, 2014b) apply it to open-domain question answering and entailment

References

Agre, Phillip and Chapman, David, 1987. “Pengi: An Implementation of a Theory of Activity.” In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*. Los Altos, CA: Morgan Kaufmann.

Baroni, Marco and Zamparelli, Roberto, 2010. “Nouns are Vectors, Adjectives are Matrices: Representing Adjective-Noun Constructions in Semantic Space.” In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Cambridge, MA: ACL, 1183–1193.

Berant, Jonathan, Dagan, Ido, Adler, Meni, and Goldberger, Jacob, 2012. “Efficient Tree-Based Approximation for Entailment Graph Learning.” In

Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Association for Computational Linguistics, 117–125.

Birch, Alexandra, Osborne, Miles, and Koehn, Philipp, 2007. "CCG SuperTags in Factored Translation Models." In *Proceedings of the 2nd Workshop on Statistical Machine Translation*. held in conjunction with ACL, Prague: ACL, 9–16.

Bos, Johan and Markert, Katja, 2005. "Recognising Textual Entailment with Logical Inference." In *Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing (EMNLP 2005)*. 628–635.

Bozşahin, Cem, 2012. *Combinatory Linguistics*. Berlin: de Gruyter.

Briscoe, Ted, 2000. "Grammatical Acquisition: Inductive Bias and Coevolution of Language and the Language Acquisition Device." *Language* 76:245–296.

Carnap, Rudolf, 1952. "Meaning Postulates." *Philosophical Studies* 3:65–73. reprinted as Carnap, 1956:222-229.

Carnap, Rudolf (ed.), 1956. *Meaning and Necessity*. Chicago: University of Chicago Press, second edition.

Çakıcı, Ruket and Steedman, Mark, 2009. "A Wide Coverage Morphemic Lexicon for Turkish." In *Proceedings of the Workshop on Parsing with Categorial Grammars, ESSLLI-09*. Bordeaux: ESSLLI.

Clark, Stephen and Curran, James R., 2004. "Parsing the WSJ using CCG and Log-Linear Models." In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*. Barcelona, Spain: ACL, 104–111.

Dowty, David, 1979. *Word Meaning in Montague Grammar*. Dordrecht: Reidel, first edition.

Gibson, James, 1966. *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton-Mifflin Co.

Goddard, Cliff, 2010. "The Natural Semantic Metalanguage Approach." In Bernd Heine and Heiko Narro (eds.), *The Oxford Handbook of Linguistic Analysis*, Oxford: Oxford University Press. 459–484.

Grefenstette, Edward and Sadrzadeh, Mehrnoosh, 2011. "Experimental Support for a Categorical Compositional Distributional Model of Meaning." In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh: ACL, 1394–1404.

Hassan, Hany, Sima'an, Khalil, and Way, Andy, 2009. "A Syntactified Direct

Translation Model with Linear-time Decoding.” In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. Singapore: ACL, 1182–1191.

Hockenmaier, Julia, 2003. “Parsing with generative models of predicate-argument structure.” In *Proceedings of the 41st Meeting of the Association for Computational Linguistics, Sapporo*. San Francisco: Morgan-Kaufmann, 359–366.

Hovy, Eduard, Gerber, Laurie, Hermjakob, Ulf, Junk, Michael, and Lin, Chin-Yew, 2001. “Question Answering in Webclopedia.” In *Proceedings of the Ninth Text Retrieval Conference (TREC-9)*. Washington, DC: NIST, 655–664.

Hrdy, Sarah Blaffer, 2009. *Mothers and Others*. Cambridge, MA: Belnap/Harvard University Press.

Hume, David, 1738. *Treatise of Human Nature*. Oxford: Clarendon Press. ed. Selby-Bigge, 1902.

Joshi, Aravind, 1988. “Tree-Adjoining Grammars.” In David Dowty, Lauri Karttunen, and Arnold Zwicky (eds.), *Natural Language Parsing*, Cambridge: Cambridge University Press. 206–250.

Kant, Immmanuel, 1781. *Kritik der reinen Vernunft (Critique of Pure Reason)*. Riga: J.F.Hartknoch.

Köhler, Wolfgang, 1925. *The Mentality of Apes*. New York: Harcourt Brace and World.

Krishnamurthy, Jayant and Mitchell, Tom, 2012. “Weakly Supervised Training of Semantic Parsers.” In *Proceedings of Joint Conference on Empirical Methods in*

Natural Language Processing and Computational Natural Language Learning. Jeju Island, Korea: ACL, 754–765.

Kwiatkowski, Tom, Goldwater, Sharon, Zettlemoyer, Luke, and Steedman, Mark, 2012. "A Probabilistic Model of Syntactic and Semantic Acquisition from Child-Directed Utterances and their Meanings." In *Proceedings of the 13th Conference of the European Chapter of the ACL (EACL 2012)*. Avignon: ACL, 234–244.

Landauer, Thomas and Dumais, S., 1997. "A Solution to Plato's Problem: The Latent Semantic Analysis of the Acquisition, Induction and Representation of Knowledge." *Psychological Review* 104:211–240.

Lashley, Karl, 1951. "The Problem of Serial Order in Behavior." In L.A. Jeffress (ed.), *Cerebral Mechanisms in Behavior*, New York: Wiley. 112–136. reprinted in Saporta (1961).

Lewis, Mike, 2015. *Natural Semantics for Wide Coverage CCG Parsers*. Ph.D. thesis, University of Edinburgh.

Lewis, Mike and Steedman, Mark, 2013a. “Combined Distributional and Logical Semantics.” *Transactions of the Association for Computational Linguistics* 1:179–192.

Lewis, Mike and Steedman, Mark, 2013b. “Unsupervised Induction of Cross-Lingual Semantic Relations.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 681–692.

Lewis, Mike and Steedman, Mark, 2014a. “A* CCG Parsing with a Supertag-factored Model.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: ACL, 990–1000.

Lewis, Mike and Steedman, Mark, 2014b. “Combining Formal and Distributional

Models of Temporal and Intensional Semantics." In *Proceedings of the ACL Workshop on Semantic Parsing*. Baltimore, MD: ACL, 28–32. Google Exceptional Submission Award.

Lin, Dekang and Pantel, Patrick, 2001. "DIRT—Discovery of Inference Rules from Text." In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data-Mining (KDD-01)*. San Francisco, 323 – 328.

Marr, David, 1969. "A Theory of Cerebellar Cortex." *Journal of Physiology* 202:437–470. Reprinted in Vaina 1991.

McCawley, James, 1968. "Lexical Insertion in a Transformational Grammar without Deep Structure." In *Papers from the 4th Regional Meeting of the Chicago Linguistics Society*. CLS, 71–80.

Mehay, Denis and Brew, Chris, 2012. "CCG Syntactic Reordering Models for

Phrase-based Machine Translation." In *Proceedings of the Seventh Workshop on Statistical Machine Translation*. Montréal: ACL, XXX–XXX.

Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Greg, and Dean, Jeff, 2013. "Distributed Representations of Words and Phrases and their Compositionality." In *Advances in Neural Information Processing Systems*. 3111–3119.

Miller, George, Galanter, Eugene, and Pribram, Karl, 1960. *Plans and the Structure of Behavior*. New York: Holt.

Mitchell, Jeff and Lapata, Mirella, 2008. "Vector-based Models of Semantic Composition." In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. Columbus, OH: ACL, 236–244.

Mourão, Kira, 2012. *Learning Action Representations using Kernel Perceptrons*. Ph.D. thesis, University of Edinburgh.

Mourão, Kira, Petrick, Ron, and Steedman, Mark, 2009. “Learning Action Effects in Partially Observable Domains I.” In *Proceedings of the ICAPS 2009 Workshop on Planning and Learning*. Thessaloniki, Greece, 15–22.

Mourão, Kira, Petrick, Ron, and Steedman, Mark, 2010. “Learning Action Effects in Partially Observable Domains II.” In *Proceedings of the 19th European Conference on AI*. Lisbon, 973–974.

Padó, Sebastian and Lapata, Mirella, 2007. “Dependency-Based Construction of Semantic Space Models.” *Computational Linguistics* 33:161–199.

Pasula, Hanna, Zettlemoyer, Luke, and Kaelbling, Leslie, 2007. “Learning Symbolic Models of Stochastic Domains.” *Journal of AI Research* 29:309–352.

Piaget, Jean, 1936. *La naissance de l’intelligence chez l’enfant*. Paris: Delachaux et Niestle. translated 1953 as *The Origin of Intelligence in the Child*, Routledge and Kegan Paul.

Prevost, Scott, 1995. *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*. Ph.D. thesis, University of Pennsylvania.

Quine, Willard Van Orman, 1951. “Two dogmas of empiricism.” *The Philosophical Review* :20–43reprinted in Quine (1953).

Quine, Willard Van Orman, 1953. *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Reddy, Siva, Lapata, Mirella, and Steedman, Mark, 2014. “Large-scale Semantic Parsing without Question-Answer Pairs.” *Transactions of the Association for Computational Linguistics* 2:377–392.

Saporta, Sol (ed.), 1961. *Psycholinguistics: A Book of Readings*. New York: Holt Rinehart & Winston.

Shieber, Stuart, 1985. “Evidence against the Context-Freeness of Natural Language.” *Linguistics and Philosophy* 8:333–343.

Steedman, Mark, 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Steedman, Mark, 2002a. “Formalizing Affordance.” In *Proceedings of the 24th Annual Meeting of the Cognitive Science Society, Fairfax VA, August*. Mahwah, NJ: Erlbaum, 834–839.

Steedman, Mark, 2002b. “Plans, Affordances, and Combinatory Grammar.” *Linguistics and Philosophy* 25:723–753.

Steedman, Mark, 2014. "Evolutionary Basis for Human Language: Comment on "Toward a Computational Framework for Cognitive Biology: Unifying Approaches from Cognitive Neuroscience and Comparative Cognition" by Tecumseh Fitch." *Physics of Life Reviews* 11:382–388.

Talmy, Leonard, 2000. *Towards a Cognitive Semantics*, volume 1 and 2. Cambridge, MA: MIT Press.

Thomason, Richmond (ed.), 1974. *Formal Philosophy: Papers of Richard Montague*. New Haven, CT: Yale University Press.

Tomasello, Michael, 1999. *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Vaina, Lucia (ed.), 1991. *From Retina to Neocortex: Selected Papers of David Marr*. Boston, MA: Birkhauser.

Van Valin, Robert (ed.), 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Vijay-Shanker, K. and Weir, David, 1994. “The Equivalence of Four Extensions of Context-Free Grammar.” *Mathematical Systems Theory* 27:511–546.

White, Michael, 2006. “Efficient Realization of Coordinate Structures in Combinatory Categorial Grammar.” *Research on Language and Computation* 4:39–75.

Willshaw, David, 1981. “Holography, Association and Induction.” In Geoffrey Hinton and James Anderson (eds.), *Parallel Models of Associative Memory*, Hillsdale, NJ: Erlbaum. 83–104.

Wittgenstein, Ludwig, 1953. *Philosophical Investigations*. London: Basil Blackwell.