Humanoids Capable of Merging

Katerina Pastra
Cognitive Systems Research Institute
http://www.csri.gr

VVV18, 9 February 2018

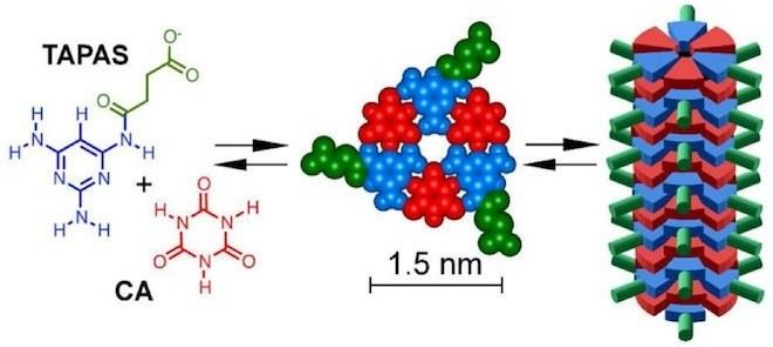
Talk Structure

- Merging: a fundamental cognitive mechanism
- Merging in Language
  - Traditional Language Grammars/Syntax
  - Examples
- Merging in Action
  - The quest for Action grammar
  - The Minimalist Grammar of Action
- Merging in Semantic Association
  - State of the art Semantic Networks
  - The PRAXICON recursive semantic network
- Humanoids with merging abilities:
  - (a) From Visual Understanding to Verbalisation
  - (b) From Verbal Requests to Action

The mechanism of Merging

"… combine two elements to create a new synthetic one…"

"… a structure- building operation…"



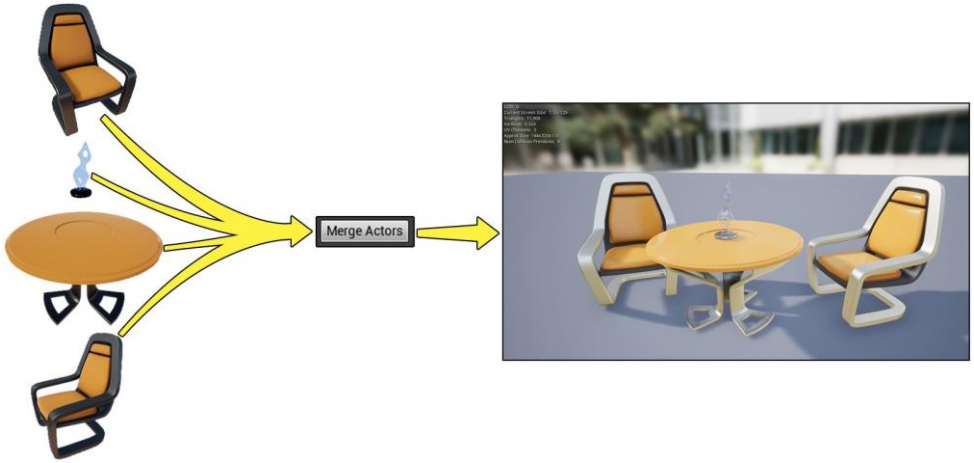
RNA lite? Chemicals known as TAPAS and CA (left) assemble together forming rosettes (middle) that then stack into genelike chains (right).
Credit: B.J. Cafferty et al., JACS (2013)

CLOSE ✕

"Step one: take a good stiff drink."

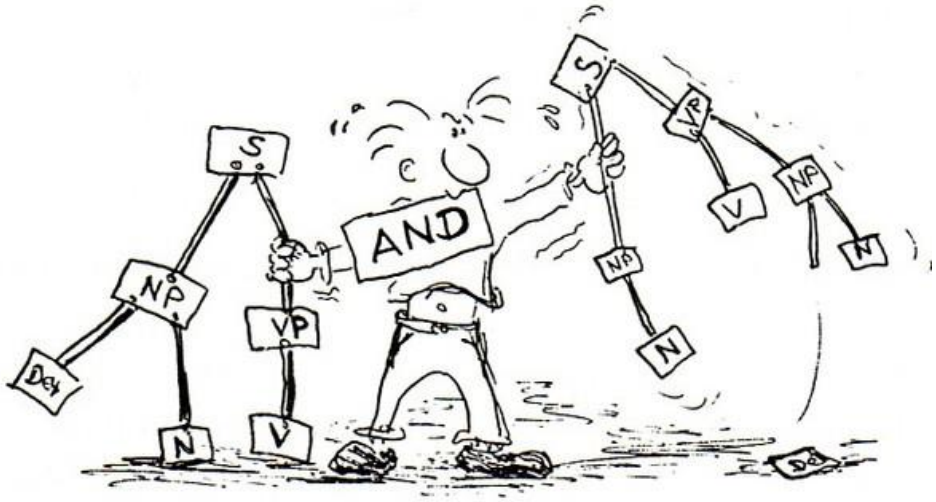
What is being merged?

(Hint: type and complexity of elements)



What is being merged?

(Hint: type and complexity of elements)



Do you see any merging in this picture?

What is being merged?

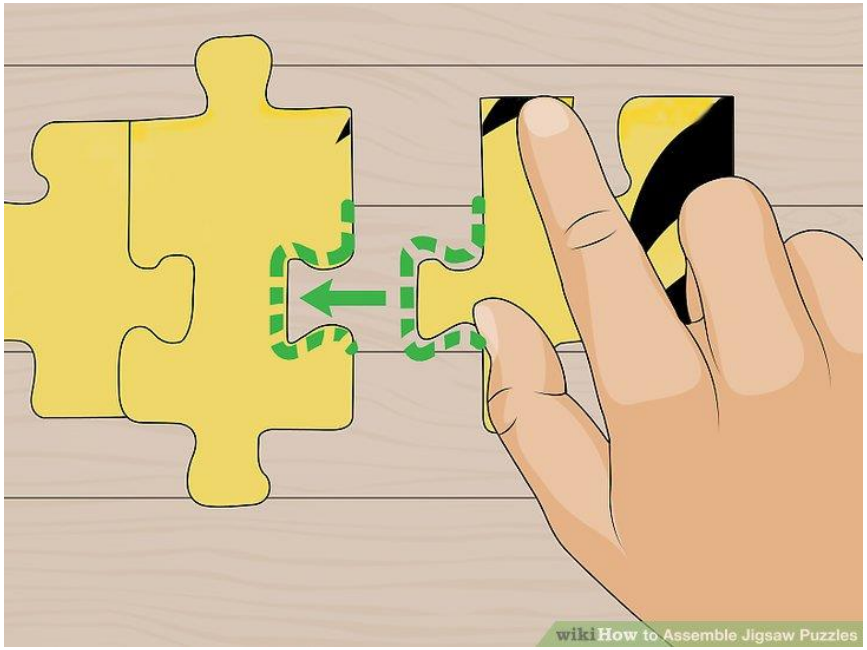
(Hint: type and complexity of elements)

http://www.freesfx.co.uk/rx2/mp3s/6/18243_1464376452.mp3

http://www.freesfx.co.uk/rx2/mp3s/3/14742_1460058298.mp3

Is Merging regulated?

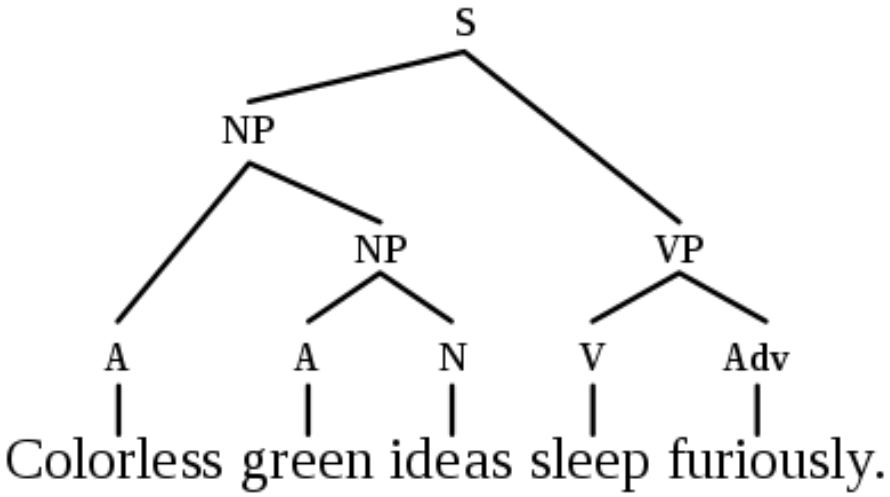
(Hint: enablement)



- Merging is driven by element feature(s)
- Enablement has degrees and so does merging; tight enablement --> dependency leads to tight merging, to tight structures

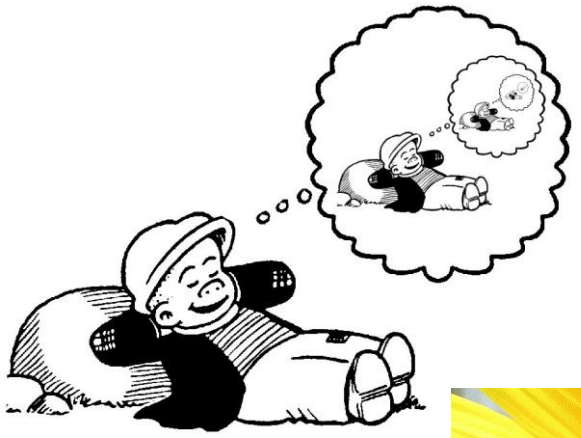
Is Merging Regulated?

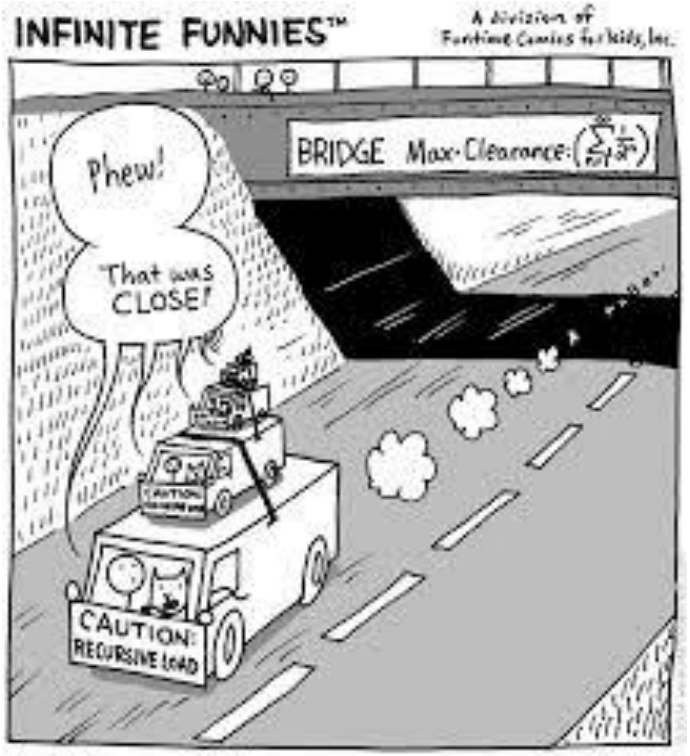
Improbable structures

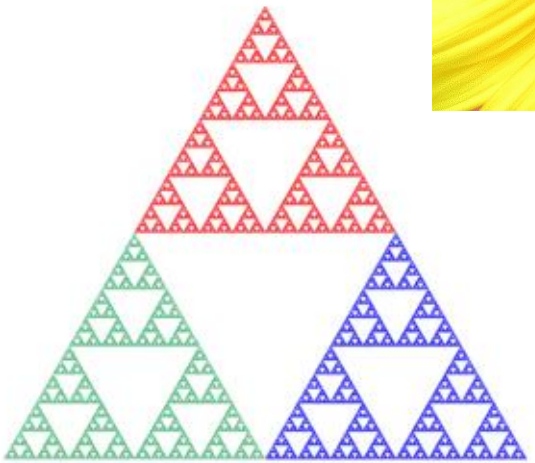


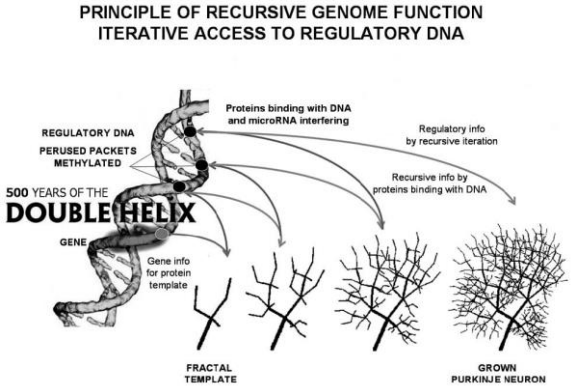
https://www.youtube.com/watch?v=NJKiqdFFv6I

Recursion in Merging









Recursion in Language

Tail Recursion:

"*the man who knows your sister who works at the bookshop*"

Nested (or true) recursion:

"*The cat the boy saw left*"

Recursion in Language (2)

Has been claimed to be:
What makes Language unique in Human Species

(References to the corresponding 'Language Evolution' Debate):

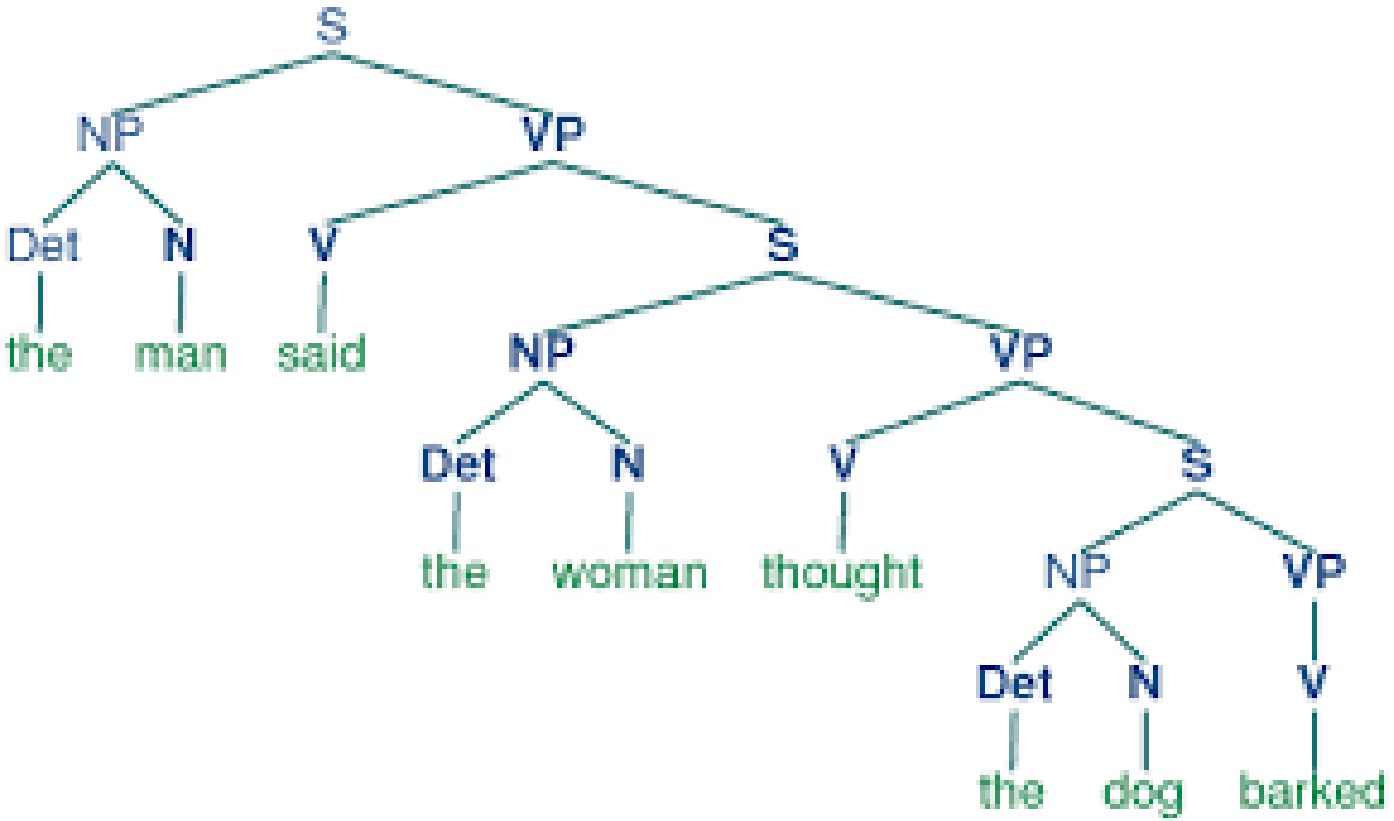
Pinker S, Jackendoff R. The Faculty of Language: What's special about it? Cognition. 2005;95:201–236.

Hauser M, Chomsky N, Fitch W. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? Science. 2002;298:1569–1579.

Fitch W, Hauser M, Chomsky N. The evolution of the language faculty: Clarifications and Implications. Cognition. 2005;97:179–210.

Recursion in Language (3)

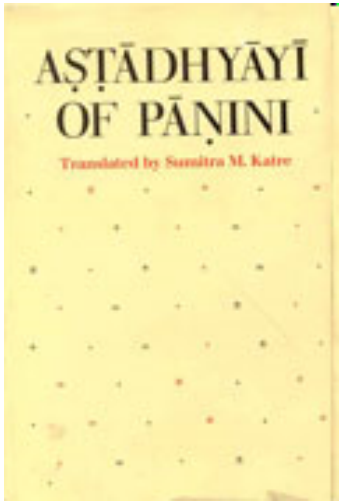
An Important Feature of Generative Grammars
(Chomskyan tradition of verbal syntactic structures)



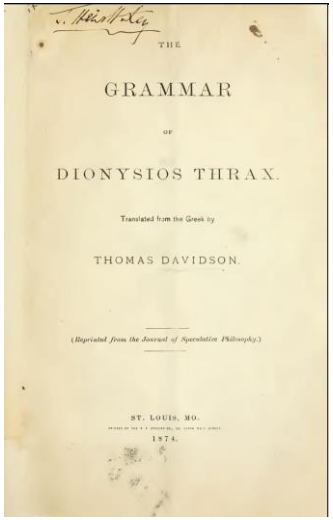
Language Grammars

A grammar is a formal analysis construct that can describe and generate all and only the correct structures in a representation space.

Natural language is the space par excellence for which grammars have been developed (cf. from Panini to Chomsky)



Panini, 4th century BC



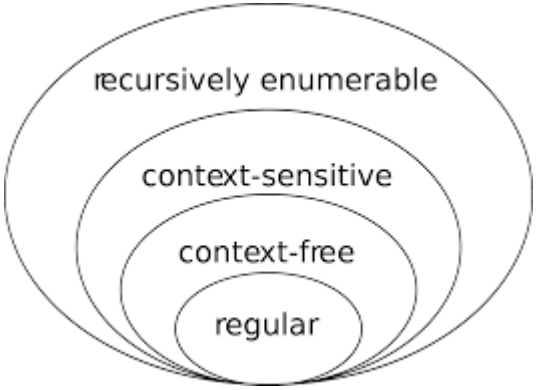
Ars Grammatica – Τέχνη Γραμματική
Dionysios Thrax, 2nd century BC

Basic Grammar Types (crash intro)

Regular grammars

$(AB)^n$

← Like regular expressions
← Implemented through
Finite State Automata
Not recursive!



Context free grammars

$A^n B^n$

← Implemented through

Pushdown Automata
They are recursive
They cannot deal with
discontinuities
and very-long
dependencies
(e.g. Agreement and
reference)

The Quest for Grammars…beyond Language

- **Kirsch, 1964**: suggested a grammar of drawings analogous to text grammar;
- **Gregory 1974**: suggested grammar of vision analogous to language grammar;
- **Lashley 1974**: suggested that syntax may apply not only to language but also to other forms of behaviour, such as goal directed action

……

- **Fadiga 2005, 2009, 2011**: supramodal syntax hypothesis and experimental evidence that Broca's area is the neural locus of (among others) language and action perception and production; suggestions for common syntactic (hierarchical/dependency-based and compositional) processes in language and action

Action Grammar Rationale

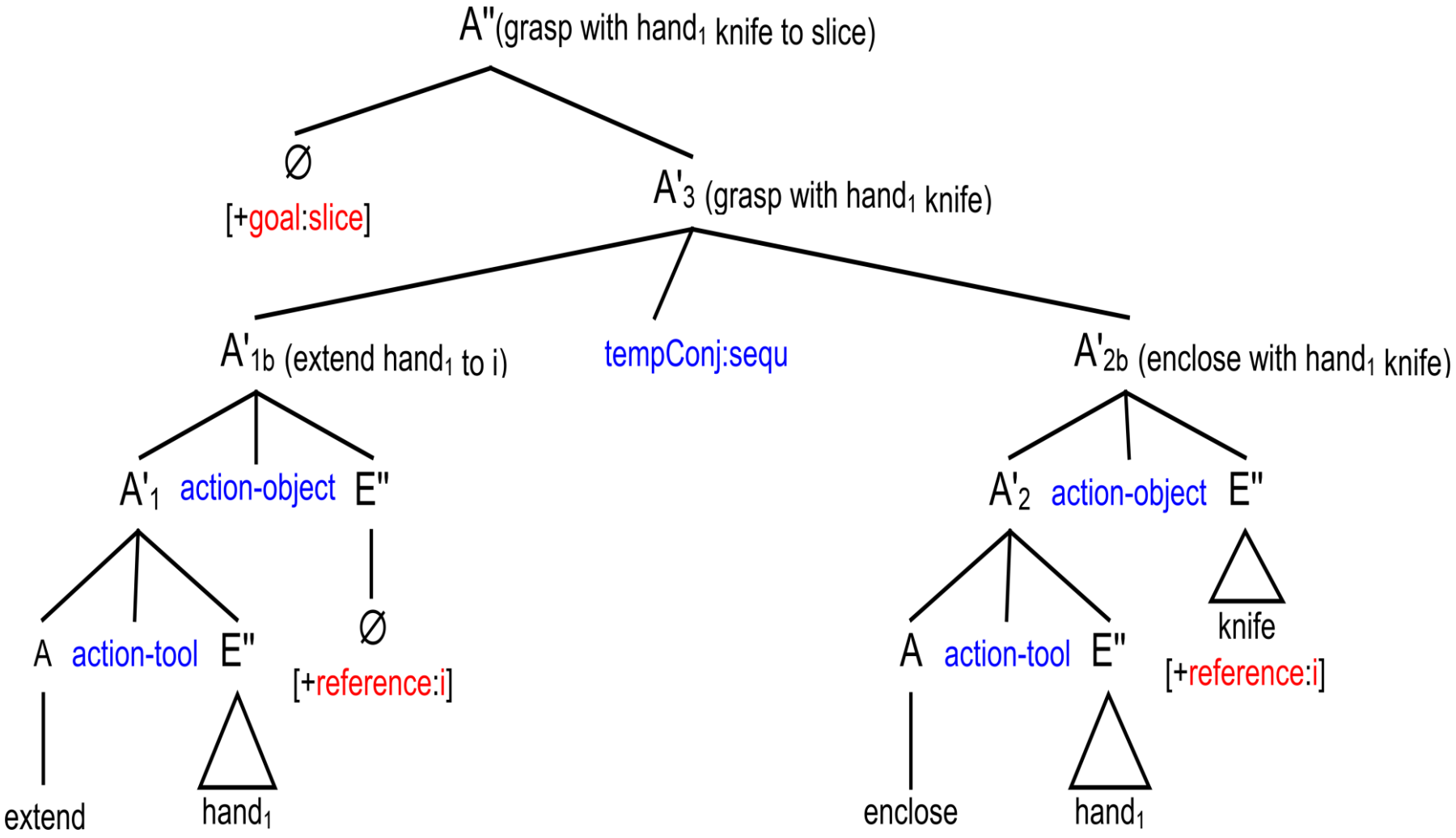
Experimental evidence that *action has a hierarchical and compositional structure*; e.g. 2-year old children have been found to be able not only to parse hierarchically organized actions (Bauer 1995), but also to copy and reproduce such actions (Whiten et al. 2006)

← Could we develop a grammar of action with descriptive and generative power, taking neurophysiological findings into consideration?

→ This would enable one to develop basic cognitive tools for the analysis and generation of human action, and

→ It would provide a way to bridge the gap between action/perception and the language space

The Minimalist Grammar of Action



The Minimalist Grammar of Action

Katerina Pastra and Yiannis Aloimonos (2012), *"The Minimalist Grammar of Action"*, Philosophical Transactions of the Royal Society B, 367(1585):103

The first **generative grammar of action** that employs the structure-building operations and principles of Chomsky's Minimalist Programme as a reference model

The grammar is based on a number of **basic findings in experimental research, and in that sense it has a biological basis**.

Action Grammar Production Rules

Production Rules	
4	A'' → g A'
3	A' → (m) A'
2	A' → A' (o_c)
1	A' → A t_c

Effects/Results → the 'static fingerprints' of actions…

Features/Constituents/Terminals:

- **Tool** Complement (**t_c**)
- **Affected Object** Complement (**o_c**)
- **Physical Space** Modifier (m)
- **Goal** Modifier (g)

Minimalist operators driven by Features:
Merge and **Move**

The operators drive the application of the rules bottom-up

Action Terminals – Non Terminals

Action Grammar Terminals (A): The simplest actions, i.e. perceptible movements carried out by an agent to achieve a goal, which have (one or more) body part tool-complements and no object-complements. Action terminals are further distinguished from each other through their perceptible motor features such as speed, force and direction

Action Grammar Non-Terminals (A′): These are perceptible action phrases, that consist of action terminals (or other non-terminals) in certain temporal configuration; they may have both tool-complements and object complements. They involve interaction with objects beyond one's own body or with other agents, for attaining a particular goal/task

Action Constituents (1)

Tool complement (t_c): the effector of a movement, this being a body part, a combination of body parts or the extension of a body part with a graspable object used as a tool. Syntactic feature. (Related Neuroscience Evidence: Iriki 1996, Fadiga et al. 2000, Mantovani et al. 2011)

Grasping with pliers vs. grasping with tweezers

Object complement (o_c): any object affected by a tool-use action. Syntactic Feature. E.g. Confer Fadiga et al. 2000.

grasping a pencil with the hand vs.
grasping a glass with the hand

Action Constituents (2)

Goal (g): the final purpose of an action sequence of any length or complexity. Inflectional feature!

1. Same movement type, same tool and affected object, but different goal:

grasping a pencil in order to displace it vs. *grasping a pencil in order to write;*

2. Same movement type, different tool or affected object, same goal:

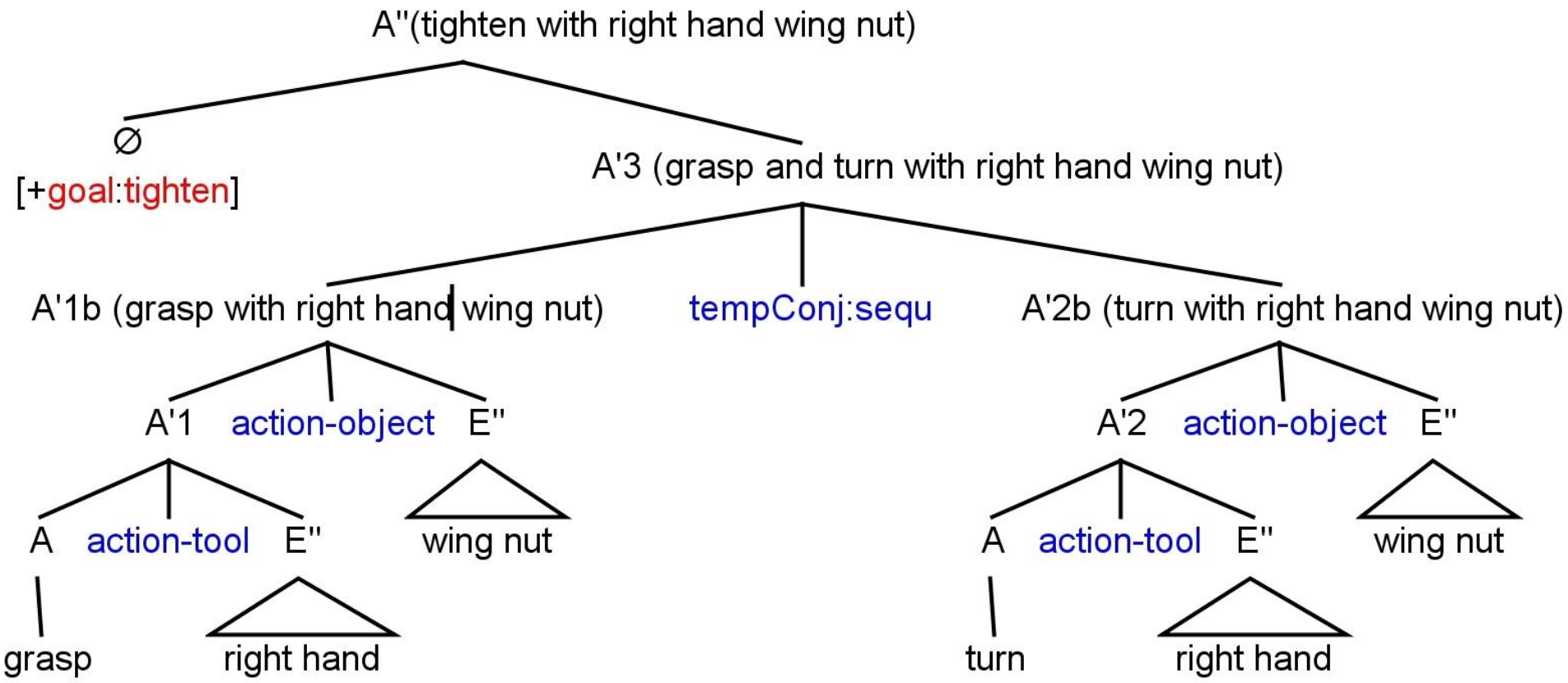
grasping an apple to displace it vs *grasping a cube to displace it*

3. Final goal of an action structure can be predicted from its first subactions

'extend hand towards pencil'

(finger preshaping during hand transfer, cf. Jeannerod et al. 1995), (grasping neurons discharge before contact with object, cf. Fadiga et al. 2000, Fogassi et al. 2005), (Cattaneo et al. 2007)

The Minimalist Grammar of Action



The Minimalist Grammar of Action

M: Movement

T: Tool

O: Object

G: Goal

A: Action

G

Create Bowline Knot

A1

M

Turn clockwise

T

Right Hand

O

Rope Free End

A2

M

Turn clockwise, in/out of Loop 1

T

Right Hand

O

Rope Free End

A3

M

Pass under Standing End, in Loop1

T

Right Hand

O

Rope Free End

Recursion in the Action Space

A)Tail Recursion:

"*extend hand$_1$ – grasp with hand$_1$ knife – cut with knife bread*",

or in language terms:

"*extend hand$_1$, which grasps knife , which cuts bread*".

(B) True Recursion:

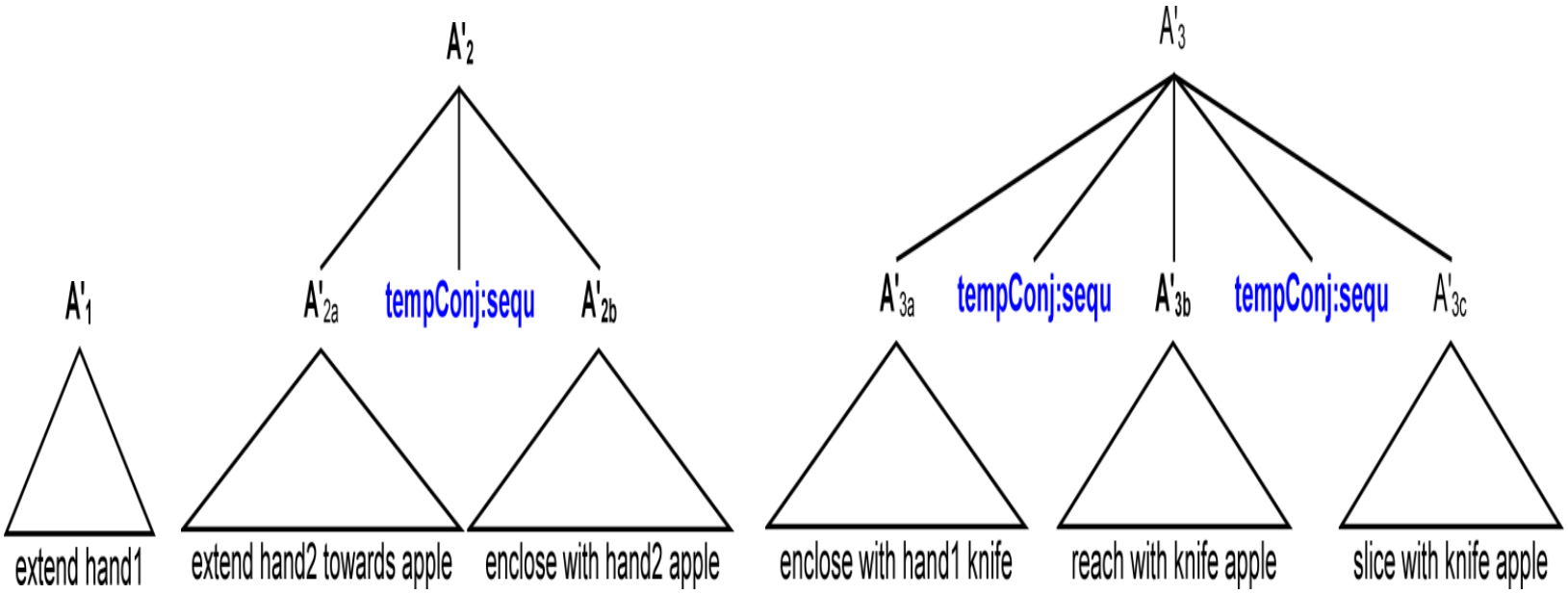
"*extend hand$_1$ – extend hand$_2$, grasp with hand$_2$ ball – grasp with hand$_1$ glass*".

"*grasp with hand$_1$ knife – grasp with hand$_2$ cutting board, press with cutting board cloth - cut with knife bread*",

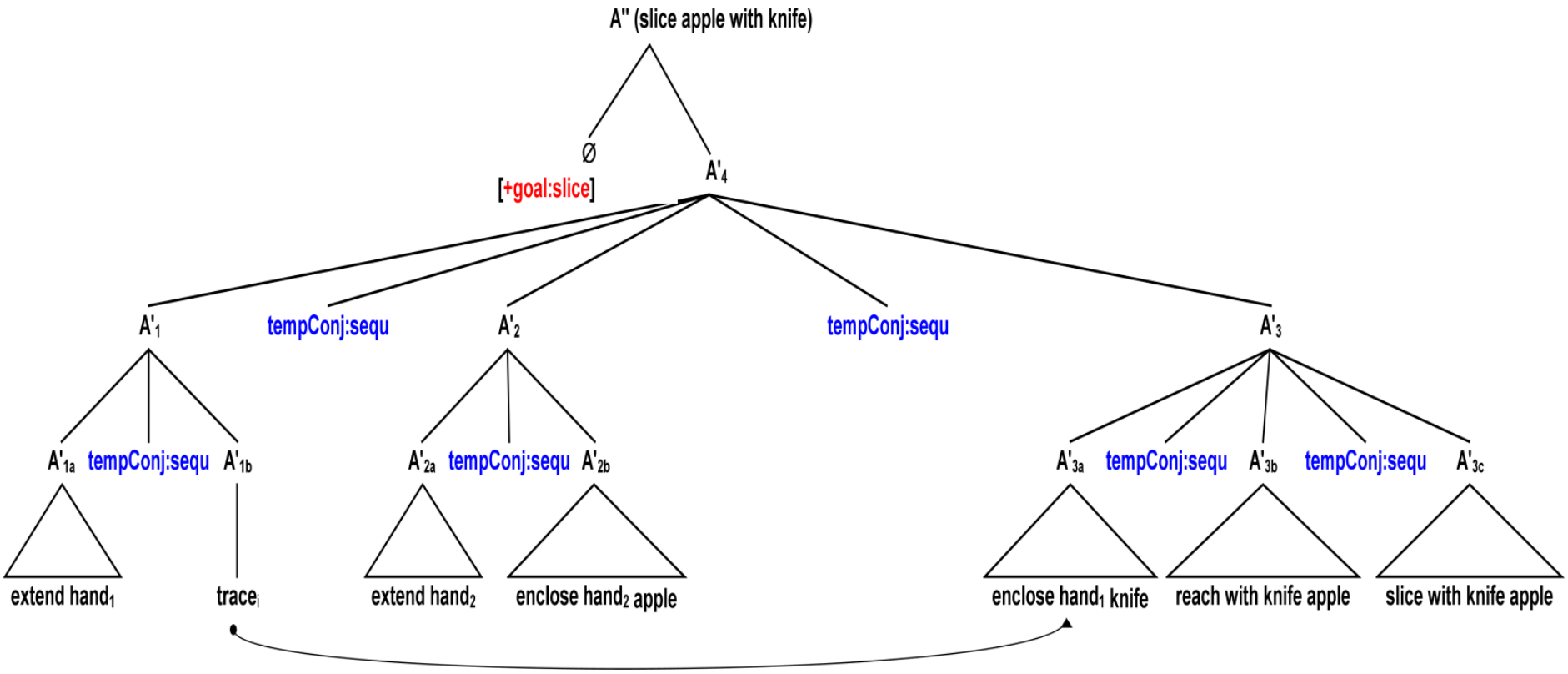
or in a more complex level:

"*grasp with hand$_1$ knife, pin with knife bread – grasp with hand$_2$ fork, pin with fork cheese, lick with tongue cheese – bite with teeth bread*".

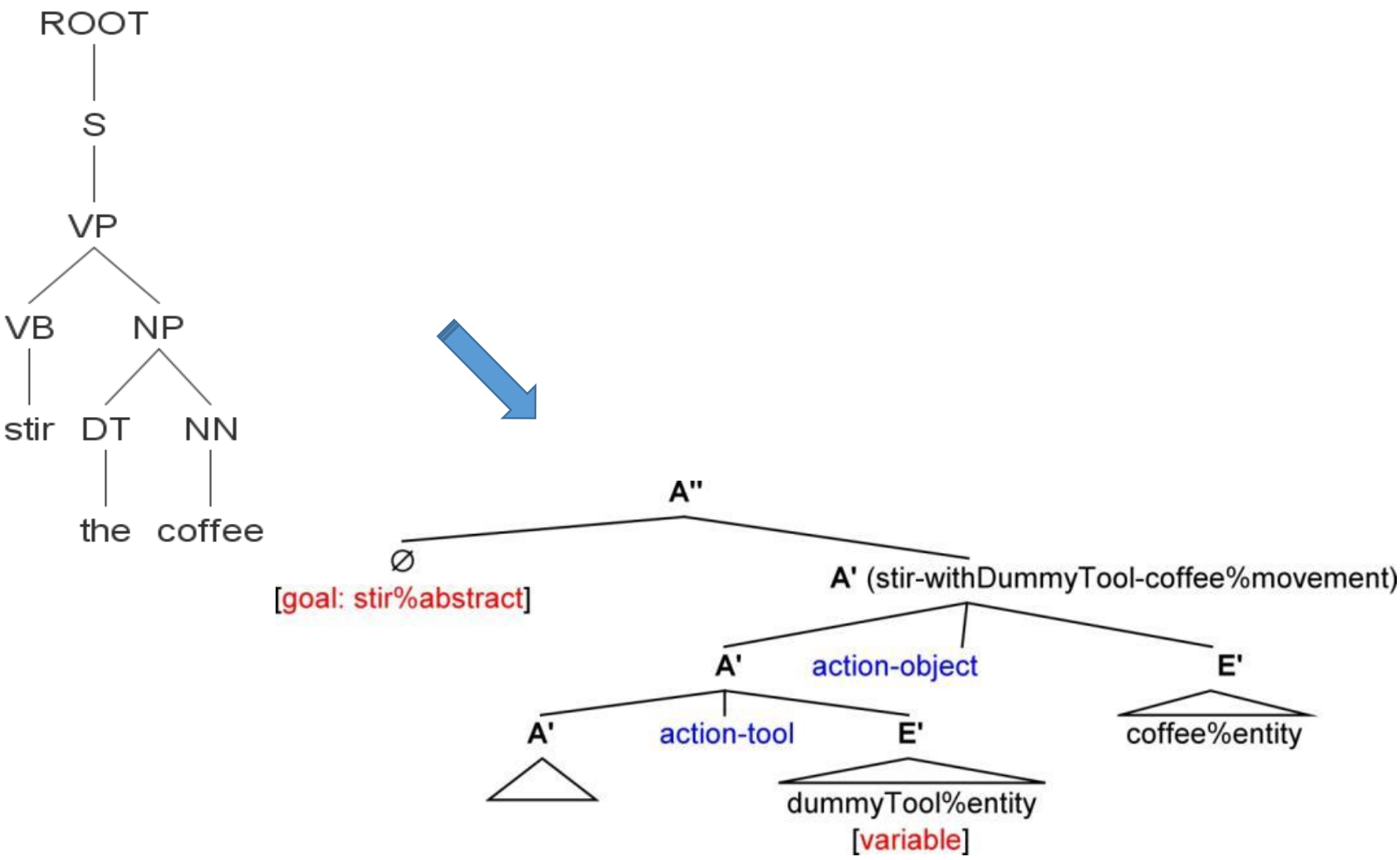
Recursion in Action (1)



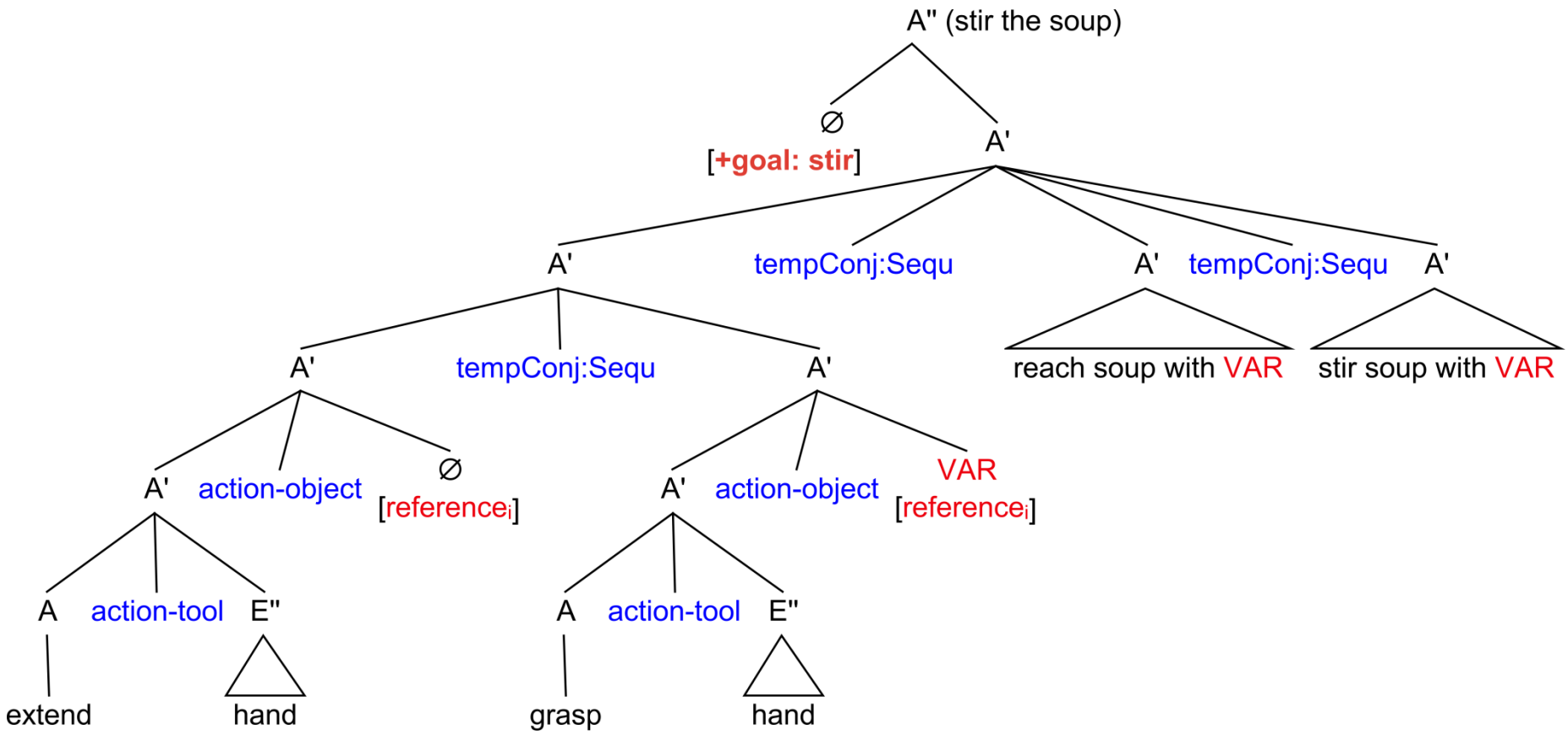
Recursion in Action (cont.)



Note: Language & Action Grammars



Note: Language & Action Grammars



Pastra et al. (2011), "Embodied Language Processing: a new generation of language technology", in Proceedings of the AAAI 2011 International Workshop on "Language-Action Tools for Cognitive Artificial Agents", San Francisco, USA.

Knowledge Representation

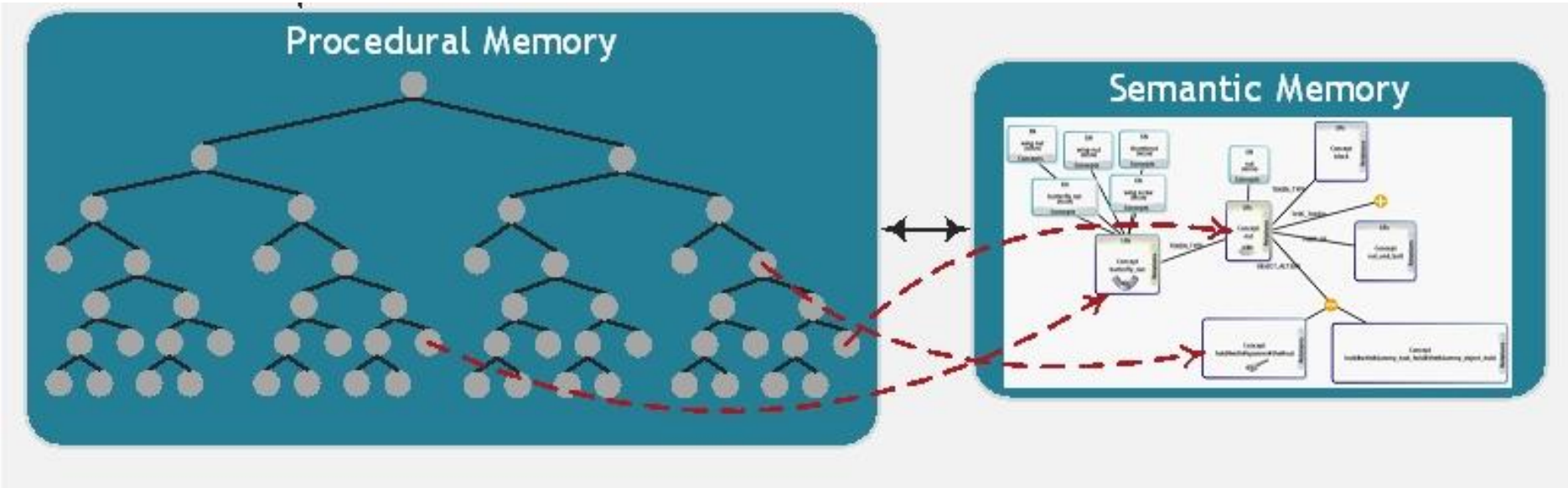
Long term Memory (see Tulvig 1972)

→ episodic (tied to specific learning experiences)

→ semantic (general knowledge of the world, and related generalisation and reasoning abilities - see also Quillian 1968 on semantic networks)

→ procedural (related to single action & action sequence learning, created through repeated learning)

Memory Interaction



Memory Issues

→ type of knowledge stored

→ structure of memory space

→ use/activations
(in memory search, retrieval, decision making)

Theories on Semantic Memory

Many theoretical accounts on structure & neural basis of SM (cf. extensive reviews in Kiefer and Pulvermueller 2012, McNorgan et al. 2011, Meteyard et al. 2012)

(1) **Concepts are flexible, distributed representations**; they **comprise modality-specific conceptual features** (latter stored in distinct sensorymotor brain areas) [Kiefer and Pulvermueller, 2012]

(2) **Much of the semantic memory content is related to perception and action and is represented in a brain region that overlaps with or corresponds to regions responsible for perception and action** (Patterson et al. 2007)

Theories on Semantic Memory (2)

How could it be implemented?

McClelland →

neuroscience evidence suggests SM to be implemented as a separate memory not subsumed to episodic memory. Suggestion that hippocampal formation and the neocortex form **complementary learning system**. Former facilitates auto and hetero-associative learning which is used to reinstate and consolidate gradually learned info in the neocortex.

Semantic Memory & Language

A number of Semantic Language Resources (SLR) around (of different types):

- Hierarchical lexical resources (e.g. WordNet)
- Common sense knowledge bases (e.g. ConceptNet, CYC)
- Ontologies, domain models, semantic lexicons, traditional dictionaries etc.

Traditional representation formats: **Semantic Networks (hierarchical or non)** (see Collins and Quillian 1969, Collins and Loftus 1975) and/or **Feature Bundles**

Categorization and story-telling…to learn / organise the world…

A number of cognitive architectures with recently incorporated semantic memory modules:

- SOAR (Laird et al. 2009), ACT-R (Anderson et al. 2004), ICARUS (Langley 2009)

SLR example

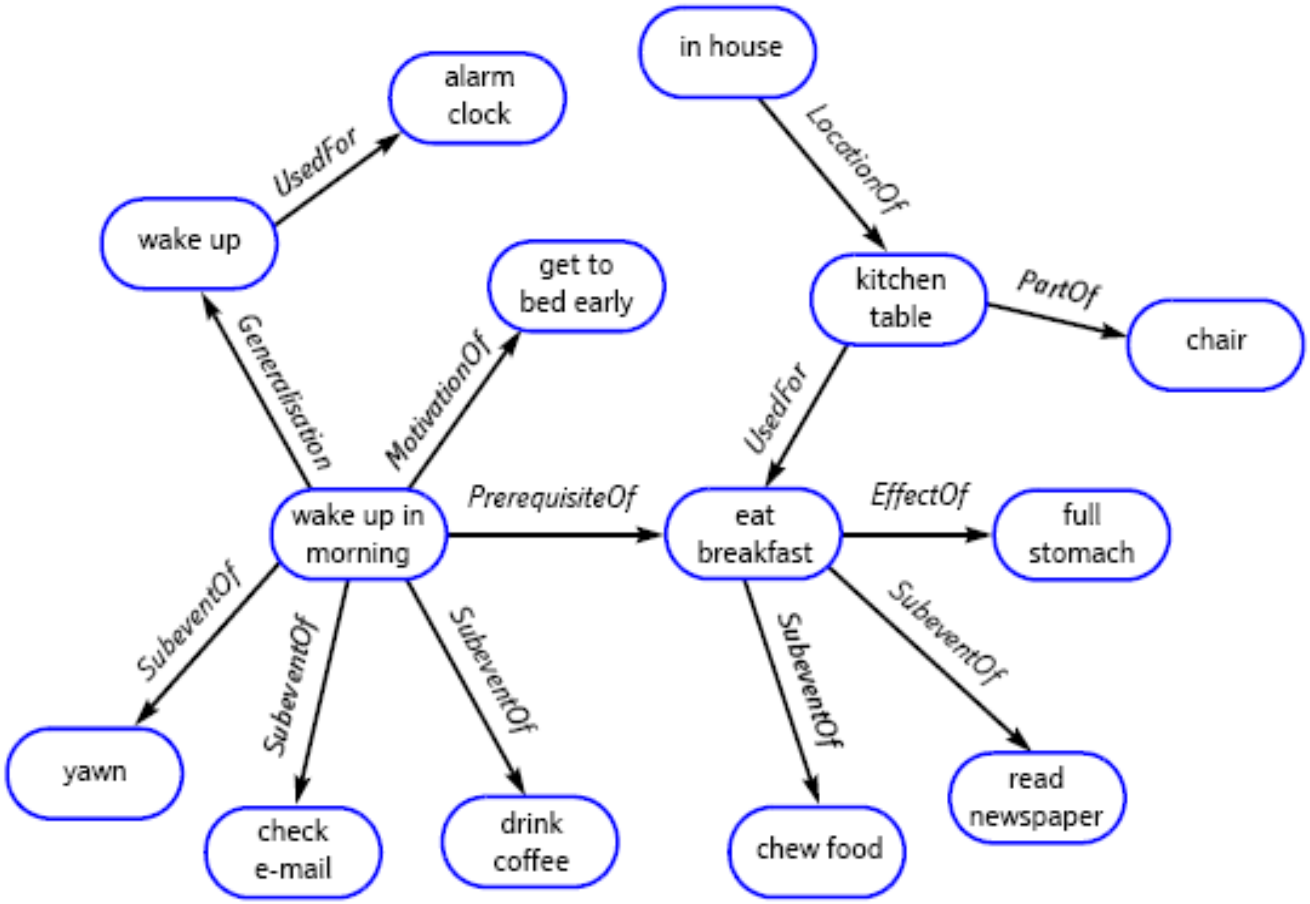


Fig 1 An excerpt from ConceptNet's semantic network of commonsense knowledge. Compound (as opposed to simple) concepts are represented in semi-structured English by composing a verb (e.g. 'drink') with a noun phrase ('coffee') or a prepositional phrase ('in morning').

SLRs and Visual Information

Multimedia Thesauri (e.g. Benitez et al. 2000)
Multimedia Ontologies (e.g. Zinger 2005 – OntoImage)
Multimedia Taxonomies (e.g. Hauptmann 2007 – LSCOM)
Multimedia Corpus (e.g. Pastra et al. 2010 – POETICONcorpus)
Labeled Image Databases (see review in Torralba 2011)

=> Long History: Ad hoc links of various types in AI systems since the late seventies (see review in Pastra and Wilks 2004)

=> Success Story for object recognition: The ImageNet Case (www.image-net.org)
14+ Million Images manually indexed to ~ 21K WN Synsets
~ 150K Images have bounding box around the object of interest
Images linked to Synsets at any level of the taxonomy; inheritance applies.

Semantic Memory & Language Assumptions

(1) Lack of Embodiment:

Common ASSUMPTION in all such resources, that agents have :

(a) sensorimotor experiences related directly or indirectly to what the language representations denote, and

(b) mechanisms for performing such link between language, perception and action

Aka: These modules/resources are NOT embodied, they are tied to language idiosyncrasies and lack structure that will unify language-perception-action.

Semantic Memory & Language Assumptions

(II) TOO OPEN:

Open world assumption implemented, however they are too open :

a) Lack principles on what comprises a concept unit, which are the available concept types, what kind of characteristics they have, which ones have modality-specific representations

b) Lack structure: how concepts are related to each other, which types of relations exist, etc.

Semantic Memory is NOT verbal

- On the relation of Cognitive Modules and Semantic Memory:

The language system (incl. mental lexicon) INTERACTS with information stored in memory, as other modules do too (i.e. Motor System and Perception)

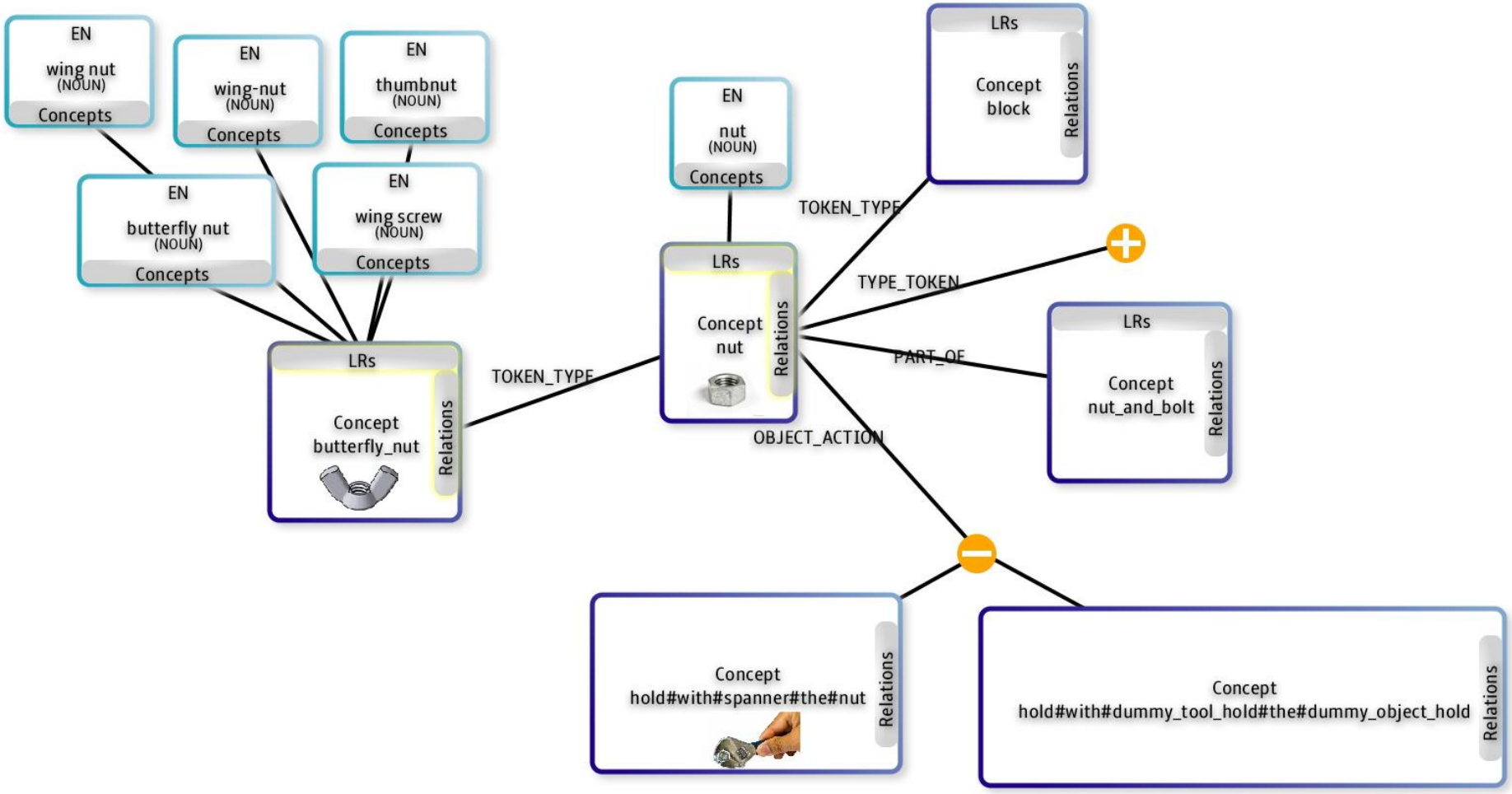
- On Symbols in Semantic Memory:

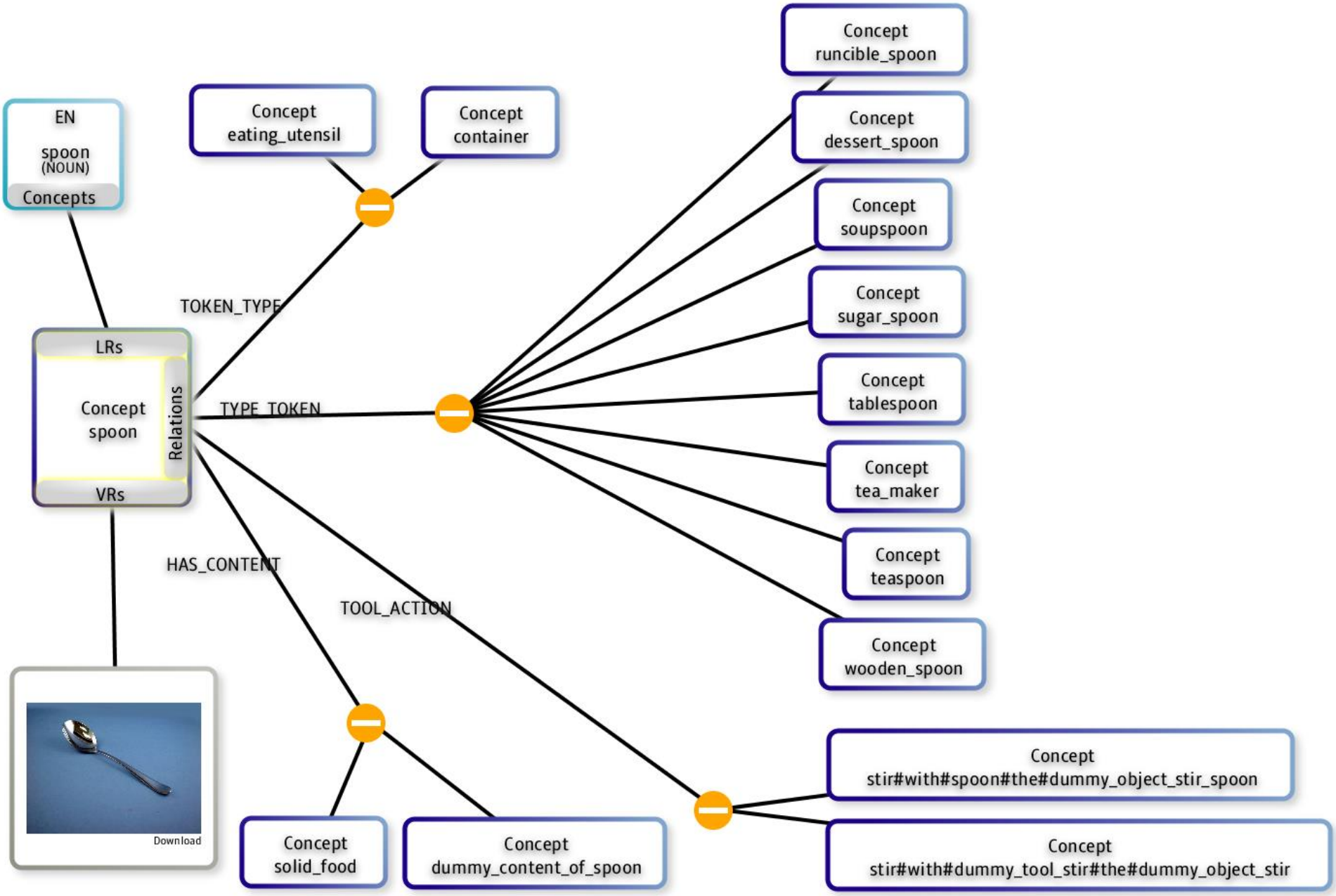
Language is not the only Symbol system; the Motor System and Perception (e.g. Vision) can become symbolic (see symbolic gestures, visual signs, certain sounds etc.).

- On Referentiality and non-Referentiality in Semantic Memory:

Language is not the only system that can be both referential and non-referential; the Motor System can be both fully referential (when objects present), conditionally referential (cf. pantomimes) and non referential (e.g. when it functions as a deictic mechanism)

The PRAXICON




EN
spoon
(NOUN)
Concepts
Concept
eating_utensil
Concept
container
Concept
runcible_spoon
Concept
dessert_spoon
Concept
soupspoon
Concept
sugar_spoon
Concept
tablespoon
Concept
tea_maker
Concept
teaspoon
Concept
wooden_spoon
TOKEN_TYPE
LRs
Concept
spoon
Relations
VRs
TYPE_TOKEN
HAS_CONTENT
TOOL_ACTION
Download
Concept
solid_food
Concept
dummy_content_of_spoon
Concept
stir#with#spoon#the#dummy_object_stir_spoon
Concept
stir#with#dummy_tool_stir#the#dummy_object_stir

The PRAXICON

- **Concepts** (nodes – multi-representational)

- **Relations** (edges – labeled, mostly bidirectional)

→One concept may have many relations to many concepts
BUT there is only one relation linking two specific concepts

Concepts: Characteristics
TYPE: entity, movement, feature (concrete or abstract)
STATUS: constant, variable
SPECIFICITY LEVEL: Basic Level, Superordinate, Subordinate
DOMAIN: e.g. Artifact, Animal etc.

PRAXICON Relations

Relations: a finite set

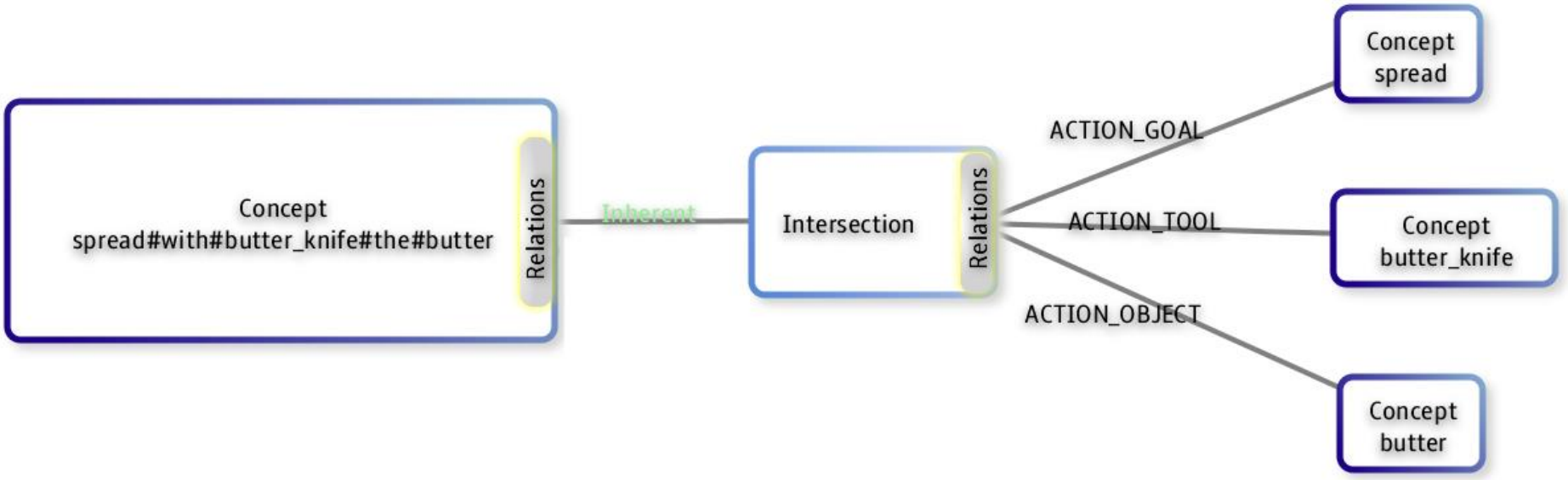
ACTION_AGENT
ACTION_GOAL
ACTION_OBJECT
ACTION_RESULT
ACTION_TOOL
ASPECT_CONCEPT
COMPARED_WITH
ENABLES

MORE
LESS
METAPHOR_OF
PRODUCER_OF
TYPE_TOKEN

HAS_ANTHROPOGENIC_EFFECT
HAS_COLOUR
HAS_CONDITION
HAS_CONTENT
HAS_DEPTH
HAS_FORCE
HAS_HEIGHT
HAS_HUE
HAS_INSTANCE
HAS_INTENSITY
HAS_LENGTH
HAS_LOCATION
HAS_LUMINANCE
HAS_MATERIAL

HAS_MEASUREMENT_UNIT
HAS_MEASUREMENT_VALUE
HAS_MEMBER
HAS_NATURAL_EFFECT
HAS_PART
HAS_PARTIAL_INSTANCE
HAS_SHAPE
HAS_SIZE
HAS_SPEED_RATE
HAS_STEP
HAS_TEMPERATURE
HAS_TEXTURE
HAS_TIME_PERIOD
HAS_VISUAL_PATTERN
HAS_VOLUME
HAS_WEIGHT
HAS_WIDTH

PRAXICON Action Example



Referential Words

Referential		
Physical Reference	**Non-Physical Reference**	**Reference is Conceptual Structure**
entity-concrete	entity-abstract	state
movement-concrete	movement-abstract	event
feature-concrete	feature-abstract	exceptional feature

Architecture, Mix, Cut, Doctor, Mother, Poverty, Sliced, Closed, Make salad, Prepare Dinner, Shopping, Wooden Egg…… People

Non-Referential Words

Non-Referential		
Phatic Communication	**Expressive Function**	**Reference Modification**
greetings	interjections	logical operator
politeness expressions	symbolic corporeal sounds	epistemic operator
		mathematical operator, temporal frame operator

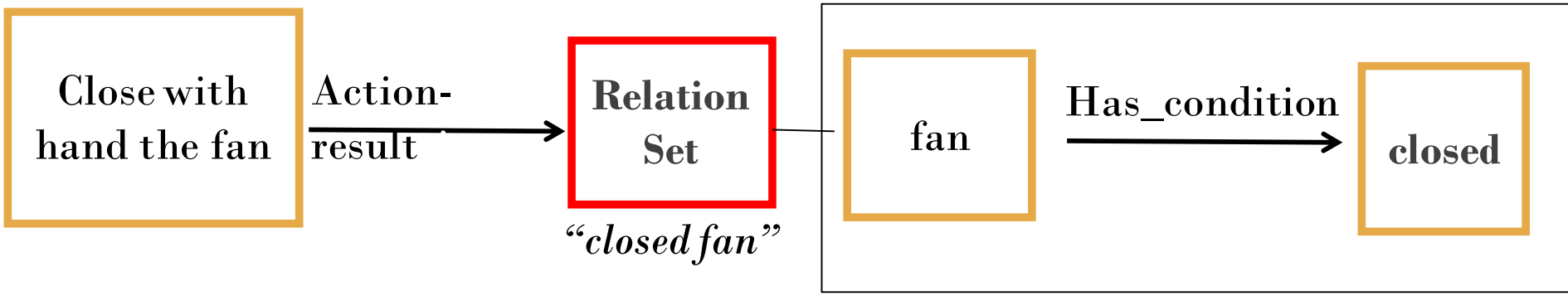
Hi, thanks, oh, and, could, more, less, finish……

Conditionally-Referential Words

Conditionally-Referential (referential when grounded)		
Indexical	**Spatiotemporal**	**Vocative**
deictics	adverbials	person names
pronouns	prepositions	animal or environmental sounds
		onomatopoeia in general

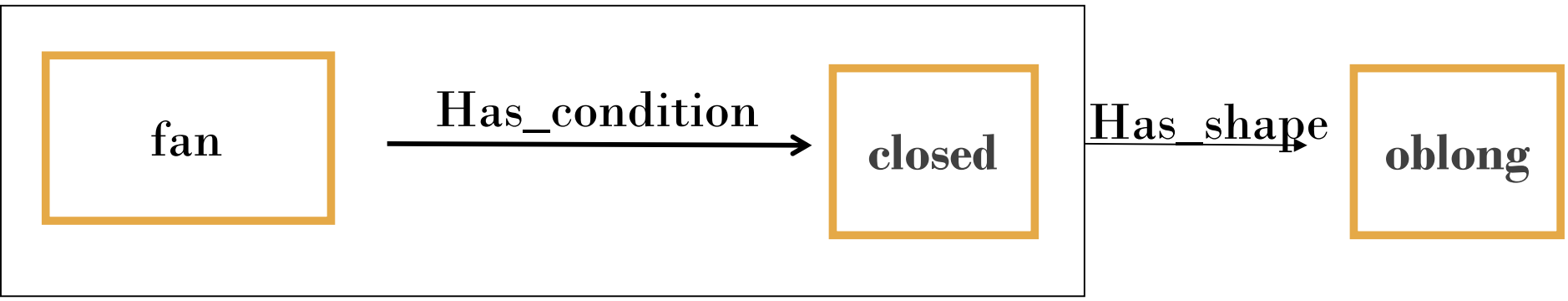
Examples: This, him, mine, now, here, outside, tomorrow, beside, Jane, woof, baa etc.

PRAXICON Relation Set Example



Why is such representation important?
Consider: "the fan is oblong"

PRAXICON Relation Set Example



Recursive Conceptual Structures

- RS → C RP | RP C
- RP → C R | R C
- R → $C1_{RT_x}$ $C2_{RT_x}$

RS: RELATION SENTENCE
RP: RELATION PHRASE
R: RELATION
C: CONCEPT
RT: RELATION TYPE

EXAMPLE:

R: fan **hasCondition** closed

RP: R **hasShape** oblong

RS: RP **tool-action** push_withFan_ball

or in language terms:

R: closed fan

RP: closed fan is oblong

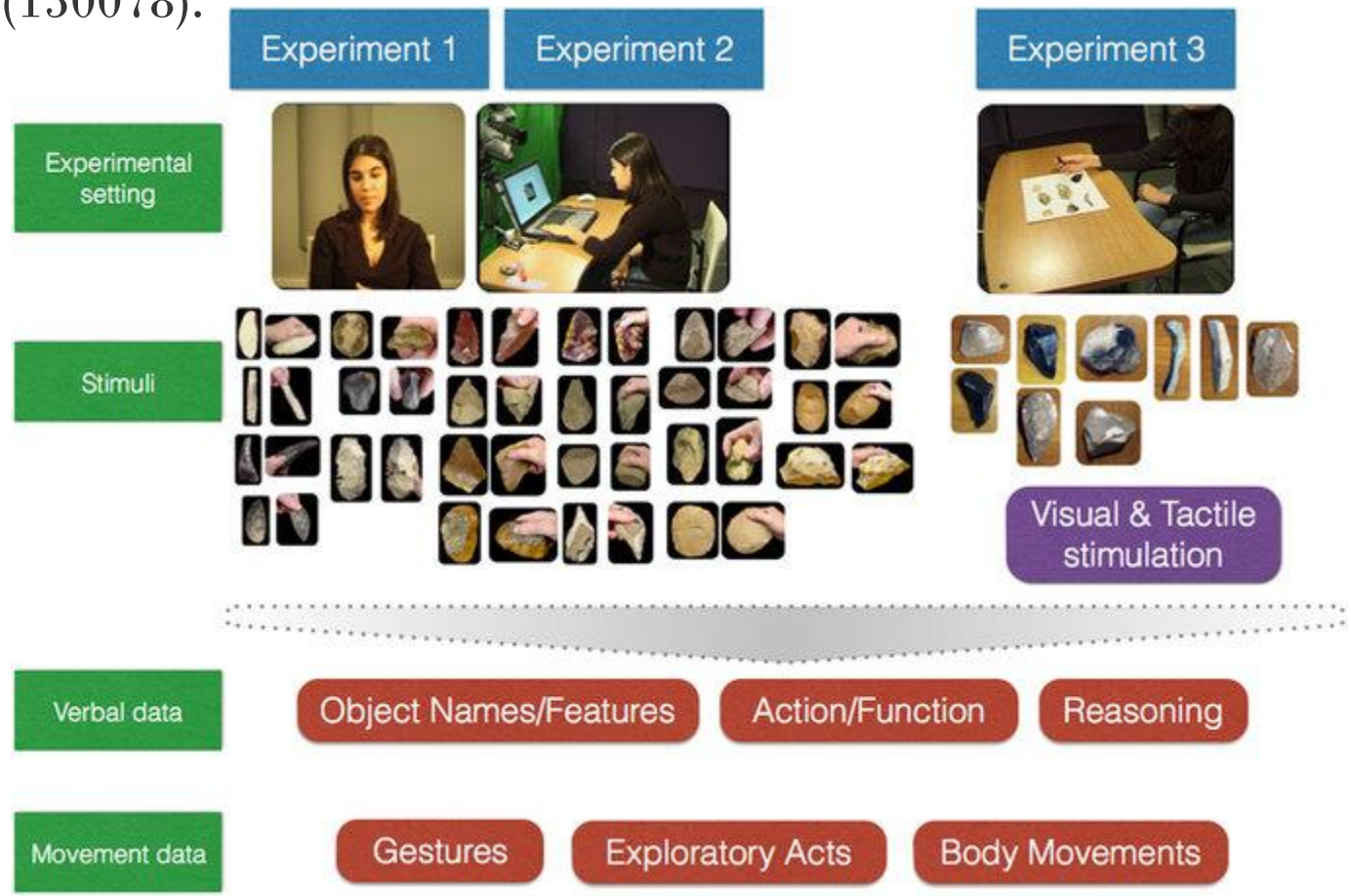
RS: push the ball with the closed fan

(action enabled by the oblong shape of the closed fan)

Rich corpus of object affordances

Vatakis, A. , & Pastra, K. (2016). **A multimodal dataset of spontaneous speech and movement production on object affordances**. Nature, *Scientific Data*, *3*(150078).

nature publishing group

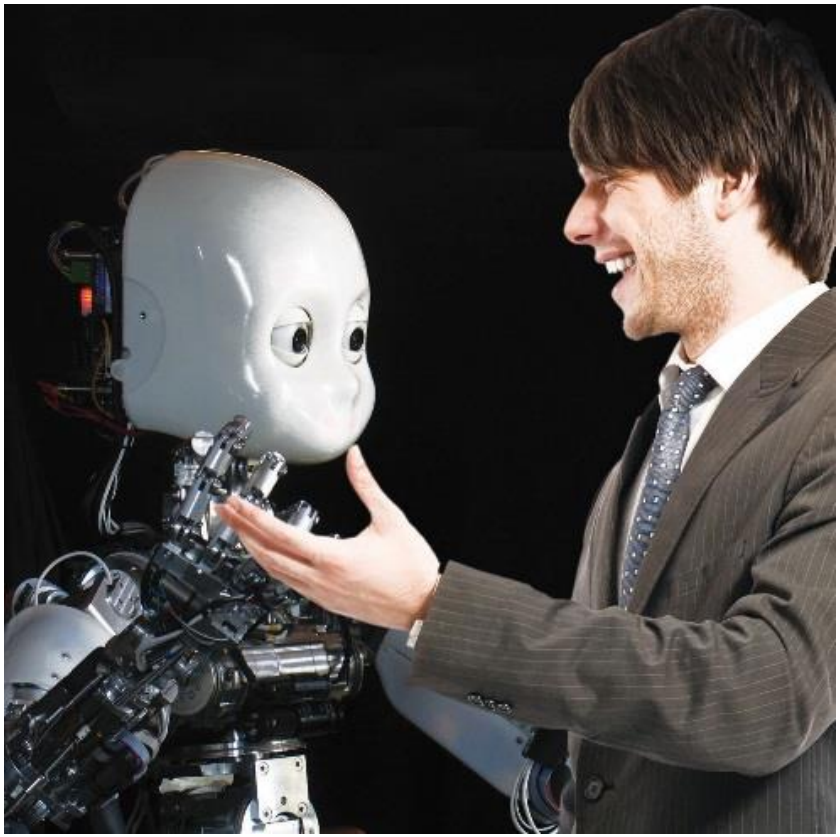


More than 120 participants, more than 90 hours of spontaneous speech…

Rich corpus of object affordances (2)

Semantic categories	Number of unique categories	Examples
Object Naming	2,942 (only 10 of those referred to the stimuli presented)	• *This is a stone, a small stone.*
		• *It has a handle here.*
Object Features	1,356	• *It's a dark green object.*
		• *It has a rounded shape.*
		• *This is quite soft.*
Object Uses	583	*I can use this to cut meat into small pieces.*
Reasoning patterns	**Number of instances**	**Examples**
Justification: X	3,060	*One could cut things with this because it is sharp.*
Comparison: X>feature>Y	622	*The upper part is softer than the bottom part.*
Conditional: X	1,183	*This could also be used as a plate if it was bigger….*
Analogy: X	702	*…it has the colour of the sand …*
Movement categories	**Number of instances**	**Examples**
Exploratory Acts	11,209	Rubbing an object's surface.
Emblems	1	The symbolic gesture of 'ok'.
Deictic	218	Pointing at various parts of the presented object.
Metaphoric	12	The gesture for 'on and on'.
Iconic-Pantomime	217	The enactment of 'writing with a pen' with the hand configured as if holding a pen while moving to write.
Pantomime-Metaphoric	14	The enactment of 'writing with a pen' with the hand having the role of a pen.
Demonstrations	556	Demonstration on how one uses a knife to cut something but without actually cutting anything.
Body Movements	74	The movement of weighting the lithic tool in one's hand.

Humanoids with merging abilities



https://www.youtube.com/watch?v=IOpn3BYqLRk

http://www.youtube.com/watch?v=FNeYixxmeTM