Embodied and Situated Language Processing 2012

28-30 August 2012

Newcastle Upon Tyne
England

Goal

- bring together researchers working on the interaction of language and visual/motor processing in embodied, situated, and language-for-action research traditions.
- unite converging and complementary evidence from three different methods (behavioral, neuropsychological, and computational).
-

2012 Focus

- focus on embodiment in language learning.
- How and when grounding occurs in learning a language?
- What role does grounding play in language evolution, and language learning in robots?

Gerry Altmann– York University, UK

- His research interests include sentence processing, ambiguity resolution, eye movements during reading and listening, and the implicit learning of grammatical information.
- Editor-in-chief Cognition
- http://www.york.ac.uk/psychology/staff/faculty/gtma1/#research
-
- **Event comprehension in the brain: Objects 'before' compete with themselves 'after'.**
- fMRI study of repr. of object state-change

Bernard Hommel, University of Leiden, the Netherlands

- **Research**
- BH is interested in how, why, and when people can do what they want, that is, how intentional actions are set up, implemented, and controlled. He also wants to know how perception and action relate to each other, whether and how what we do is affected by what we perceive, and vice versa. Finally, he is interested in how the events we perceive and the actions we do are represented in our brains, and how distributed brain codes can be integrated into coherent cognitive structures. In particular, BH's research focuses on these issues:
- Cognitive representation of action plans
- Acquisition of intentional action
- Representation of self and other
- Interactions between perception and action
- Integration of distributed information
- Attentional control and control of attention
- Impact of religion and culture on cognition and action
- Development of cognitive functions across lifespan
- Creativity
- Neuromodulation of cognitive processes
- Cognitive neurorobotics

Integrating perception and action: The Theory of Event Coding

- Hommel, B. (2013). Ideomotor action control: On the perceptual grounding of voluntary actions and agents. In W. Prinz, M. Beisert & A. Herwig (Eds.), *Action science: Foundations of an emerging discipline* (pp. 113-136). Cambridge, MA: MIT Press.

Jesse Snedeker, Harvard University, USA

- https://software.rc.fas.harvard.edu/lds/research/snedeker/jesse-snedeker
-
- Embodied cognition(s), development and language: An outsider's perspective.

Empirical papers (selected)

- Anuenue Kukona (University of Dundee, UK), Gerry Altmann (University of York, UK) and Yuki Kamide (University of Dundee, UK): **Competition dynamics in the representation of location in a situated language context**.
- Kenny Coventry (University of East Anglia and Northumbria University, UK), Debra Griffiths (Northumbria University, UK) and Colin Hamilton (Northumbria University, UK): **Language and the perceptual parameters affecting the representation of space**.
- Maria Staudte (Saarland University, Germany), Matthew Crocker (Saarland University, Germany), Alexander Koller (University of Potsdam, Germany) and Konstantina Garoufi (University of Potsdam, Germany): **Grounding spoken instructions using listener gaze in dynamic virtual environments**
- Eiling Yee (Basque Centre on Cognition, Brain and Language, Spain), Lisa Musz (University of Pennsylvania, USA) and Sharon Thompson-Schill (University of Pennsylvania, USA): **Mapping the similarity space of concepts in sensorimotor cortex**.

Art Glenberg

- http://glial.psych.wisc.edu/index.php/psychsplashfacstaff/103

Modeling paper

- Kerstin Fischer (University of Southern Denmark, Denmark), Davide Marocco (University of Plymouth, UK), Anthony Morse (University of Plymouth, UK) and Angelo Cangelosi (University of Plymouth, UK): **Embodied language learning and tacit distributional analyses: A comprehensive framework for learning new words**.
- http://www.youtube.com/watch?v=5l4LHD2lYJk&feature=plcp

Language learning with meanings as stored sensorimotor sequences: a connectionist model

Martin Takac[1,2], Lubica Benuskova[2] and Alistair Knott[2]
[1]Dept. of Applied Informatics, Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia. takac@ii.fmph.uniba.sk
[2]Dept. of Computer Science, University of Otago, Dunedin, New Zealand, lubica@cs.otago.ac.nz, alik@cs.otago.ac.nz

Introduction

In accordance with the embodied view on cognition, we conjecture that **high-level semantic representations of concrete episodes are delivered by sensorimotor routines.**

We focus on representation of concrete episodes (events or states) that can be described by transitive sentences, e.g. reaching to grasp an object. We argue that **experience of a transitive episode involves a canonical sequence of sensorimotor operations – a *deictic routine*.** A deictic routine involves a sequence of attentional or motor operations interleaved with transitory sensory consequences.

In our model, an observer stores recently experienced episodes in working memory as prepared sensorimotor sequences that can be internally replayed.

Sustained signals	Transient signals			
	Initial context	Operation	Reafferent signal	New context
plan(attend_agent, attend_cup, grasp)	C_1	attend_agent	agent_rep	C_2
plan(attend_agent, attend_cup, grasp)	C_2	attend_cup	cup_rep	C_3
plan(attend_agent, attend_cup, grasp)	C_3	grasp	agent_rep	C_4
plan(attend_agent, attend_cup, grasp)	C_4		cup_rep	

Sequence of signals produced during a replay of the cup-grabbing deictic routine. Note that the agent and patient are each attended more than once.

Conjecture for language: **A speaker needs to internally replay a stored episode representation in order to express it verbally.**

Problem

Episode representations are cross-linguistically universal, yet particular languages differ in their grammar and surface forms. Learning task for an infant: acquire from a sample of a language general syntactic rules and particular surface regularities (such as idioms) of their language.

Our goals

- **Propose a connectionist model of sentence generation able to learn a mapping from sensorimotor sequences (episodes) to phonological sequences (utterances).** Train the model on a language that contains a mixture of syntactically regular and idiomatic sentences. Answer the following questions:
- **Is the model able to learn different word-ordering conventions?**
- **Does the model generalize well to previously unseen episodes?**
- **How well does the model produce idioms?**
- **Does the learning in the model follow a developmentally realistic trajectory?**

Method

Generate a basic artificial language of transitive sentences for each of the following word orders: SVO (subject verb object), SOV, VSO, VOS, OSV, OVS.

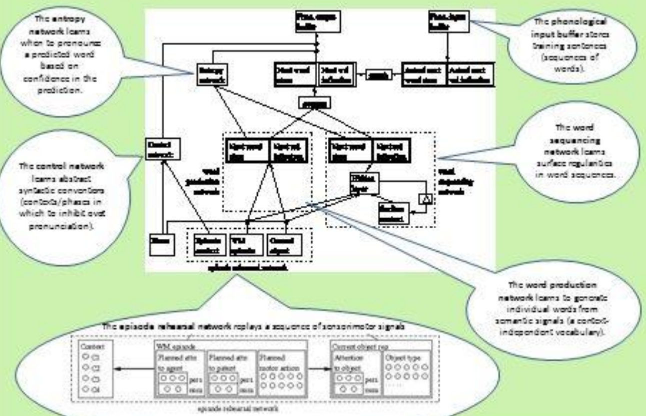
In each case, train a model on sentences paired with their meanings (sequences of sensorimotor signals) and test it for a sentence generation task. For each word order, 10 model instances with different initial random weights are trained on a random subset of the target language and tested on a different subset of the language.

Model

The sentence generation task involves mapping an episode representation onto a sequence of words. The episode is presented to the network as a sequence of sensorimotor representations in the episode-rehearsal system. In each stage of episode rehearsal, the word sequencing network is given a chance to produce more than one word for a current sensorimotor signal, which enables production of idiomatic phrases.

The network alternates between two modes of iteration:

1) The episode rehearsal system iterates through the sequence of SM signals until it reaches a context at which the control network allows a word to be overtly pronounced.
2) The word production/sequencing networks work together to produce a word for a currently active SM signal. If they can confidently predict the next word, the word is pronounced, the word sequencing network updates its surface context layer and the model carries on in this mode until the networks can no longer confidently predict the next word.



The SVO language

Vocabulary: 105 words from the Child Development Inventory (Fenson et al., 1994)

Morphology: number (SG,PL) inflections of nouns, number and person (1st, 2nd, 3rd) inflections on verbs, subject-verb agreement, irregular plurals (leaves, fish, teeth, women, etc.), personal pronouns.

Syntax: 127088 sentences of three types :

- **Syntactically regular sentences (Reg, 80.6%)**, like *"Mice bite-3pl dog-sg."*,
- **Sentences with continuous idioms (Cont, 13.0%)**, like *"Mia-sg lick-3sg ice cream-sg."* or *"You tickle-2sg Winnie the Pooh-sg."*,
- **Sentences with discontinuous idioms (Disc, 6.4%)**, like *"Grandpa-sg give-3sg grandma-sg a hug."* or *"Daddy-sg kiss-3sg teddy bear-sg good bye."*,

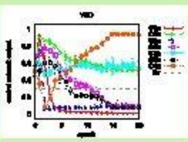
Training set: 4000 sentences (5% of language) paired with meaning, e.g.

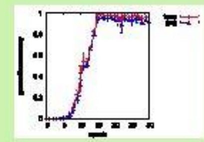
Meaning							Sentence
Context							
C_1(AG)		C_2(PAT)		C_3(AG)		C_4(PAT)	
a	b	a	b	a	b	a	
PERSPRON/1/PL	(ACT)	MUMMY/3/SG	(ACT)	PERSPRON/1/PL	(ACT)	MUMMY/3/SG	We tickle-1pl mummy-sg .
WM episode: 1/PL | 3/SG | TICKLE							
POOH/3/SG	(ACT)	HELEN/3/SG	(ACT)	POOH/3/SG	(ACT)	HELEN/3/SG	Winnie the Pooh-sg kiss-3sg Helen-sg good bye .
WM episode: 3/SG | 3/SG | FAREWELL							

Test set: 4000 previously unseen sentences/meanings.

Results

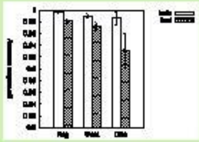
1. The model learns all word-ordering conventions.





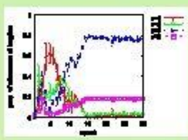
The control network learns the correct inhibition patterns for all the word orders in 100% of cases (the left picture shows the VSO case).
The generation accuracy is the proportion of correctly generated sentences.
The model is able to generalize to unseen episodes from just about 3% of data.

2. The model learns to correctly produce sentences with idiomatic phrases.



The SVO case (N=10)

3. The model goes through developmental stages, from producing single-word utterances (1) through proto-syntactic 'item-based' constructions (2) to sentences with full syntax and agreement morphology (3).



Examples of sentences generated for the meaning DOG/3/SG (AG), CHASE|3/SG|3/PL (ACT), RABBIT/3PL (PAT) during training (in the SVO language):

(1) *"dog-sg" ... "rabbit-pl" ... "chase" .. "chase-3sg"* (epoch 5)

(2) *"dog-sg chase-3sg" ... "rabbit-pl chase-3sg" .. "rabbit-sg chase-3sg"* (epoch 10)

(3) *"dog-sg chase-3sg rabbit-pl"* (epoch 15)

Conclusions

We implemented a connectionist model of sentence generation from meanings represented as replayed sensorimotor sequences. The model is able to learn abstract syntactic conventions as well as surface regularities of a particular language. Different learning tasks bootstrap each other from scratch in a developmentally plausible way.

Reference:
Takac, M., Benuskova, L., Knott, A.: Mapping sensorimotor sequences to word sequences: A connectionist model of language acquisition and sentence generation. **Cognition** 2012. http://dx.doi.org/10.1016/j.cognition.2012.06.006

Nomen omen ☺

- Francesca Citron, Michael Kucharski (Freie Universität Berlin, Germany) and Adele Goldberg (Princeton University, USA): An Effect of language on embodied metaphor: Sour vs. Sauer
- Laura Speed (University College London, UK) and Gabriella Vigliocco (University College London, UK): The multimodal meaning of speed in language.

Nomen omen ☺

- Francesca Citron, Michael Kucharski (Freie Universität Berlin, Germany) and Adele Goldberg (Princeton University, USA): An Effect of language on embodied metaphor: Sour vs. Sauer
- Laura Speed (University College London, UK) and Gabriella Vigliocco (University College London, UK): The multimodal meaning of speed in language.