Episode representations as stored sensorimotor sequences: A case study in sentence generation

Martin Takac, Alistair Knott and Lubica Benuskova

Department of Computer Science, University of Otago, Dunedin, New Zealand, takac@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.

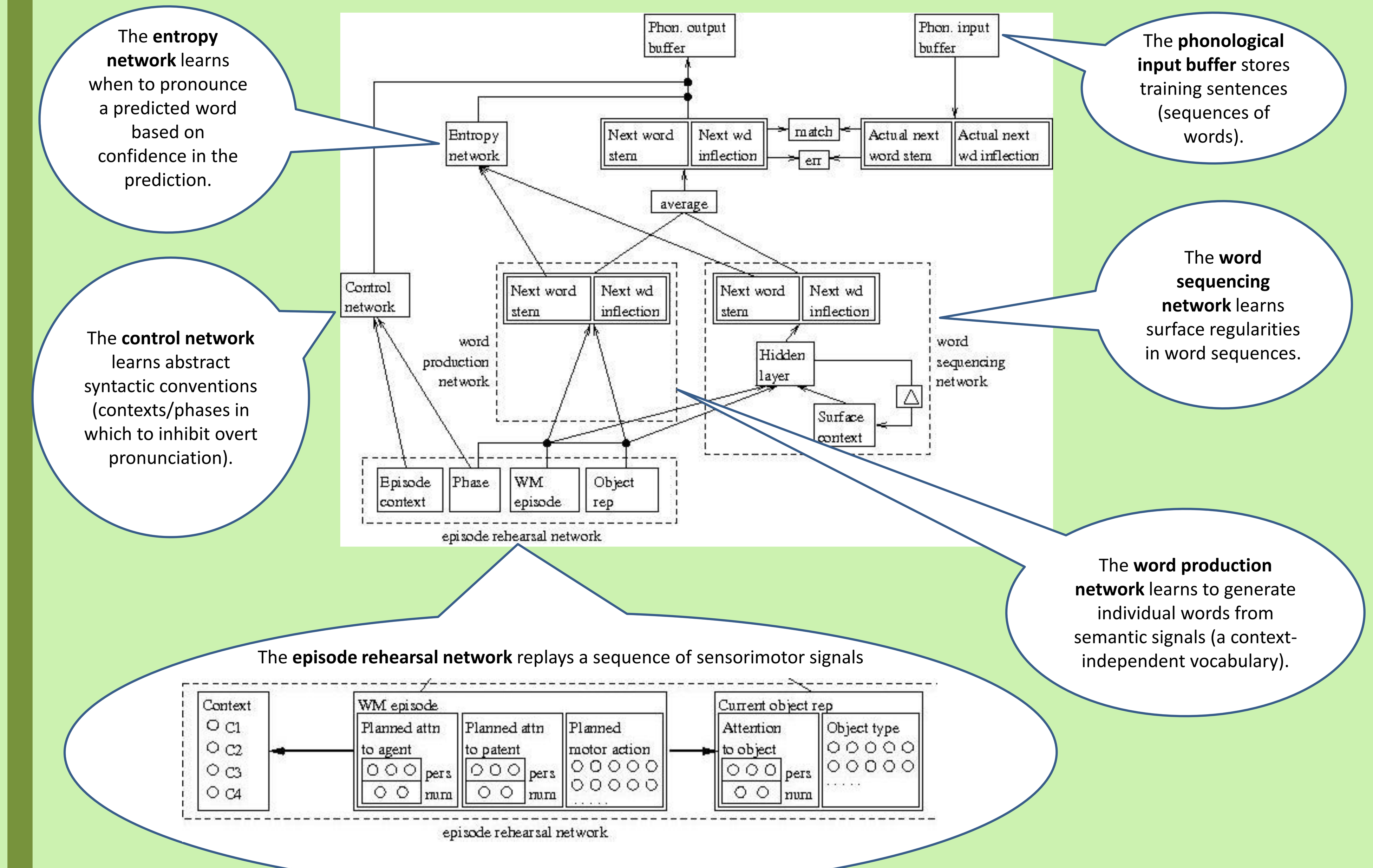
In each case, train a model on sentences paired with their meanings (sequences of sensorimotor signals) and test it for a sentence generation task. For each word order, 10 model instances with different initial random weights are trained on a random subset of the target language and tested on a different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

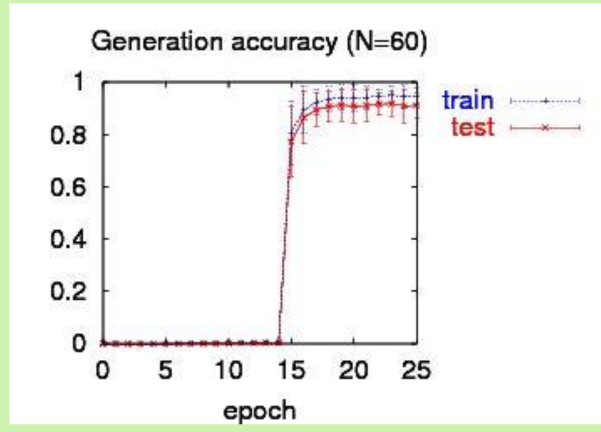
Training set: 4000 sentences (3% of language) paired with meaning, e.g.

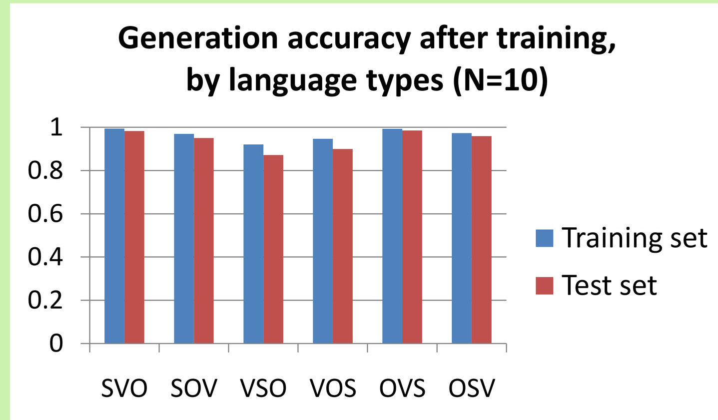
Meaning							Sentence
Context							
C_1(AG)		C_2(PAT)		C_3(AG)		C_4(PAT)	
a	b	a	b	a	b	a	
PERSPRON/1/PL	(ACT)	MUMMY/3/SG	(ACT)	PERSPRON/1/PL	(ACT)	MUMMY/3/SG	*We tickle-1pl mummy-sg .*
WM episode: 1/PL | 3/SG | TICKLE							
POOH/3/SG	(ACT)	HELEN/3/SG	(ACT)	POOH/3/SG	(ACT)	HELEN/3/SG	*Winnie the Pooh-sg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

Test set: 4000 previously unseen sentences/meanings.

Results

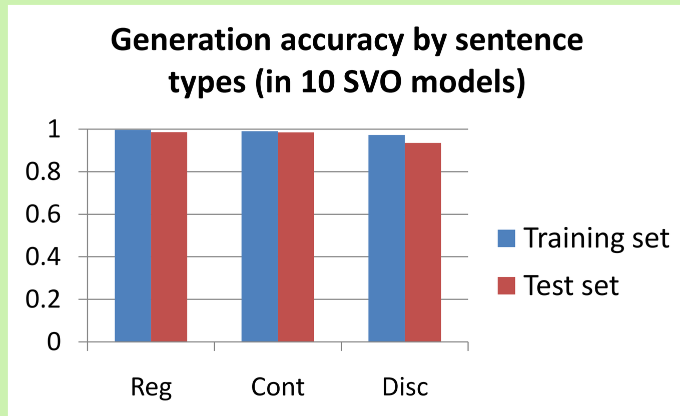
1. The model learns all word-ordering conventions.



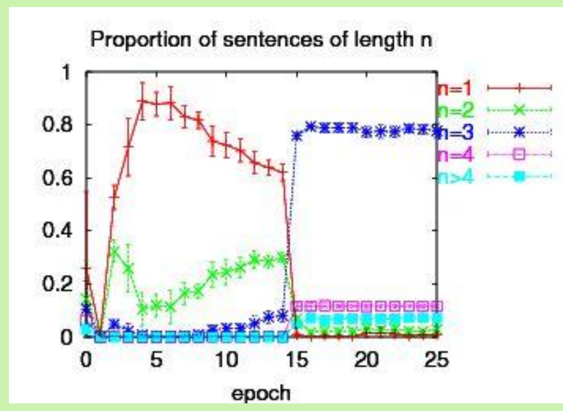


Generation accuracy is the proportion of correctly generated sentences.
The model is able to generalize to unseen episodes from just about 3% of data.

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



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.