Connectionist Semantic Systematicity in Language Production

Jesús Calvillo, Harm Brouwer, Matthew W. Crocker
Saarland University, Germany

Systematicity

"The ability to produce/understand some sentences is intrinsically connected to the ability to produce/understand certain others"

Fodor & Pylyshyn (1988, p. 37)

If you understand…

Brad kisses Angelina.

Perhaps because you had heard/read the sentence before.

…then you should also understand:

Angelina kisses Brad.

Even though you might have never heard/read it before.

→NOT just memorization.

According to Fodor and Pylyshyn (1988):

- Connectionist models are not able to display systematicity without implementing a classical symbol system.

But…

- Frank et al.(2009) present a connectionist model of comprehension that achieves relevant levels of systematicity.
  - Unseen sentences
  - Unseen situations

This talk…

Is Frank et al.(2009)'s approach suitable to model **language production**??

Can we also achieve **systematicity**??

Frank et al. (2009)'s model

Sentence Comprehension

"charlie plays soccer"

play(charlie,soccer)

Zwaan & Radvansky (1998). *Psychol. Bull.*

Distributed Situation Space (DSS) model

A richer representational scheme

We can *represent* that Charlie is outside, on a field, playing with a ball, and with others, etc.

With knowledge about the world

We *know* that Charlie is probably outside on a field, because soccer is typically played on a field, with a ball, with others, etc.

DSS—The main idea

Take a snapshot of the world ("a sample") at many different times, and for each snapshot write down the *full state-of-affairs* in the world.

Next: extract regularities—*world knowledge*—from the full set of observations, and construct meaning representations (vectors) that encode this world knowledge.

Problem: How to record full state-of-affairs in the world for each snapshot?

> use a confined *microworld* (which limits the scope of the world)

Defining a Microworld

A *state-of-affairs* (situation) in a microworld is defined in terms of *basic events* that can be assigned a state (i.e., they can be *the case* or not *the case*)

Class	Variable	Class members (concepts)	#
People	p	charlie, heidi, sophia	3
Games	g	chess, hide&seek, soccer	3
Toys	t	puzzle, ball, doll	3
Places	x	bathroom, bedroom, playground, street	4
Manners of playing	m_{play}	well, badly	2
Manners of winning	m_{win}	easily, difficultly	2
Predicates	—	play, win, lose, place, manner	5

Event name		#
play(p, g)	$3 \times 3 =$	9
play(p, t)	$3 \times 3 =$	9
win(p)		3
lose(p)		3
place(p, x)	$3 \times 4 =$	12
manner(play(p), m_{play})	$3 \times 2 =$	6
manner(win, m_{win})		2
	Total	44

More specifically, states-of-affairs are combinations of these 44 basic events

Example—"heidi loses at chess": $play(heidi, chess) \wedge lose(heidi)$

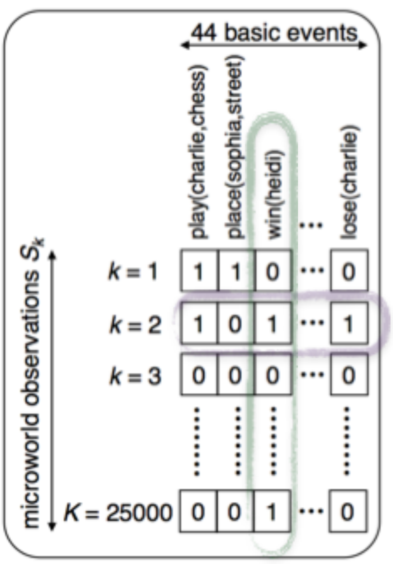
> 2^44 (≈10^13) possible situations, but world knowledge precludes many

Note: there are hard (being there) and probabilistic (preferences) constraints

Frank et al. (2009). *Cognition*

Situation-State Space

Many samples of microworld situations constitute a "*situation-state space*"



Rows represent observations (states-of-affairs)

Columns represent situation vectors for basic events:

Using (fuzzy) logic, *complex event* vectors can be derived:

$$\vec{v}(\neg a) = 1 - \vec{v}(a)$$

$$\vec{v}(a \wedge b) = \vec{v}(a)\vec{v}(b) \quad \text{where} \quad \vec{v}(a \wedge a) = \vec{v}(a)$$

Finally, a dimensionality reduction is applied in order to go from **25k** dimensions to **150**.

Frank et al. (2009). *Cognition*

So now we have a way to represent events (basic and complex) in terms of the **situations** in which they are true.

$$play(heidi, chess) \wedge lose(heidi) \quad \longleftrightarrow \quad \begin{bmatrix} 0.1 \\ 0 \\ 1.0 \\ 0.03 \\ \ldots \\ 0.8 \\ 0.9 \end{bmatrix}$$

Our model: Belief Vectors

- No dimensionality reduction.
- Instead of defining the meaning of an event in terms of the situations in which it is true, define it in terms of the **basic events** with which it appears.
  - *belief* vectors
- Dimensionality := # basic events
- Each dimension:
  - P(basic event | complex event)

Defining a Microlanguage - Lexicon

Class	Words	#
proper nouns	*charlie, heidi, sophia*	3
(pro)nouns	*boy, girl, someone, chess, hide-and-seek, soccer, football, game, puzzle, ball, doll, jigsaw, toy, ease, difficulty, bathroom, bedroom, playground, shower, street*	20
verbs	*wins, loses, beats, plays, is, won, lost, played*	8
adverbs	*well, badly, inside, outside*	4
prepositions	*with, to, at, in, by*	5
	Total	40

Frank et al. (2009), *Cognition*

Defining a Microlanguage - Semantics

charlie plays chess	play(c, chess)
chess is played by charlie	play(c, chess)
girl plays chess	play(h, chess) ∨ play(s, chess)
heidi plays game	play(h, chess) ∨ play(h, hide&seek) ∨ play(h, soccer)
heidi plays with toy	play(h, puzzle) ∨ play(h, ball) ∨ play(h, doll)
sophia plays soccer well	play(s, soccer) ∧ manner(play(s), well)
sophia plays with ball in street	play(s, ball) ∧ place(s, street)
someone plays with doll	play(c, doll) ∨ play(h, doll) ∨ play(s, doll)
doll is played with	play(c, doll) ∨ play(h, doll) ∨ play(s, doll)
charlie plays	play(c, chess) ∨ play(c, hide&seek) ∨ play(c, soccer) ∨ play(c, puzzle) ∨ play(c, ball) ∨ play(c, doll)

> Propositional logic semantics are then translated into situation vectors

Frank et al. (2009), *Cognition*

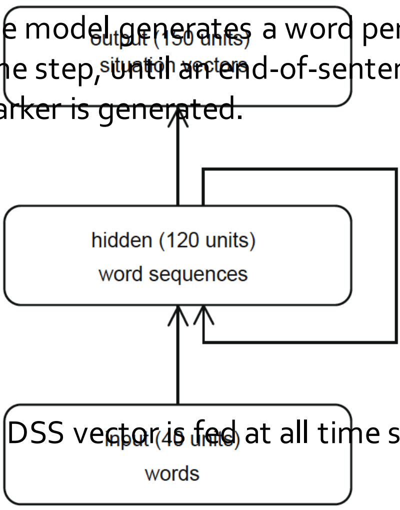
Microlanguage

- 40 original words
  - *+ 2 determiners and end-of-sentence marker → 43 words in our model.*
- Grammar generates 13556 sentences, but only 8201 are lawful according to the microworld.
- Out of the 8201 sentences:
  - 6782 in active voice
  - 1419 in passive voice
- 782 unique DSS representations:
  - 424 related to active and passive sentences
  - 358 related only to active sentences

The grammar defined by Frank et al. (2009) does not define passive sentences for situations where the object of the action is either a person ("Heidi beats Charlie.") or undefined ("Charlie plays.").

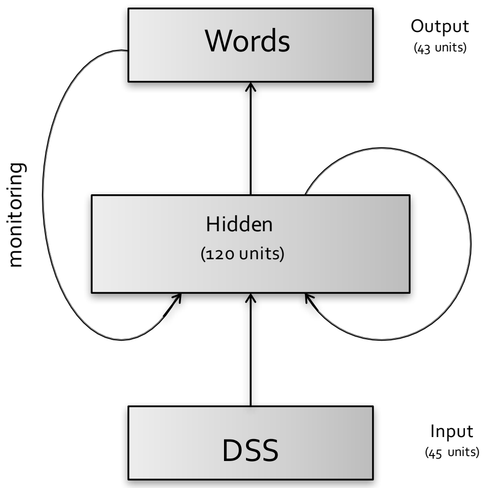
Model Architecture

The model generates a word per time step, until an end-of-sentence marker is generated.



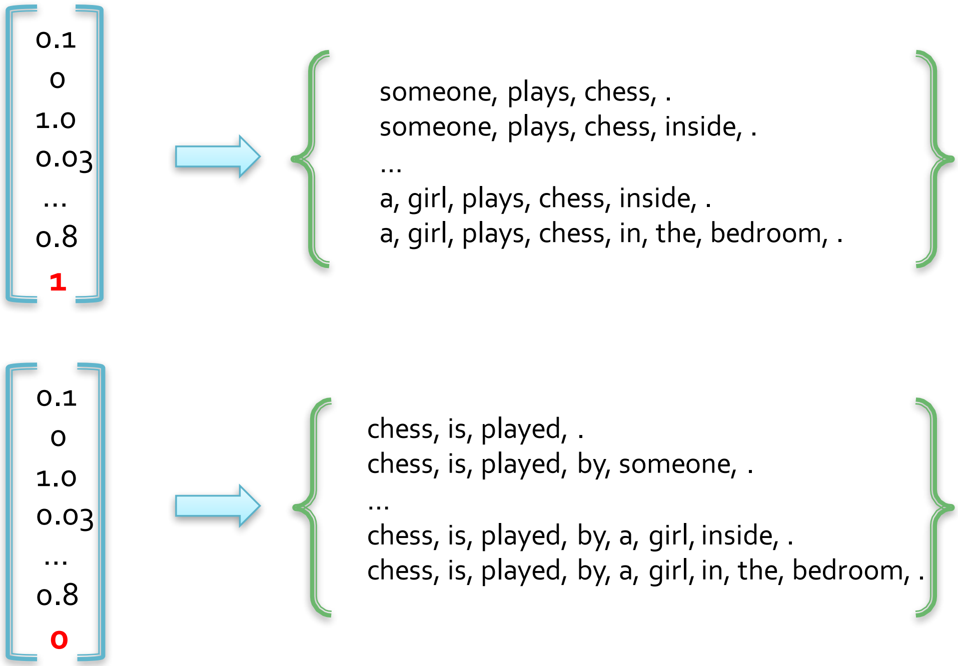
The DSS vector is fed at all time steps.

Comprehension Model
Frank et al. (2009)

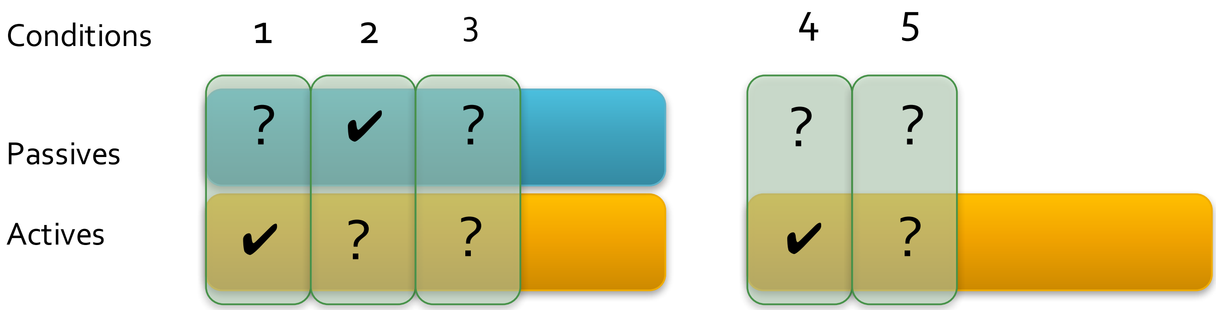


Production Model

Examples Set



Testing Conditions



setAP
(424 situations)

setA
(358 situations)

Training/Testing

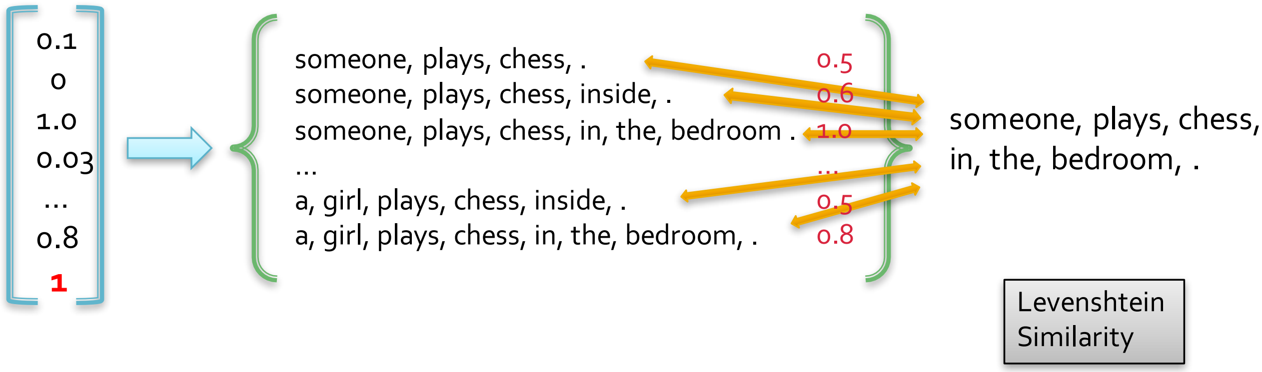
- 10-fold cross validation
  - 90% for training (714 situations), 10 % for testing (70 situations).

  - Each fold's testing set was further divided into the conditions.
    - 14 situations per condition/per fold.

Training Procedure

- Cross-Entropy Backpropagation (Rumelhart, Hinton & Williams, 1986).
- Weight updates after each word.
- Weight initialization with random values drawn from N(0,0.1).
- Bias units weights initialized to zeros.
- At time t, monitoring units were set to what the model was supposed to produce at t-1 given the training item.
- Initial learning rate of 0.124 which has halved each time there was no improvement of performance on the training set during 15 epochs.
- Training halted after 200 epochs or if there was no performance improvement on the training set over a 40-epoch interval.

Sentence Level Evaluation



$$sim(s_1, s_2) = 1 - \frac{distance(s_1, s_2)}{max(length(s_1), length(s_2))}$$

$$sim(\hat{s}_i) = \max_{s \in \varphi_i} sim(\hat{s}_i, s)$$

Results

Condition	Query	Similarity (%)	Perfect Match (%)
train	-	99.43	98.23
1	pas	97.66	92.86
2	act	97.58	93.57
3	act	98.35	93.57
3	pas	96.79	83.57
5	act	95.08	85.0
Average Test	-	**97.1**	**88.57**

*10-fold cross validation averages

Qualitative Analysis

- With a couple of exceptions, all sentences are **syntactically correct and semantically felicitous.**

- Mistakes occur when the model produces a sentence that is **semantically highly similar** to the one expected.

The errors of 5 folds were manually inspected (38 errors).

39.9%	underspecification
23.5%	overspecification
31.6%	very highly similar situations (pp-attach)

	Output	Expected	
1	Sophia beats Heidi with ease at hide_and_seek.	Sophia beats Heidi with ease at hide_and_seek **in the bedroom**.	
2	**a girl** plays with a doll inside.	**Heidi** plays with a doll inside.	undersp.
3	Charlie plays **a game** in the street.	Charlie plays in the street.	
4	Sophia wins with ease at a game **in the street**.	Sophia wins with ease at a game **outside**.	oversp.
5	Sophia beats someone at hide_and_seek in the bedroom.	someone loses to Sophia at hide_and_seek in the bedroom.	
6	someone wins in the bedroom at hide_and_seek.	someone loses in the bedroom at hide_and_seek.	PP-attach

Conditions 4-5 (passives?)

Output of 3 folds was manually inspected (84 situations).

- Mostly correct and coherent with the given semantics.
- Model learns that:
  - passive sentences begin by the object of the action.
  - the object is never a person.

	Passive Output	Active Sentence
1	**hide_and_seek** is won with ease by **Heidi** in the playground.	**Heidi** beats Sophia with ease in the playground at **hide_and_seek**.
2	**a game** is won with ease by **Sophia**.	**Sophia** beats Charlie with ease.
3	**a toy** is played with.	someone plays.
4	**a toy** is played with in the playground by Sophia.	Sophia plays in the playground.
5	a game is lost with difficulty by Charlie.	a girl beats Charlie with difficulty **in the street**.
6	chess is lost by **Heidi** in the bedroom.	the boy loses to **Heidi** at chess in the bedroom.

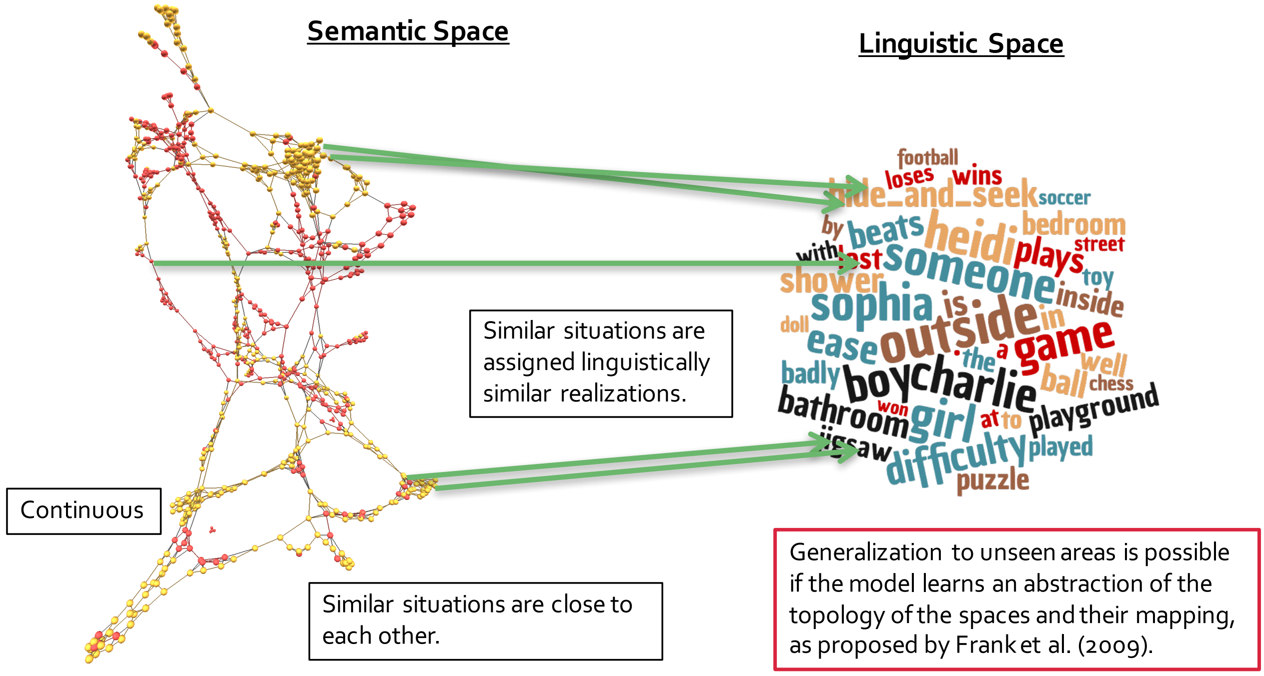
win/lose 92.9%

u. object 7.1%

From the results...

- Model is able to **describe situations** for which it has **no experience**, while being as informative as possible.
- **Not** just memorization. In all test conditions the model is prompted to generate **novel sentences**.
- Only difficulty: **highly similar situations**.
- Even for those, the output is largely **correct**.

Discussion



Discussion

- Cond. 1 & 2:
  - the model is able to generate novel sentences for semantically known situations but with a different voice, showing syntactic systematicity.
- Cond. 3 & 5:
  - the model is able to generate sentences for unseen areas in the semantic space, showing semantic systematicity.
- Cond. 4 & 5:
  - the model is able to produce coherent sentences even if the grammar that was used to generate the train/test sets does not associate passive constructions with these situations.

Conclusion

- The overall high performance of the production model shows that the representations described by Frank et al. (2009) are **suitable to model language production**.

- The model can generate alternative unseen encodings (active/passive) for a particular semantics, showing
  - **syntactic systematicity**.

- Furthermore, the model can generate novel sentences for previously unseen situations, showing
  - **semantic systematicity**.