Miriam Schulz

PhD position interview for SFB 1102 – Project C3

12 July 2022

Contents

Academic background

- M.A. in **Theoretical and Experimental Linguistics**, Université Paris Diderot
  - Courses: syntax (e.g. HPSG), semantics (e.g. predicate logic; event-based semantics), pragmatics, language acquisition, *etc.*
  - Master thesis: A Rational Speech Act model of cross-linguistic pronoun resolution preferences (Prof Dr. Barbara Hemforth, Dr. Heather Burnett)
- Summer schools in 2016: *ESSLLI* and *Trieste Encounters on Cognitive Science*
  - Intro to distributional semantics, CCG, Information Theory
- M.Sc in **Language Science and Technology**, Saarland University
  - Courses: computational psycholinguistics, connectionist language processing, neural networks, statistical natural language processing, statistics, *etc.*
  - Current master's thesis: a Self-Paced-Reading study of forward and backward inference generation

A Rational Speech Act model of cross-linguistic pronoun resolution preferences (Master's thesis 1)

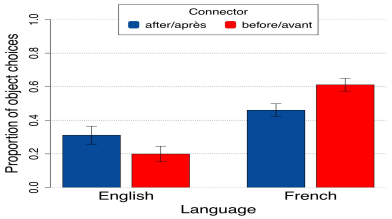
Different **pronoun resolution preferences** for **ambiguous pronouns** across languages:

- ➔ Subject preference in English + German
- ➔ Object preference in French

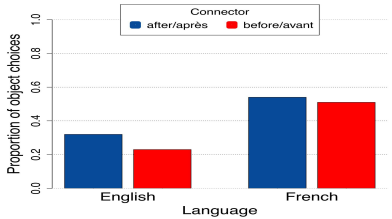
A **Gricean reasoning + frequency-based** approach:

1. **Corpus study**: collect the **frequency of an alternative construction** with obligatory subject co-reference
2. Integrate corpus study results into an **RSA model of cross-linguistic pronoun resolution** as **language- and construction-specific costs**
3. Compare model predictions with **new experimental data**

Experimental data



Model predictions

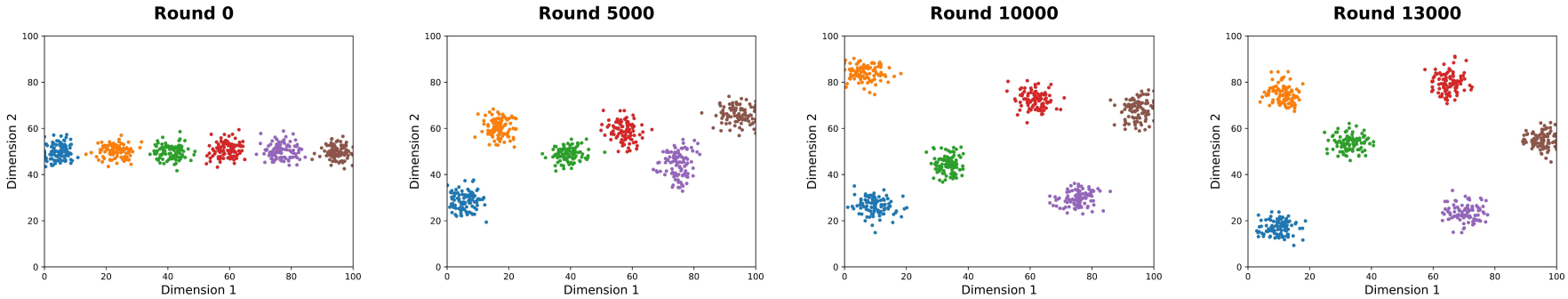


Schulz, Burnett & Hemforth (2019; 2021)

A multi-level exemplar model of the evolution of sound-category systems

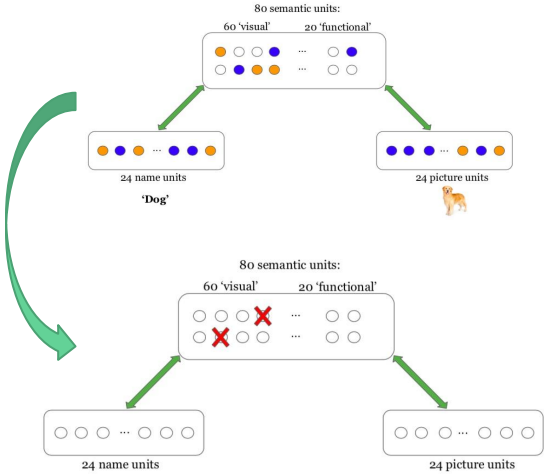
- Term paper for the seminar *Exemplar Theory* by Prof. Dr. Bernd Möbius
- A Python implementation of the **exemplar model of variation & evolution** in the sound system by **Winter & Wedel (2016)**
- **Hidden (sublexical) variation** creates pathways for future change of the linguistic system

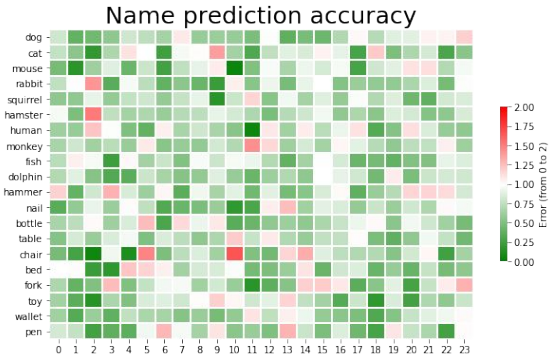
Evolution of the sound-category system over 13000 dialogue turns in the model:

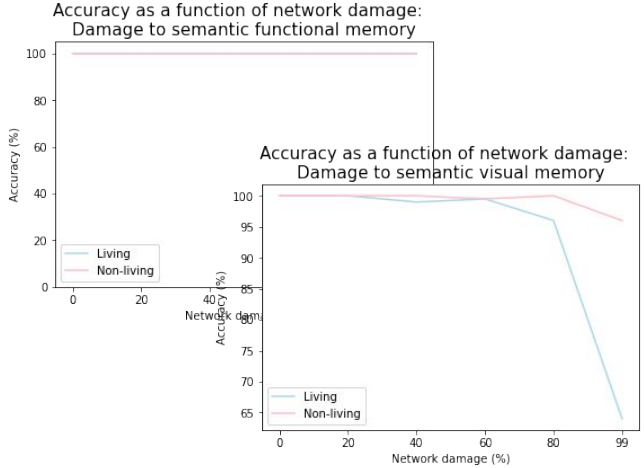


A simplified re-implementation of a connectionist model of semantic memory impairment

- ➔ Course presentation+project for the course *Cognitive Neuropsychology* by Prof. Dr. Axel Mecklinger
- ➔ Simplified reimplementation of the **Farah & McClelland (1991) model** in **PyTorch**
- ➔ Brain lesions simulated by **setting connection weights to zero** in visual *vs.* functional semantic memory







German noun frequency tool for stimuli creation

- ➔ A python program that lets the user search through a list of German nouns sorted by frequency (constructed from the deWaC corpus), controlling for gender, case, number
- ➔ https://github.com/miriamschulz/German_noun_frequency_tool

```
GERMAN NOUN FREQUENCY TOOL

Please enter either
- a noun (e.g. Haus), or
- a search frequency (int or float, e.g. 5.5)
Your input: Regenschirm

Analysis of the input noun 'Regenschirm':
        Frequency rank:         0.81 per million
        Word length:            11 characters
        Possible genders:       masc
        Possible cases:         nom, acc, dat
        Possible numerus:       sing

Automatically defined search criteria:
        * Search frequency:     0 to 1.81 per million
        * Word length:          9 to 13 characters
        * Gender(s):            masc
        * Case(s):              acc, dat
        * Numerus:              sing

Press 'c' to change these criteria, otherwise press Enter.


Searching for nouns...

Found the following 71393 nouns with similar frequency:
```

NOUN	FREQUENCY	GENDERS	CASES	NUMERUS
Tagungsband	0.81	masc	acc/dat	sing
Maastrichter	0.81	masc	acc/dat	sing
Automatische	0.81	masc	acc/dat	sing
Langfristige	0.81	masc	acc	sing
Küchentisch	0.81	masc	acc/dat	sing
Datenschützer	0.81	masc	acc/dat	sing
Weinstock	0.81	masc	acc/dat	sing
Weißensee	0.81	masc	acc/dat	sing
Schlachter	0.81	masc	acc/dat	sing
Kriegszustand	0.81	masc	acc/dat	sing
Terrorakte	0.81	masc	acc/dat	sing
Mitstreitern	0.81	masc	acc/dat	sing
Historismus	0.81	masc	acc/dat	sing
Eigenbetrieb	0.81	masc	acc/dat	sing
Waldbesitzer	0.81	masc	acc/dat	sing
Mutterschutz	0.81	masc	acc/dat	sing
Horrorfilm	0.81	masc	acc/dat	sing

```
Please enter a verb (infinitive) to check for co-occurrence with the retrieved nouns: öffnen

Out of the 71393 search results, 763 nouns can occur with 'öffnen':
```

BIGRAM COUNT	NOUN	FREQUENCY	GENDERS	CASES	NUMERUS
74	Fallschirm	0.98	masc	acc/dat	sing
39	Doppelklick	0.88	masc	acc/dat	sing
33	Muttermund	0.76	masc	acc/dat	sing
20	Brustkorb	1.33	masc	acc/dat	sing
18	Verschluß	0.79	masc	acc/dat	sing
18	Texteditor	0.26	masc	acc/dat	sing
17	Wasserhahn	1.09	masc	acc/dat	sing
16	Radverkehr	0.91	masc	acc/dat	sing
13	Garagentor	0.16	masc	acc/dat	sing
13	Briefumschlag	0.63	masc	acc/dat	sing
12	Sprungtor	0.07	masc	acc/dat	sing
12	Fensterladen	0.09	masc	acc/dat	sing

Current master's thesis: Inference generation during reading

Research questions:

1. Do reading times for the generation of bridging inferences covary with the P600 effect observed by Burkhardt (2007)?
2. [How] do the behavioral correlates of inference generation differ from a failure to achieve coherence?
3. Can the use of a forward inference encouraging task facilitate the production of forward inferences?

Method:

- Two self-paced reading studies
- SPR1: bridging (backward) inferences
- SPR2: predictive (forward) inferences (= SPR1 with a different task)

Summary: Research interests

- ➔ Going beyond the *what* towards the *why* of linguistic structure and language use:
  - ◆ Which **cognitive factors and biases** contribute to shape human language?
  - ◆ How can the **human brain** process language?
  - ◆ Which role do **extra-linguistic context, general world knowledge** as well as **prediction** play in language processing?
- ➔ Opportunity to zoom in on a single problem in depth during the PhD
- ➔ Integration of Surprisal Theory, psycholinguistic research, and cognitive modeling in Project C3

References

Burkhardt, Petra. 2007. The P600 reflects cost of new information in discourse memory. *Neuroreport* 18, no. 17: 1851-1854.

McClelland, James L., and M. J. Farah. 1991. A Computational Model of Semantic Memory Impairment: Modality Specificity and Emergent Category Specificity. *Journal of Experimental Psychology*.

Schulz, Miriam, Heather Burnett & Barbara Hemforth. 2019. A Rational Speech Act model of cross-linguistic differences in pronoun resolution preferences. Poster presented at *CUNY Conference on Human Sentence Processing*, Colorado Boulder. DOI: https://doi.org/10.13140/RG.2.2.24487.80805

Schulz, Miriam, Heather Burnett & Barbara Hemforth. 2021. Corpus, experimental and modeling investigations of cross-linguistic differences in pronoun resolution preferences. *Glossa: a journal of general linguistics* 6(1): 66. DOI: https://doi.org/10.5334/gjgl.1142

Winter, Bodo, and Andrew Wedel. 2016. The Co-evolution of Speech and the Lexicon: The Interaction of Functional Pressures, Redundancy, and Category Variation. *Topics in cognitive science* 8, no. 2: 503-513

Thank you