Gendered language and the mental lexicon: Computational insights

Dr Dominic Schmitz
English Language and Linguistics
Heinrich Heine University Düsseldorf

dominic.schmitz@hhu.de
dmncschmtz.bsky.social

What this talk is about

contexts

What can context information reveal about…

semantics

What role do (very similar) meanings play in…

discriminative learning

What does discriminative learning tell us about the…

acoustics of masculines

What connects the acoustic of masculines to their contexts?

Who do you picture?

Künstler

‘artist’

- a male artist?
- a female artist?
- a nonbinary or genderqueer artist?

Who do you picture?

Darin sind jene frühen Arbeiten des **Künstlers** zu sehen, die er lange vor seinem Aufstieg schuf.

‘It features the **artist's** early works, created long before his rise to fame.’

- a male artist?
- a female artist?
- a nonbinary or genderqueer artist?

Who do you picture?

Heute geht es um die Verkehrslage und eine tolle **Künstlerin** im Konzerthaus.

‘Today we're talking about the traffic situation and a great **artist** at the concert hall.’

- a male artist?
- a female artist?
- a nonbinary or genderqueer artist?

Who do you picture?

Für jeden Song fallen 0,4 Cent für den **Künstler** ab,
mit der Zeit läppert es sich.

'For each song, 0.4 cents go to the **artist**, and over time it adds up.'

- a male artist?
- a female artist?
- a nonbinary or genderqueer artist?

German role nouns

- most German role nouns come in two flavours: **masculine** and **feminine**

  Darin sind jene frühen Arbeiten des **Künstlers** zu sehen, …

  'It features the artist's early works, …'

  Heute geht es um eine tolle **Künstlerin** im Konzerthaus.

  'Today we're talking about a great artist at the concert hall.'

- masculine role nouns show two distinct usage cases
  - **specific**: referring to male individuals
  - **generic**: referring to individuals of any/unknown/irrelevant gender

    Für jeden Song fallen bei 0,4 Cent für den **Künstler** ab.

    'For each song, 0.4 cents go to the artist.'

  - specific and generic masculines are **identical in form**: *Künstler*

The role of context

- even though *Künstler*$_{\text{specific}}$ and *Künstler*$_{\text{generic}}$ share their form, you were rather certain who you're thinking of, at least in the specific case
- how can this be?
- the answer: context
- context may reveal whether a masculine role noun is intended to be specific or generic
- apparently, context is rather important for the meaning of masculine role nouns!

Distributional semantics

- meaning is reflected in patterns of use
- Harris (1954)

  "Words that occur in similar contexts tend to have similar meanings"

- Firth (1957)

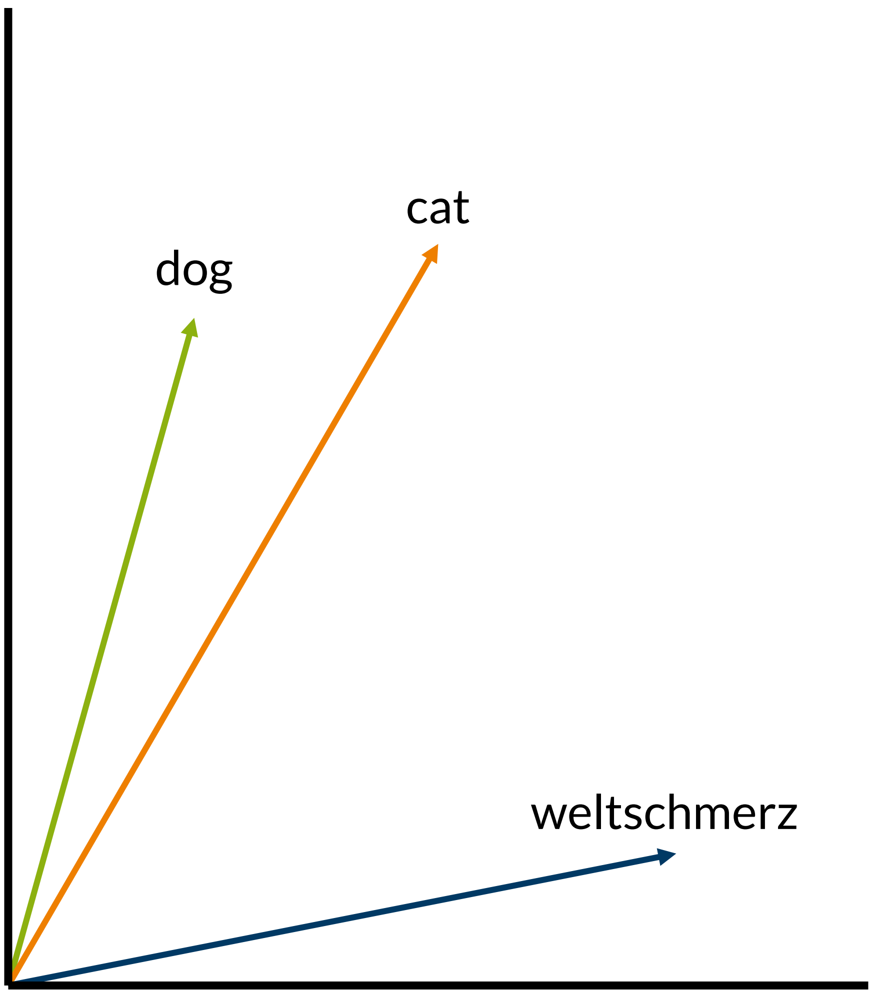
  "You shall know a word by the company it keeps"

- distributional models reverse-engineer meaning by tracking contextual regularities in large corpora
- output: high-dimensional semantic vectors representing patterns of linguistic behaviour

Boleda (2020)

Vector space models of meaning

- each word becomes a point in a multidimensional semantic space
- dimensions capture statistical properties of contexts (often not interpretable individually)
- semantic similarity corresponds to geometric closeness (e.g. cosine similarity)
- distances encode graded, continuous semantic relatedness



Boleda (2020)

What counts as “context”?

- context can mean many things
  - words in a window (±*n* words)
  - sentence, paragraph, document
  - syntactic relations
  - multimodal contexts (e.g. visual information)
- different models operationalise context differently, leading to different semantic spaces
- key idea: semantically similar words have overlapping context distributions

Boleda (2020)

Defining context

The apple is the pomaceous fruit of the apple tree, species Malus domestica in the rose family (Rosaceae). […] There are more than 7,500 known cultivars of apples, resulting in a range of desired characteristics. […] The apple forms a tree that is small and deciduous, reaching 3 to 12 metres (9.8 to 39 ft) tall, with a broad, often densely twiggy crown. […] The apple tree was perhaps the earliest tree to be cultivated, and its fruits have been improved through selection over thousands of years. […]

Thater (2011)

Defining context

The **apple** is the pomaceous fruit of the apple tree, species Malus domestica in the rose family (Rosaceae). […] There are more than 7,500 known cultivars of apples, resulting in a range of desired characteristics. […] The apple forms a tree that is small and deciduous, reaching 3 to 12 metres (9.8 to 39 ft) tall, with a broad, often densely twiggy crown. […] The apple tree was perhaps the earliest tree to be cultivated, and its fruits have been improved through selection over thousands of years. […]

Thater (2011)

Defining context

The apple is the pomaceous fruit of the **apple** tree, species Malus domestica in the rose family (Rosaceae). […] There are more than 7,500 known cultivars of apples, resulting in a range of desired characteristics. […] The apple forms a tree that is small and deciduous, reaching 3 to 12 metres (9.8 to 39 ft) tall, with a broad, often densely twiggy crown. […] The apple tree was perhaps the earliest tree to be cultivated, and its fruits have been improved through selection over thousands of years. […]

Thater (2011)

Defining context

The apple is the pomaceous fruit of the apple tree, species Malus domestica in the rose family (Rosaceae). […] There are more **than 7,500 known cultivars of apples, resulting in a range of** desired characteristics. […] The apple forms a tree that is small and deciduous, reaching 3 to 12 metres (9.8 to 39 ft) tall, with a broad, often densely twiggy crown. […] The apple tree was perhaps the earliest tree to be cultivated, and its fruits have been improved through selection over thousands of years. […]

Thater (2011)

Defining context

The apple is the pomaceous fruit of the apple tree, species Malus domestica in the rose family (Rosaceae). […] There are more than 7,500 known cultivars of apples, resulting in a range of desired characteristics. […] The apple forms a tree that is small and deciduous, reaching 3 to 12 metres (9.8 to 39 ft) tall, with a broad, often densely twiggy crown. […] The apple tree was perhaps the earliest tree to be cultivated,
selection o

	tree	fruit	forms	perhaps	apple	…
…						
apple	3	2	1	1	0	…
tree	0	1	1	1	3	…
…						

Thater (2011)

Defining context

The apple is the pomaceous fruit of the apple tree, species Malus domestica in the rose family (Rosaceae). […] There are more than 7,500 known cultivars of apples, resulting in a range of … forms a tree that is … etres (9.8 to 39 ft) … y crown. […] The apple tre… perhaps the earliest tree to be cultivated, … selection o…

rows
the vocabulary we are interested in

columns
"features" of the vocabulary, i.e. semantic dimensions

	tree	fruit	forms	perhaps	apple	…
…						
apple	3	2	1	1	0	…
tree	0	1	1	1	3	…
…						

Thater (2011)

From co-occurrence to vectors

- start with a large vocabulary and context inventory
- build a matrix: rows = target words, columns = contextual features
- fill the matrix with counts or derived statistics
- each row → a word vector, e.g. $v_{apple} = \langle 3, 2, 1, 1, 0, \ldots \rangle$

	tree	**fruit**	**forms**	**perhaps**	**apple**	**...**
...						
apple	3	2	1	1	0	...
tree	0	1	1	1	3	...
...						

Thater (2011)

Design choices in distributional models 1

Pre-processing

- do we use word-forms (*teachers*, *teacher*, *Teacher*) or simplified forms (teacher)?
- do we keep function words (*the*, *of*, *to*) or remove them?
- how do we treat punctuation, compounds, or multi-word expressions?

Context

- a fixed window (e.g. the five words around the target)?
- whole sentences or paragraphs?
- grammatically defined relations (e.g. subject–verb–object links)?

Boleda (2020)

Design choices in distributional models 2

Associative strength

- some methods simply count co-occurrences
- others give more weight to informative contexts (rare but meaningful associations)
- some methods learn these weights automatically

Model type

- **count-based models**
  build a large co-occurrence table and transform it
- **predictive models**
  learn vectors by predicting missing words from context
- **subword models**
  include information from letter sequences to handle morphological richness

Boleda (2020)

Predictive models: *word2vec*

- instead of counting co-occurrences, predictive models learn meaning by **guessing** words
- the model sees sentences and repeatedly asks itself questions like “Given this context, which word is likely to appear here?”
- to guess well, the model must learn which words occur in similar situations
- words with similar contexts end up having similar vectors

The ______ grows on trees.

Mikolov et al. (2013)

Subword models: *fastText*

- many languages have rich morphology: *Student*, *Studentin*, *Studierende*, …
- traditional models treat each word as unrelated, even if they clearly share meaning
- *fastText* improves this by breaking words into small letter chunks
- these 'subword' pieces also get vectors, and a word's meaning is built from these pieces, e.g. $v_{student\_n3} = \langle stu, tud, ude, den, ent \rangle$
- result
  - better handling of rare (and even novel) forms
  - better handling of inflected words
  - more realistic similarity between related forms

Bojanowski et al. (2016)

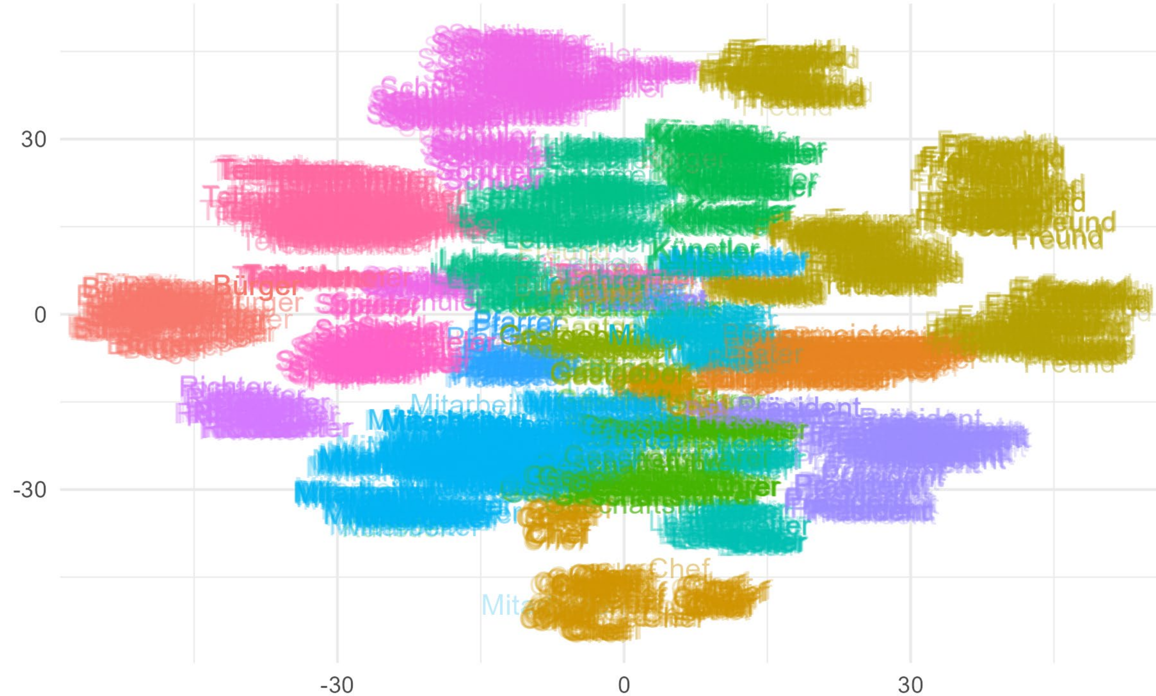
Subword models: *fastText*

- example: *Student*
  - $v_{student\_n2} = \langle st, tu, ud, de, en, nt \rangle$
  - $v_{student\_n3} = \langle stu, tud, ude, den, ent \rangle$
  - $v_{student\_n4} = \langle stud, tude, uden, dent \rangle$
  - $v_{student\_n5} = \langle stude, tuden, udent \rangle$
  - $v_{student\_n6} = \langle studen, tudent \rangle$
  - $v_{student\_full} = \langle student \rangle$
  - $v_{student} = \left\langle \begin{matrix} v_{student\_n2} + v_{student\_n3} + v_{student\_n4} + v_{student\_n5} + v_{student\_n6} + \\ v_{student\_full} \end{matrix} \right\rangle$

Bojanowski et al. (2016)

The final semantic space

- t-Distributed Stochastic Neighbour Embedding
- aim: reduce high-dimensional vectors to 2 or 3 dimensions without loosing local patterns or structure between the vectors



van der Maaten & Hinton (2008)

How vector similarity is measured

- to compare two words, we measure the angle between their vectors

  = **cosine similarity**

  - small angle → vectors point in similar directions → high similarity
  - large angle → vectors diverge → low similarity

- the resulting values range from:
  - +1 (= identical direction)
  - 0 (= unrelated)
  - −1 (= opposite directions)

Boleda (2020)

Guess the similarity!

	surgeon	empathy
nurse		

	classroom	male
teacher		

	newspaper	man
editor		

	queen	castle
king		

remember: 1 = identical meaning; 0 = unrelated meaning; -1 = opposites

Guess the similarity!

	surgeon	empathy
nurse	0.52	0.23

	classroom	male
teacher		

	newspaper	man
editor		

	queen	castle
king		

remember: 1 = identical meaning; 0 = unrelated meaning; -1 = opposites

Guess the similarity!

	surgeon	**empathy**
nurse	0.52	0.23

	classroom	**male**
teacher	0.61	0.22

	newspaper	**man**
editor		

	queen	**castle**
king		

remember: 1 = identical meaning; 0 = unrelated meaning; -1 = opposites

Guess the similarity!

	surgeon	empathy
nurse	0.52	0.23

	classroom	male
teacher	0.61	0.22

	newspaper	man
editor	0.50	0.22

	queen	castle
king		

remember: 1 = identical meaning; 0 = unrelated meaning; -1 = opposites

Guess the similarity!

	surgeon	empathy
nurse	0.52	0.23

	classroom	male
teacher	0.61	0.22

	newspaper	man
editor	0.50	0.22

	queen	castle
king	0.71	0.38

remember: 1 = identical meaning; 0 = unrelated meaning; -1 = opposites

Back to role nouns

- *Künstler*$_{\text{specific}}$ and *Künstler*$_{\text{generic}}$ share their form, but apparently context can help us disambiguate them
- so, in theory, the differences in contexts should also be represented by pertinent semantic vectors - because semantic vectors make use of contextual information
- let’s see whether this is actually the case…

Oh no!

- before we can look into this, we first need to solve a problem
- German masculine role nouns pose a methodological issue:

  one form - two meanings
- *fastText* doesn’t know this, it only sees the one spelling *Künstler*
- so, both meanings collapse into one vector $v_{Künstler}$
- with one vector for two meanings, we cannot do anything meaningful…

Schmitz (2024)

A solution: instance vectors

- following Lapesa et al. (2018) we can compute instance vectors
  - instead of one vector per word type, compute one vector per token based on the actual words around it
- instance vector computation
  1. take a target token (e.g. *Arbeiter*)
  2. take the *n* content words before and after it
  3. average their *fastText* vectors
  4. the result = an instance vector capturing the meaning in this sentence
- instance vectors are thus a method of contextual semantic disambiguation that remains purely distributional, without resorting to grammars or lexicons

Schmitz (2024)

Study: instance vectors and role nouns

1. **corpus**

   30,000 manually annotated attestations of role nouns (generic vs specific use)

2. **target paradigms**

   76 role nouns from Gabriel et al. (2008)

3. **context vectors**

   pre-trained German *fastText* vectors (subword-based)

4. **instance vector computation**

   for each target token, compute one instance vector at window sizes $n = 2, 5, 8$

5. **analysis**

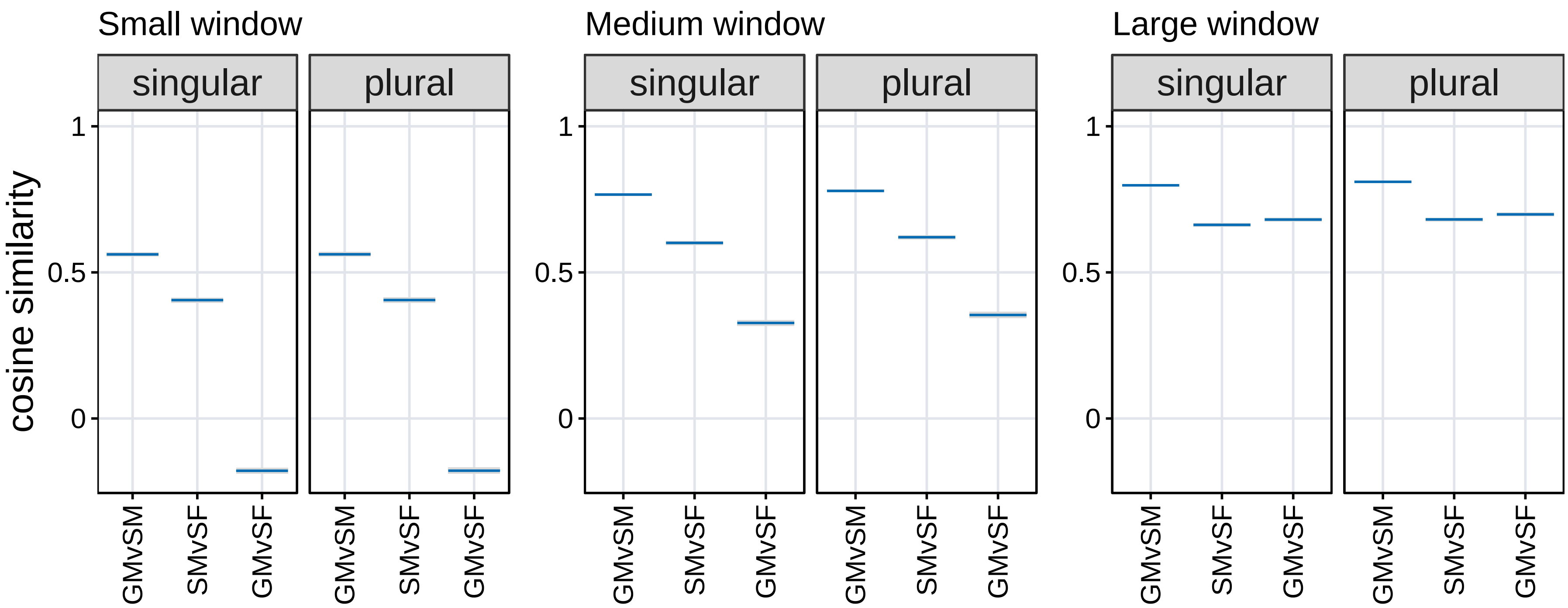
   compute cosine similarities within paradigms (generic $\leftrightarrow$ specific masc;

   generic masc $\leftrightarrow$ specific fem; specific masc $\leftrightarrow$ specific fem)

   model using beta-regression

Schmitz (2024)

Study: instance vectors and role nouns



- across number and window sizes, generic masculines are always more similar to specific masculines than to specific feminines

Schmitz (2024)

Study: instance vectors and role nouns

- a single *fastText* vector cannot solve the generic/specific ambiguity
- but instance vectors can
- instance vectors allow us to
  - treat each token as its own semantic event,
  - recover the contextual meaning of *Arbeiter* as used in that sentence,
  - and measure similarity patterns that reveal systematic male bias in generic uses
- …but is any of this cognitively plausible?

Schmitz (2024)

From distributional semantics to discriminative learning

- distributional models capture semantic similarity through patterns of co-occurrence
- they give us representations, but they are not motivated by cognitive insights or theories
- discriminative learning models address exactly this
- they implement error-driven learning, cue competition, and discriminability, providing a cognitive model of how lexical knowledge emerges from usage patterns

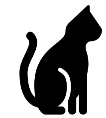
Chuang & Baayen (2021)

Discriminative learning: the basic idea

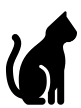
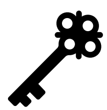
- learning is **error-driven**: what is learned is the difference between predicted and observed outcomes
- **cues** compete to predict **outcomes**
- correct predictions strengthen cue → outcome links;
  incorrect predictions weaken cue → outcome links
- learning is incremental, usage-based, and continuous across the lifespan
- lexical knowledge emerges from these learned mappings
- **naive discriminative learning** (NDL) follows the Rescorla–Wagner model of associative learning (Rescorla & Wagner, 1972)

Baayen et al. (2011); Chuang & Baayen (2021)

The learning mechanism: error-driven updating

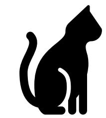
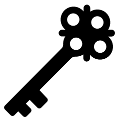
- imagine the model is trying to learn that the sound /kæt/ means ( not  

Step 1: Look at the cues in the input

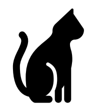
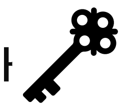
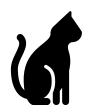
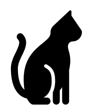
- When the model hears /kæt/, it identifies the cues: /k/, /æ/, /t/
- These cues each have weights pointing to many possible meanings ( ,   ..)

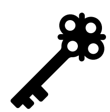
Chuang & Baayen (2021)

The learning mechanism: error-driven updating

- imagine the model is trying to learn that the sound /kæt/ means  not  

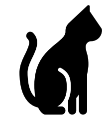
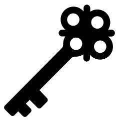
Step 2: Add up the current evidence for each meaning

- The model adds the cue → meaning weights
  - maybe /k/ gives weak support for  and 
  - /æ/ gives strong support for 
  - /t/ gives modest support for 
- This creates the model's **prediction**:

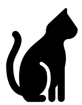
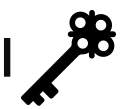
   might get a medium score, but  a low score, etc.

Chuang & Baayen (2021)

The learning mechanism: error-driven updating

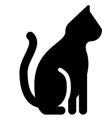
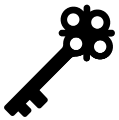
- imagine the model is trying to learn that the sound /kæt/ means  not  

Step 3: Compare prediction with the actual outcome

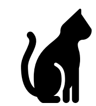
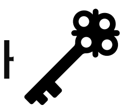
- Reality is:  **= correct**, all others = incorrect
- If the model did **not predict**  **strongly enough**, this is an error
- If the model gave **too much support to**  that is also an error

Chuang & Baayen (2021)

The learning mechanism: error-driven updating

- imagine the model is trying to learn that the sound /kæt/ means  not  

Step 4: Adjust the weights to reduce this error next time

- Strengthen the weights from the present cues (/k/, /æ/, /t/) **towards C**  because  should have been predicted more strongly
- Weaken the weights from these same cues **towards**   because the model wrongly predicted them
- After many such steps, the cue → meaning weights come to reflect the statistical structure of the input

Chuang & Baayen (2021)

NDL as a model of semantic structure

- each outcome has a vector of incoming cue weights
- if we specify outcomes and cues to be words, we compute word embeddings
- each embedding reflects the cues that reliably predict the meaning of its outcome
- outcomes with similar patterns of predictive cues are semantically similar
- NDL thus provides a purely usage-based semantic space aligned with psychological learning principles

Baayen et al. (2011); Chuang & Baayen (2021)

Why NDL is useful beyond theory

- **transparent**: mathematically equivalent to linear regression
- **flexible** cue choices: orthography, phonology, morphology, syntax, prosody, social meaning signals
- **no morpheme segmentation or rule system required** — structure emerges from data
- **connects semantics and processing**
  - predicts activation, competition, confusability, semantic neighbourhood effects
  - can feed into behavioural models (RTs, choices, acoustic durations)

Baayen et al. (2011); Chuang & Baayen (2021)

Back to role nouns (again)

- *Künstler*$_{\text{specific}}$ and *Künstler*$_{\text{generic}}$ share their form, but apparently context can help us disambiguate them
- so, in theory, the differences in contexts should also be represented by pertinent semantic vectors - because semantic vectors make use of contextual information
- let's see whether this is actually the case…
- NDL lets us treat the features of these forms as distinct cues/outcomes with separate semantic representations
- by learning from real corpus data, the model reconstructs how the two meanings differ in their contexts

Schmitz et al. (2023)

Study: NDL and role nouns

1. **corpus**

   30,000 manually annotated attestations of role nouns (generic vs specific use)

2. **target paradigms**

   76 role nouns from Gabriel et al. (2008)

3. **NDL setup**

   - cues and outcomes: all content words in a sentence, reduced to base form, and grammatical gender, number, generic/specific
   - each sentence is a learning event; all cues → each outcome present in that sentence

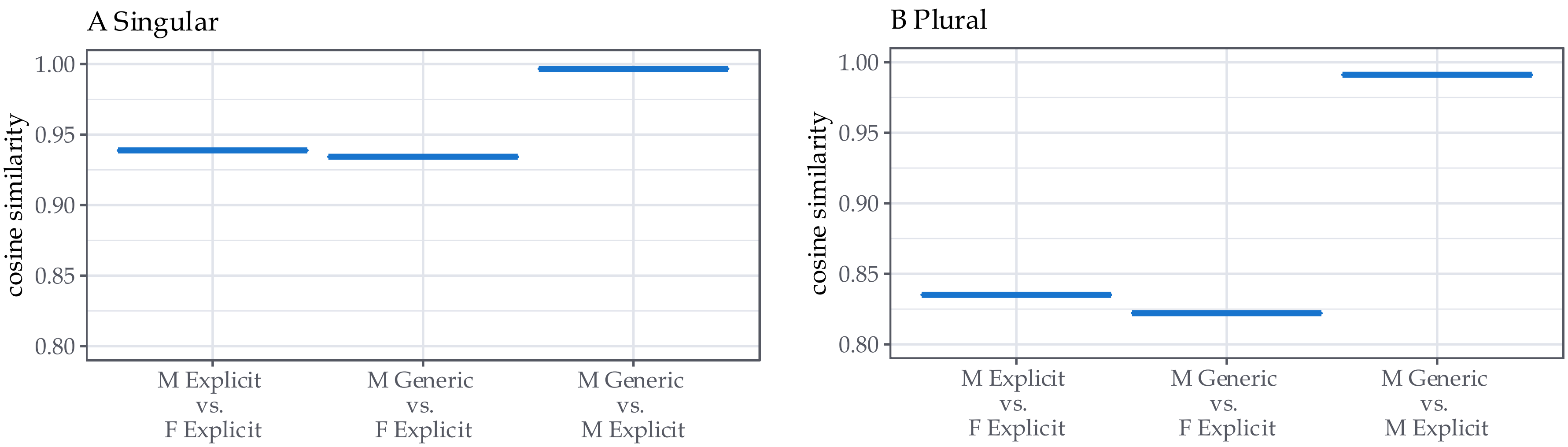
4. **vectors**

   - sum of parts vectors for target words, e.g.

$$v_{Arbeiter\_masc\_sg\_g} = v_{Arbeiter} + v_{masculine} + v_{singular} + v_{generic}$$

Schmitz et al. (2023)

Study: NDL and role nouns



- across number, generic masculines are more similar to specific masculines than to specific feminines
- …and what about the mental lexicon?

Schmitz et al. (2023)

From learned semantics to lexical processing

- NDL and other algorithms of distributional semantics give us semantic embeddings: vectors describing how cues relate to outcomes
- but language processing involves (at least) two mappings
  - comprehension: form → meaning
  - production: meaning → form
- **linear discriminative learning** (LDL) models both mappings directly
- LDL uses the same learning principles as NDL but extends them to high-dimensional mappings between form and meaning

Baayen et al. (2019); Chuang & Baayen (2021)

What is LDL?

a linear learning model in which **form vectors** and **meaning vectors** are linked through **linear mappings** that are learned from experience

Baayen et al. (2019); Chuang & Baayen (2021)

What is LDL?

a linear learning model in which **form vectors** and **meaning vectors** are linked through **linear mappings** that are learned from experience

Baayen et al. (2019); Chuang & Baayen (2021)

Form vectors

- LDL represents a word's form as a **binary vector** encoding which sublexical cues it contains
- standard cues are *n*-grams
- each unique *n*-gram across the lexicon becomes a column in the form matrix C; each word is a row, marked with 1s where its *n*-grams occur
- this avoids assuming phonemes: speech is contextual, gradients matter, and *n*-grams capture this better than discrete phonemes
- C can be built from orthography, phonology, syllables, or even acoustic vectors
- because only a few cues are present per word, C is a sparse matrix, optimised for efficient computation

Baayen et al. (2019); Chuang & Baayen (2021)

Form vectors

- example: *Student*, *Studentin*, *Ärztin*, *resistent*

	#st	stu	tud	ude	den	ent	nt#	nti	tin	in#
student	1	1	1	1	1	1	1	0	0	0
studentin	1	1	1	1	1	1	0	1	1	1
ärztin	0	0	0	0	0	0	0	0	1	1
resistent	0	0	0	0	0	1	1	0	0	0

Baayen et al. (2019); Chuang & Baayen (2021)

What is LDL?

a linear learning model in which **form vectors** and **meaning vectors** are linked through **linear mappings** that are learned from experience

Baayen et al. (2019); Chuang & Baayen (2021)

What is LDL?

a linear learning model in which **form vectors** and **meaning vectors** are linked through **linear mappings** that are learned from experience

Baayen et al. (2019); Chuang & Baayen (2021)

Meaning vectors

- meanings are represented by semantic vectors
- vectors come from NDL or other distributional methods

	D1	D2	D3	D4	D5	…
student	0.2	0.4	0.3	0.9	0.8	…
studentin	0.1	0.5	0.2	0.8	0.7	…
ärztin	0.9	0.1	0.1	0.3	0.2	…
resistent	0.5	0.9	0.8	0.4	0.4	…

Baayen et al. (2019); Chuang & Baayen (2021)

What is LDL?

a linear learning model in which **form vectors** and **meaning vectors** are linked through **linear mappings** that are learned from experience

Baayen et al. (2019); Chuang & Baayen (2021)

What is LDL?

a linear learning model in which **form vectors** and **meaning vectors** are linked through **linear mappings** that are learned from experience

Baayen et al. (2019); Chuang & Baayen (2021)

Linear mappings

- linear mappings allow transparent, interpretable learning
- they implement discriminative learning in vector spaces
- efficient enough to scale to full lexicons (tens of thousands of words)
- empirically successful in modelling
  - lexical decision (Chuang et al. 2020),
  - auditory recognition (Arnold et al. 2017),
  - morphological processing (Baayen & Smolka 2020),
  - semantic priming (Baayen & Smolka 2020),
  - subphonemic durational differences (Schmitz et al. 2021),
  - and more

Baayen et al. (2019); Chuang & Baayen (2021)

Linear mappings: comprehension

- the model learns which form features reliably point to which areas of meaning space
- LDL learns a transformation matrix F so that $S = C \cdot F$
- because S and C are high-dimensional, $C \cdot F$ never results in S, but in $\hat{S}$
- $\hat{S}$ is the best approximation to S possible
- $\hat{S}$ reflects the outcome of the comprehension process, i.e. differences between S and $\hat{S}$ represent the doings of the simulated mental lexicon
- based on $\hat{S}$, meaningful measures based on semantics in the mental lexicon can be derived

Baayen et al. (2019); Chuang & Baayen (2021)

Linear mappings: production

- the model learns which pieces of form are most likely for what it wants to say
- LDL learns a transformation matrix G so that $C = S \cdot G$
- because S and C are high-dimensional, $S \cdot G$ never results in C, but in $\hat{C}$
- $\hat{C}$ is the best approximation to C possible
- $\hat{C}$ reflects the outcome of the comprehension process, i.e. differences between C and $\hat{C}$ represent the doings of the simulated mental lexicon
- based on $\hat{C}$, meaningful measures based on forms in the mental lexicon can be derived

Baayen et al. (2019); Chuang & Baayen (2021)

What is LDL?

a linear learning model in which **form vectors** and **meaning vectors** are linked through **linear mappings** that are learned from experience

Baayen et al. (2019); Chuang & Baayen (2021)

What is LDL?

a linear learning model in which **form vectors** and **meaning vectors** are linked through **linear mappings** that are learned from experience

- form vectors in C, meaning vectors in S
- comprehension via F for $\hat{S}$, production via G for $\hat{C}$
- once trained, LDL can
  - map form $\longrightarrow$ meaning
  - map meaning $\longrightarrow$ form
  - compute different measures based on comprehended semantics and produced form

Baayen et al. (2019); Chuang & Baayen (2021)

Study: LDL and role nouns

C matrix

- orthographic trigrams used as form cues
- each word type corresponds to a row; each trigram cue is a column

S matrix

- the NDL semantic embeddings for each word serve as meaning vectors

idea

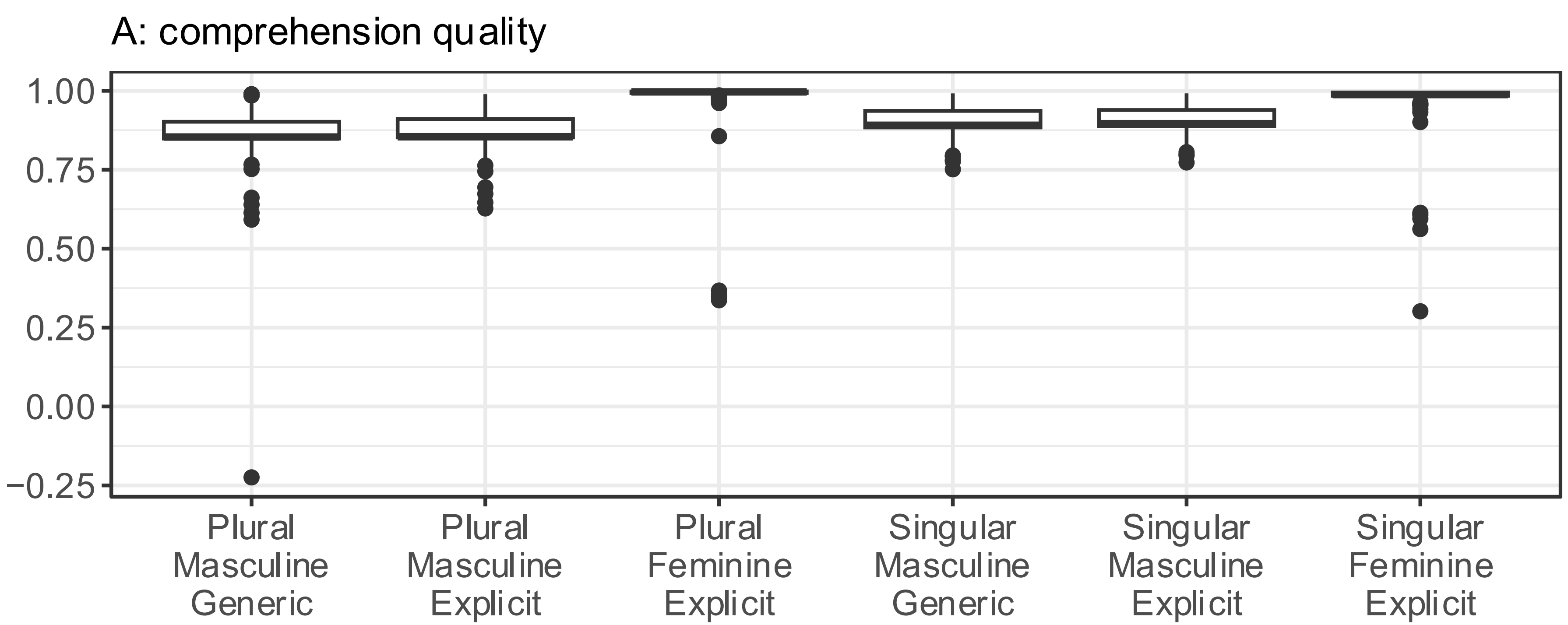
- this time we're not interested in cosine similarities, but in **measures** extracted from the comprehended semantics of our target words - are they different between generic masc, specific masc, and feminines?

Schmitz et al. (2023)

Study: LDL and role nouns

Measure 1: comprehension quality

- how well is the input semantic vector comprehended?
  = correlation of input vector and comprehended vector



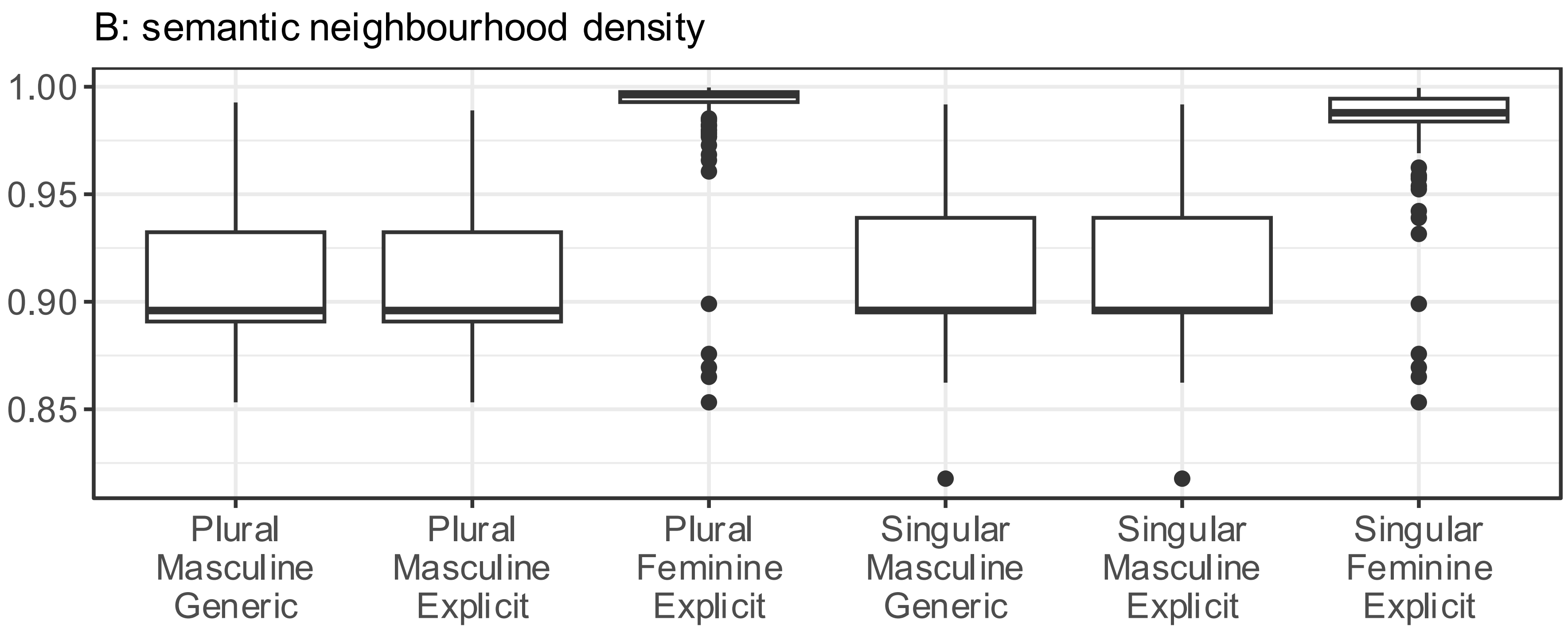
Schmitz et al. (2023)

Study: LDL and role nouns

Measure 2: semantic neighbourhood density

- how dense is the semantic neighbourhood of a target?

  = mean correlation of 10 nearest neighbours



Schmitz et al. (2023)

Study: LDL and role nouns

Measure 3: semantic activation diversity

- how strongly are semantic dimensions activated by the target?

  = Euclidean norm of the comprehended semantic vector

Schmitz et al. (2023)

Role nouns in the lexicon

- by now we know that
  - semantically, generic and specific masculines are highly similar
  - when it comes to their form, they are identical
  - the interplay of very similar meaning and identical form leads to almost no difference in their comprehension
  - feminine forms, in contrast, are different than masculine forms
- but what happens when we introduce a new third form to the paradigm?

Study: LDL and the gender star

- a paradigm now has four members
  - specific masculine *Künstler*
  - generic masculine *Künstler*
  - specific feminine *Künstlerin*
  - gender star form *Künstler*in*
- based on a manually annotated corpus, semantic embeddings were computed
- these embeddings - and the forms of the targets - are the input for yet another LDL model
- … or rather, 20 of them

Schmitz (submitted)

Study: LDL and the gender star

C matrix

- orthographic trigrams used as form cues

S matrix

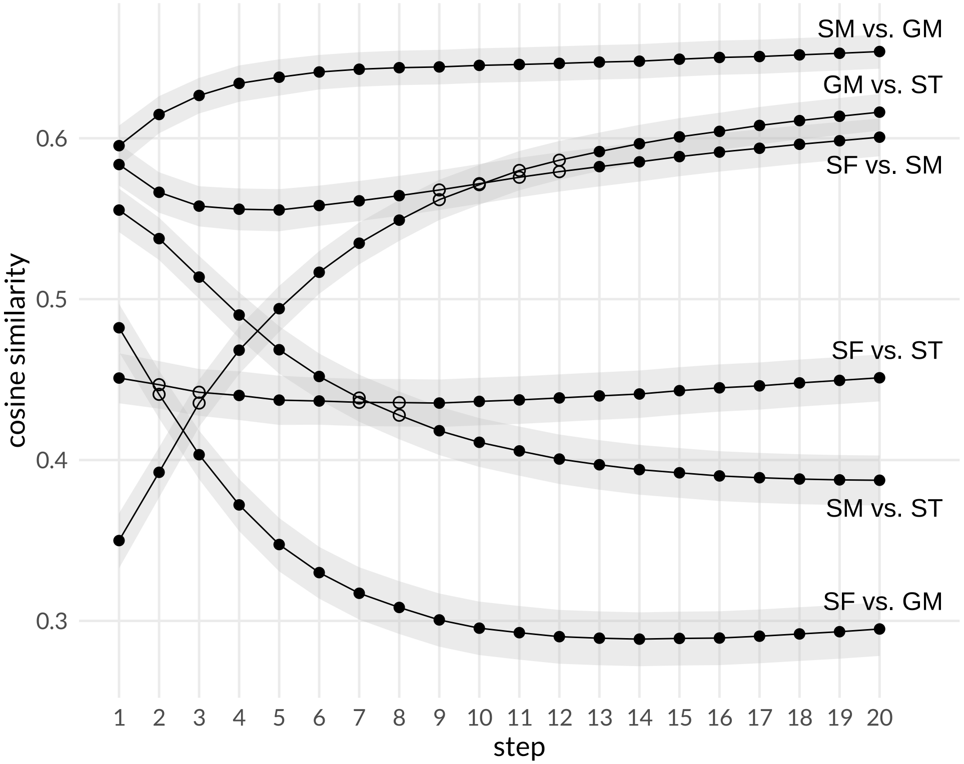
- *fastText* semantic embeddings for each word serve as meaning vectors

mapping

- this time we're using another flavour of matrix mapping: frequency-informed
- across the 20 LDL models, the frequency of the gender star form approaches the frequency of its generic masculine counterpart

Schmitz (submitted)

Study: LDL and the gender star



- generic ↔ specific masculine: consistently most similar at all steps
- gender-star ↔ generic masculine: similarity increases as star frequency rises
- gender-star ↔ specific masculine: similarity decreases over steps
- gender-star ↔ specific feminine: relatively stable, and from about step 9 onward, gender star is more similar to feminine than to specific masculine

Schmitz (submitted)

Study: LDL and the gender star

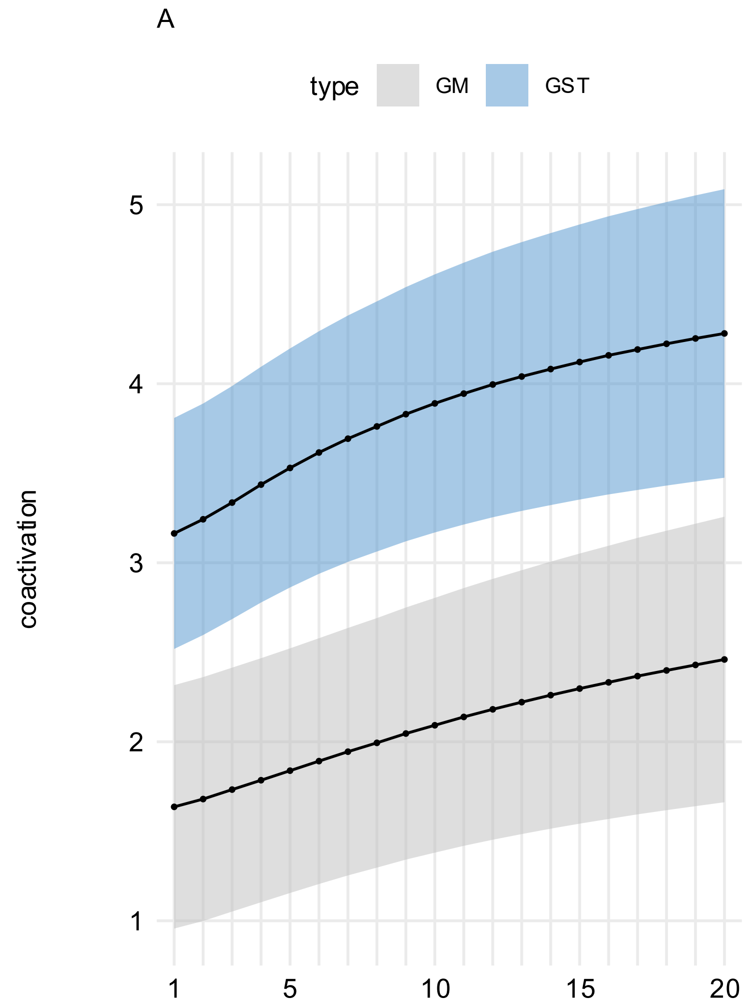
- gender star forms are pulled toward generic masculines in terms of semantics, but their form overlap with feminines shapes their position relative to specific masculines and feminines
- so, what about the underlying semantic measures of gender star forms?
- let's look at
  - semantic coactivation
  - word-level certainty
  - lexicon-level uncertainty

Schmitz (submitted)

Study: LDL and the gender star

semantic coactivation

- both GM and GST increase over steps
- GST always higher than GM
- accessing a gender-star form activates more semantic dimensions

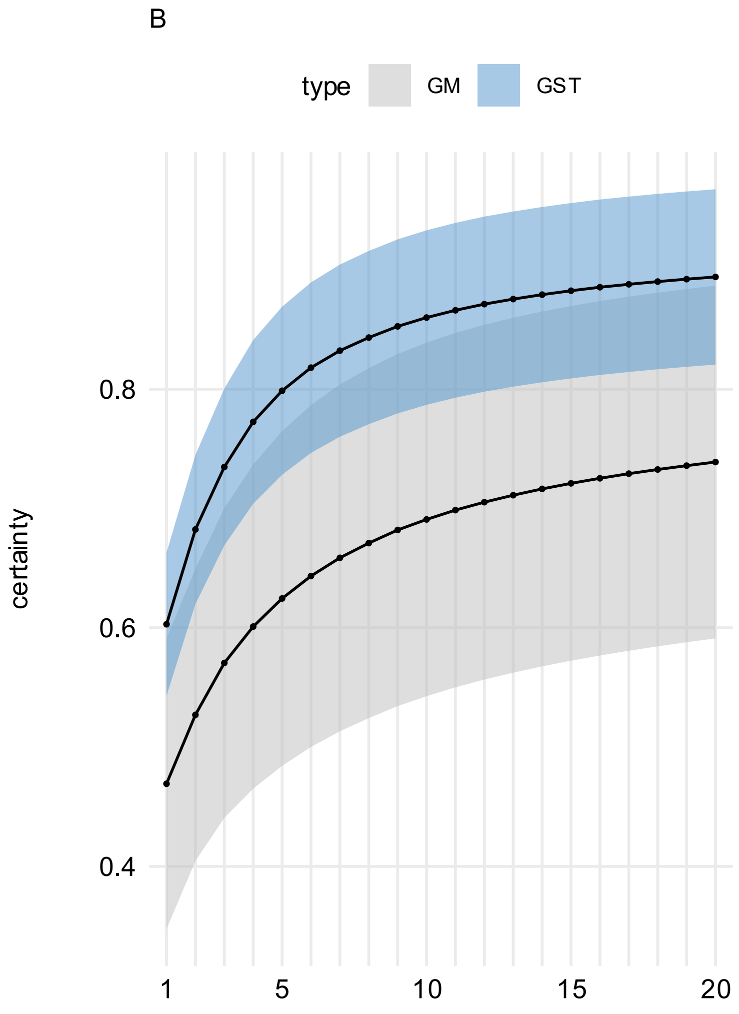


Schmitz (submitted)

Study: LDL and the gender star

word-level certainty

- both GM and GST become more certain with frequency
- GST consistently more certain than GM
- LDL predicts GST semantics reliably once frequency is sufficient

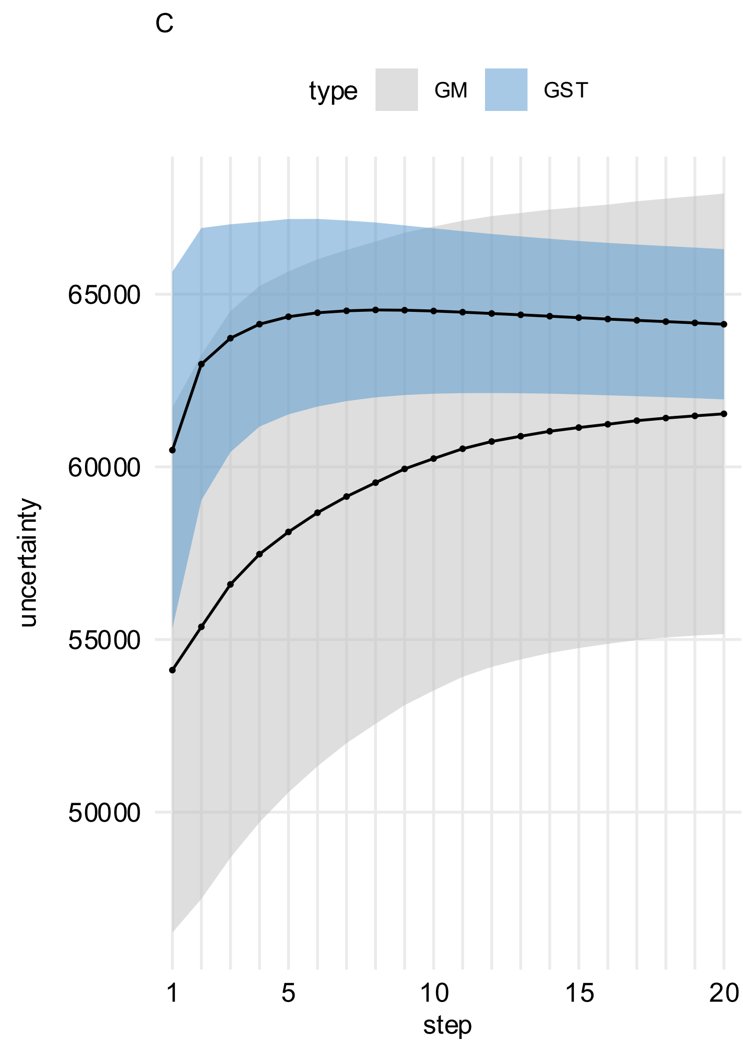


Schmitz (submitted)

Study: LDL and the gender star

lexicon-level uncertainty

- both GM and GST show high uncertainty values
- GST has higher lexicon-level uncertainty than GM, especially in early steps, then plateau/slight decrease
- GST sits in a densely populated semantic region (many near neighbours)



Schmitz (submitted)

Study: LDL and the gender star

- what do these findings mean for actual language use and processing?
- let’s re-analyse some psycholinguistic studies!
  - Körner et al. (2022): continuation judgement RTs,
  - Kurz & Mulder (2023): exemplar retrieval,
  - Schunack & Binanzer (2022): perceived gender ratios
- idea: replace simple categorical predictor type (GM vs GST) with LDL measures (coactivation, certainty, uncertainty)
- result: LDL measures, esp. coactivation, meaningfully account for patterns like
  - slower RTs for male continuations after GST,
  - increased proportion of female exemplars with GST,
  - higher estimated female ratios for GST paradigms

Schmitz (submitted)

Study: ???

Study: Masculines, /ɐ/, and the mental lexicon

- previous research found durational differences conditioned by morphologically different but phonologically identical elements
  - homophonous free and bound (pseudo-)stems (e.g. Seyfarth et al. 2017)

    frees vs. *freeze*

  - homophonous prefixes (e.g. Ben Hedia & Plag 2017)

    impossible vs. *implant* (negative vs. locative)

  - types of /s/ (e.g. Plag et al. 2017, Schmitz et al. 2021)

    bus vs. *cats* vs. *cat's* (non-morphemic vs. suffix vs. clitic)

- homophonous forms show differences in their phonetic realisation

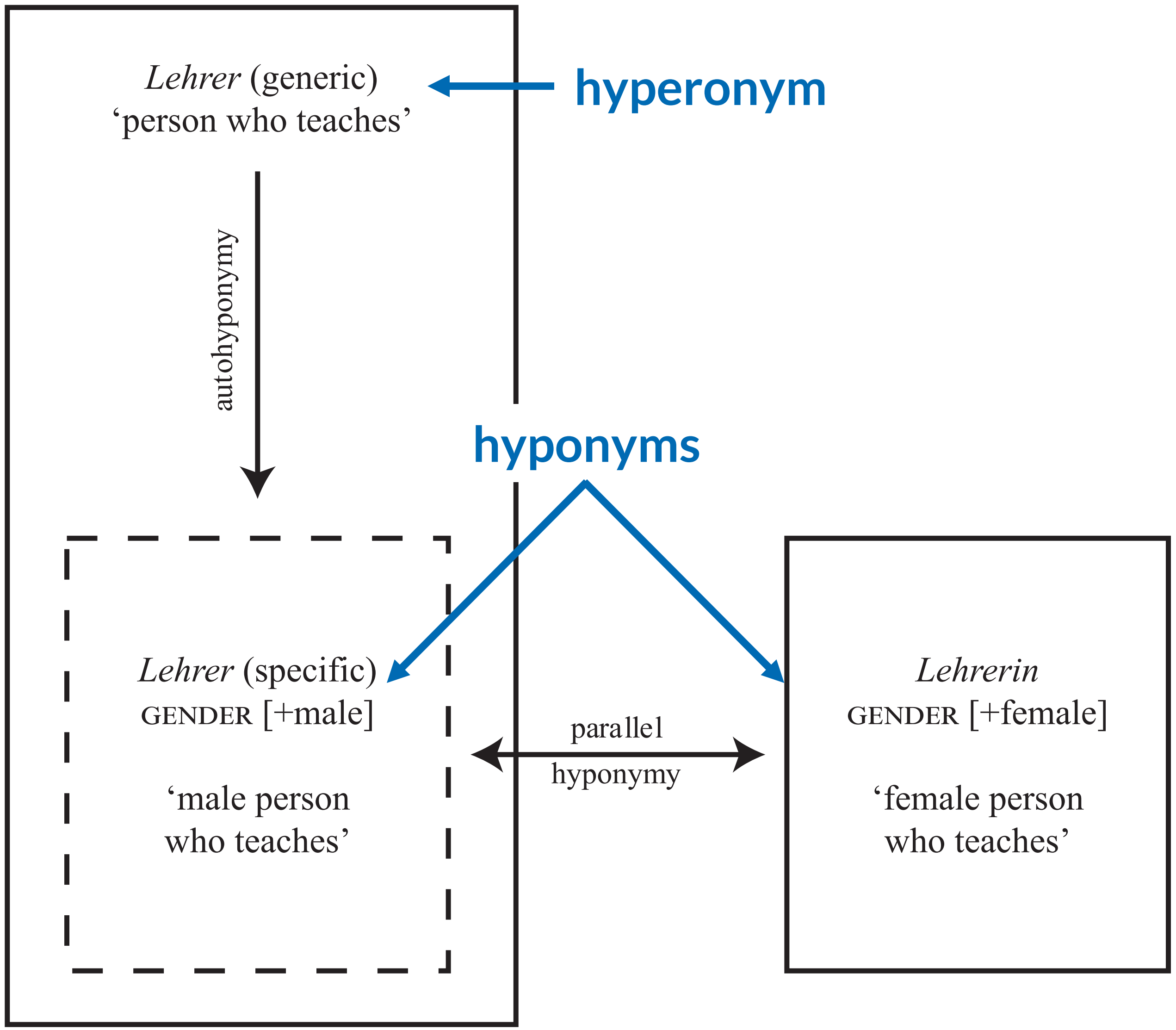
Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

- according to prominent linguistic theories, morphology and phonetics are strangers (e.g. Levelt et al. 1999, Chomsky & Halle 1968) – just as semantics and phonetics are
- but: we just heard of counterevidence for the connection of morphology and phonetics via homophonous elements
- so, what about another type of lexical ambiguity: polysemy?

Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon



Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

Items

- **targets**: 20 role nouns ending in the -*er* suffix, i.e. /ɐ/

stereotypically **female** (Misersky et al., 2014)				
Balletttänzer 'ballet dancer'	*Eiskunstläufer* 'ice skater'	*Flugbegleiter* 'flight attendant'	*Geburtshelfer* 'obstetrician'	*Haushälter* 'housekeeper'
Hellseher 'clairvoyant'	*Kosmetiker* 'beautician'	*Pfleger* 'carer'	*Schneider* 'tailor'	*Verkäufer* 'salesperson'
stereotypically **male**				
Bauarbeiter 'construction worker'	*Elektriker* 'electrician'	*Fußballspieler* 'football player'	*Kranführer* 'crane operator'	*Maurer* 'mason'
Programmierer 'programmer'	*Rennfahrer* 'race driver'	*Reporter* 'reporter'	*Schreiner* 'carpenter'	*Wahrsager* 'fortuneteller'

- **fillers**
  - feminine forms of target items, e.g. *Balletttänzerin*, *Bauarbeiterin*
  - used with female referents only

Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

Contexts - Reading Task

1. phrase or sentence introducing the referent
2. phrase or sentence containing the target item

specific	***Matteos Vater*** *kann richtig gut nähen.* ***Er*** *ist* ***Schneider*** *von Beruf.*
	'**Matteo's father** is really good at sewing. **He** is a **tailor** by profession.'
generic,	***Mein Kind*** *kann richtig gut nähen.* ***Es*** *ist* ***Schneider*** *von Beruf.*
gender unspecificed	'**My child** is really good at sewing. **It** is a **tailor** by profession.'
generic,	***Marias Mutter*** *kann richtig gut nähen.* ***Sie*** *ist* ***Schneider*** *von Beruf.*
gender specificed	'**Maria's mother** is really good at sewing. **She** is a **tailor** by profession.'

Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

Participants - Reading Task

- 40 participants
- L1 German
- age: mean 29.1 years, range: 20 – 64 years

Procedure - Reading Task

- 1 set of context and target phrase/sentence per trial
- instructions: read quietly before reading aloud
- self-paced

Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

Reading Task

Matteos Vater kann richtig gut nähen.
Er ist Schneider von Beruf.

Ton aufnehmen

Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

Contexts - Recall Task

1. sentence introducing the referent
2. sentence introducing the referent's occupation
3. question about the referent's occupation

specific	*Das ist **Lasse**. **Lasse** ist **Pfleger** im Hospiz. Was ist **Lasse**?* 'This is **Lasse**. **Lasse** is a **nurse** at the hospice. What is **Lasse**?'
generic	*Das ist **Lisa**. **Lisa** ist **Pfleger** im Hospiz. Was ist **Lisa**?* 'This is **Lisa**. **Lisa** is a **nurse** at the hospice. What is **Lisa**?'
feminine	*Das ist **Lisa**. **Lisa** ist **Pflegerin** im Hospiz. Was ist **Lisa**?* 'This is **Lisa**. **Lisa** is a **nurse** at the hospice. What is **Lisa**?'

Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

Participants - Recall Task

- 210 participants
- L1 German
- age: mean 42.3 years, range: 22 – 64 years

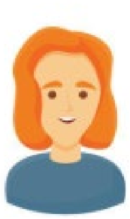
Procedure - Recall Task

- 1 context per trial
- instructions: read quietly, then answer the question
- self-paced

Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

Recall Task



Das ist Iva.
Iva ist Balletttänzer beim städtischen Theater.

Weiter

Schmitz (in prep.)

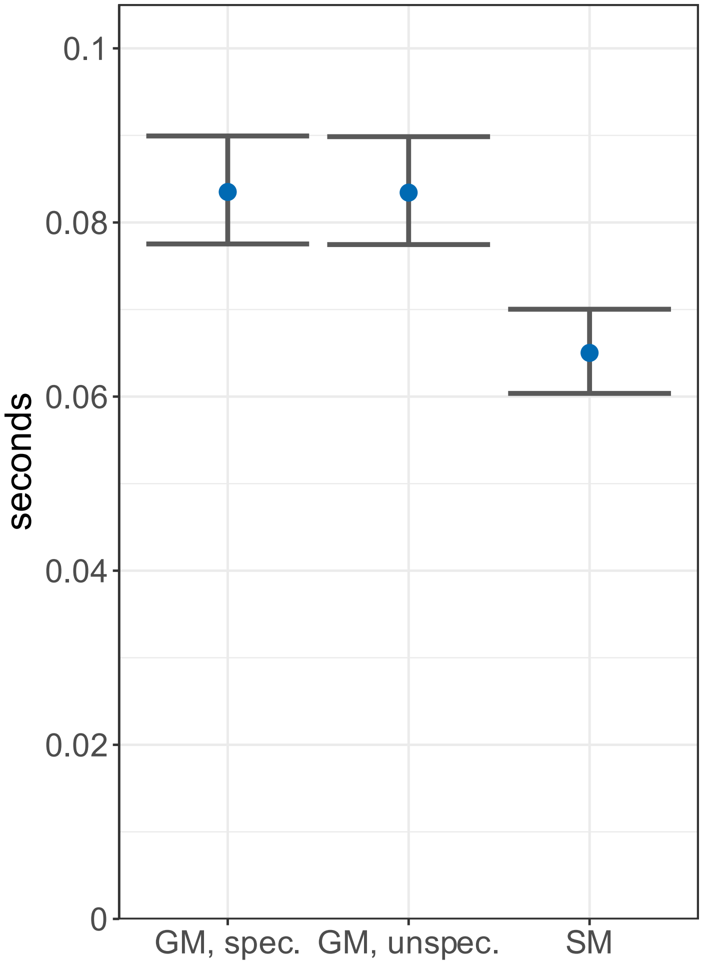
Study: Masculines, /ɐ/, and the mental lexicon

Recall Task

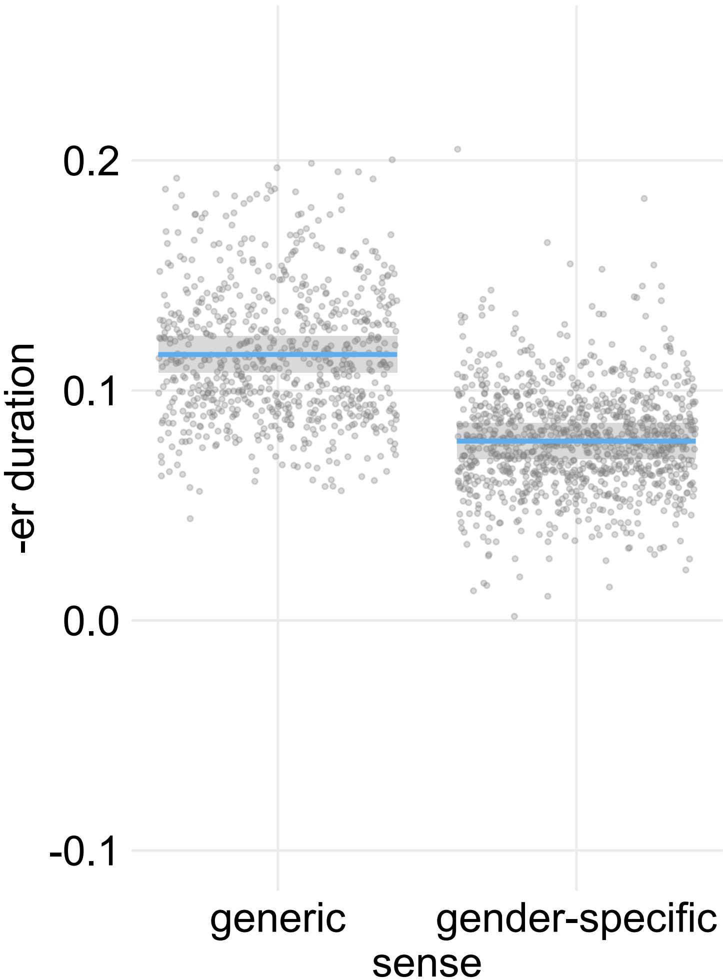
Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

Reading Task



Recall Task



Schmitz (in prep.)

the crucial question

Why is a generic masculine /ɐ/ longer than a specific masculine /ɐ/?

Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

C matrix

- phonological trigrams used as form cues

S matrix

- context-informed vectors from BERT, an LLM

mappings

- comprehension: form $\rightarrow$ meaning
- production: meaning $\rightarrow$ form

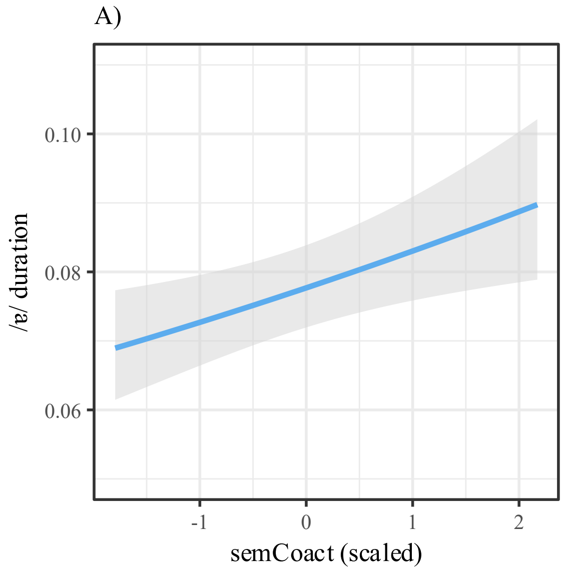
idea

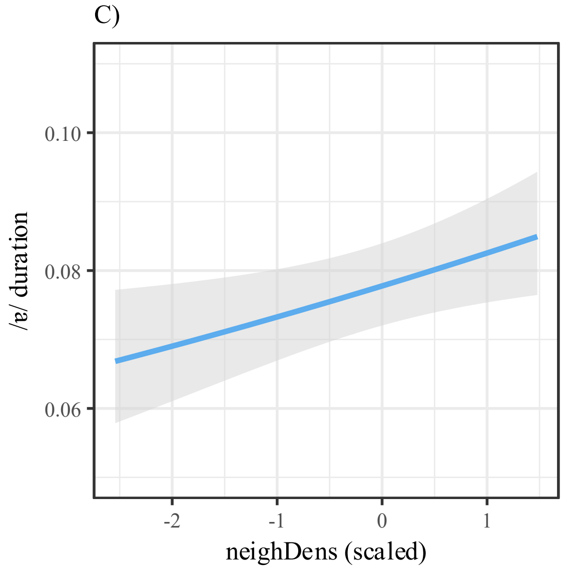
- can LDL derived measures on semantics and form predict /ɐ/ duration?

Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

- in both tasks, measures derived from the LDL model explain the duration of /ɐ/
- for example, in the **reading** data, generic masculines come with higher levels of semantic co-activation and denser neighbourhoods
- higher values of semantic co-activation and denser neighbourhoods, in turn, come with longer /ɐ/ durations
- this is in line with the analysis independent of LDL, in which generic masculines show longer /ɐ/ durations





Schmitz (in prep.)

Study: Masculines, /ɐ/, and the mental lexicon

- generic and gender-specific masculines differ systematically in word-final /ɐ/ duration across tasks
- LDL measures show why: generic senses produce broader semantic activation and denser semantic neighbourhoods → longer duration
- semantic structure does not stay abstract: it shapes the phonetic signal
- the mental lexicon is not symbolic storage: it is a predictive, discriminative system
- polysemy is not just cognitive - it's acoustically real

Schmitz (in prep.)

Bringing things together

- **semantic level**
  - distributional semantics shows GM ≈ SM ≠ SF
  - gender star forms carve out new positions in vector space
- **lexical-discriminative level**
  - LDL explains similarity patterns, learning trajectories, and processing variance
  - novel morphology integrates through cue-based mapping
- **acoustic level**
  - competing senses (generic vs specific) produce measurable differences in word-final /ɐ/ duration
  - lexical ambiguity → phonetic consequences

Schmitz et al. (2023), Schmitz (2024), Schmitz (submitted), Schmitz (in prep.)

THANK YOU!

References 1

Arnold, D., Tomaschek, F., Sering, K., Lopez, F., & Baayen, R. H. (2017). Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PLOS ONE*, *12*(4), e0174623.

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, *2019*, 4895891.

Baayen, R. H., & Smolka, E. (2020). Modelling morphological priming in German with naive discriminative learning. *Frontiers in Communication, Section Language Sciences*, *5*(17), 1–23.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438–481

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, *6*(1), 213–234.

Chuang, Y.-Y., & Baayen, R. H. (2021). Discriminative learning and the lexicon: NDL and LDL. *Oxford Research Encyclopedia of Linguistics*.

Chuang, Y. Y., Vollmer, M. L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., & Baayen, R. H. (2020). The processing of nonword form and meaning in production and comprehension: A computational modeling approach using linear discriminative learning. *Behavior Research Methods*, *53*(3), 945–976.

Firth, J.R. (1957). *A synopsis of linguistic theory 1930-1955*. Studies in Linguistic Analysis: 1–32. London: Longman.

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., & Oakhill, J. (2008). Au pairs are rarely male: Norms on the gender perception of role names across English, French, and German. *Behavior Research Methods*, *40*(1), 206–212.

Harris, Z. S. (1954). Distributional structure. *WORD*, *10*(2–3), 146–162.

References 2

Lapesa, G., Kawaletz, L., Plag, I., Andreou, M., Kisselew, M., & Padó, S. (2018). Disambiguation of newly derived nominalizations in context: A Distributional Semantics approach. *Word Structure*, *11*(3), 277–312.

Maaten, L. van der, & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*(86), 2579–2605.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13) - Volume 2*.

Müller-Spitzer, C., Ochs, S., Rüdiger, J. O., & Wolfer, S. (2025). *Geschlechtsübergreifende Maskulina im Sprachgebrauch: Eine korpusbasierte Untersuchung zu lexemspezifischen Unterschieden*.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Schmitz, D. (2024). Instances of bias: The gendered semantics of generic masculines in German revealed by instance vectors. *Zeitschrift für Sprachwissenschaft*, *43*(2).

Schmitz, D. (in prep.). Autohyponymy and the phonetic signal: The duration of word-final -er in German.

Schmitz, D., Ochs, S., Rüdiger, J. O., & Müller-Spitzer, C. (submitted). Context fails to neutralise the male bias of generic masculines.

Schmitz, D., Plag, I., Baer-Henney, D., & Stein, S. D. (2021). Durational differences of word-final /s/ emerge from the lexicon: Modelling morpho-phonetic effects in pseudowords with linear discriminative learning. *Frontiers in Psychology*, *12*.

Schmitz, D., Schneider, V., & Esser, J. (2023). No genericity in sight: An exploration of the semantics of masculine generics in German. *Glossa Psycholinguistics*, *2*(1).

Thater, S. (2011). Distributionelle Semantik [PowerPoint slides]. Seminar Allgemeine Linguistik (Computerlinguistik), Universität des Saarlandes. https://www.coli.uni-saarland.de/courses/ds-11/material/01-Intro.pdf

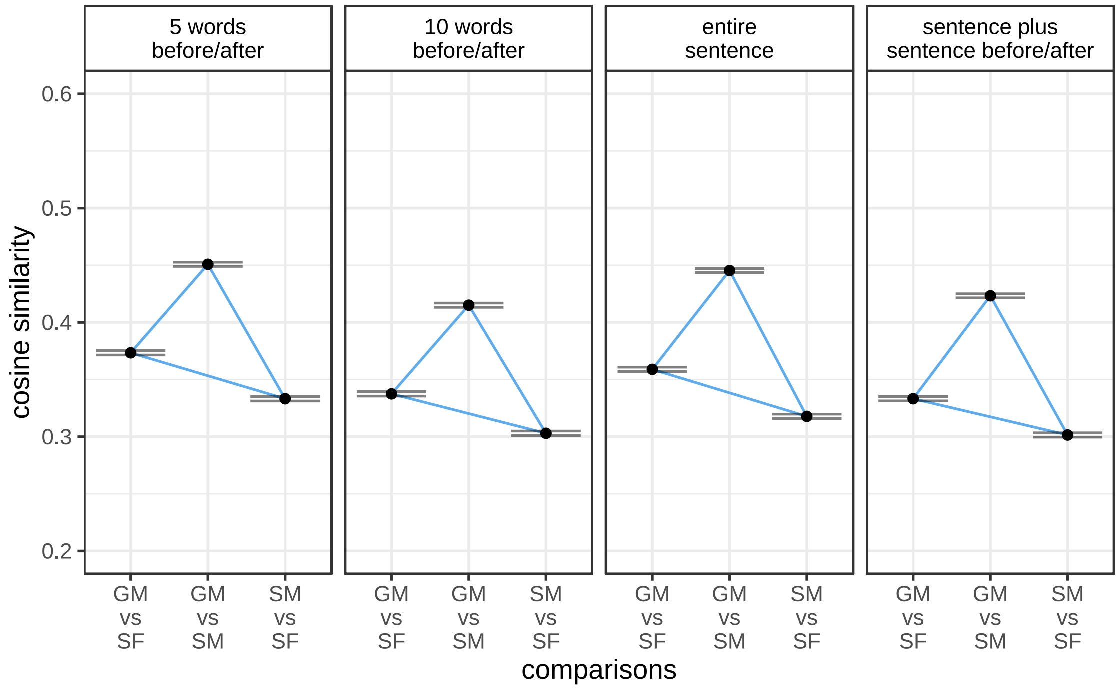
LLMs: using BERT to model generic masculines

- Large language models (e.g. GPT2, BERT, etc.) provide contextualised token embeddings
- Each token receives a vector that already integrates
  - its local morphosyntax,
  - its sentence semantics,
  - broader discourse context
- LLMs do not require us to specify genericity for the model
  - Those distinctions (if present) must emerge from the context itself
  - This eliminates the asymmetry: all tokens receive equally rich contextual representations

Schmitz et al. (submitted)

LLMs: using BERT to model generic masculines

- Even with rich discourse context, generic masculines remain semantically closest to specific masculines
- Context might enlarge semantic distinction overall, but does not selectively pull generic masculine toward a gender-neutral meaning



Schmitz et al. (submitted)