Semantic activation and semantic neighbourhood shape the genericness of role nouns in German

Dominic Schmitz[1], Viktoria Schneider[1], Janina Esser[2]

[1] Heinrich-Heine-Universität Düsseldorf

[2] Association for Diversity in Linguistics

Masculine Generics in German

- in German, role nouns such as *Anwalt* 'lawyer' can be used as generic forms

target word paradigm

word	referent gender(s)	grammatical gender	number
Anwalt	male	masculine	singular
Anwalt	male or female	masculine	
Anwältin	female	feminine	
Anwälte	male	masculine	plural
Anwälte	male and/or female	masculine	
Anwältinnen	female	feminine	

- generic forms are not different from explicit masculine forms in their orthographic or phonological form
- they are used to describe individuals of all genders in singular and plural contexts
- generic forms are traditionally assumed to "abstract away" notions of gender; to be "gender-neutral" (Doleschal, 2002)

Previous Research

- however, previous research has cast doubt on the gender-neutral use of masculine generics
- most (if not all) behavioural studies on the subject find one overall result

  → masculine generics are not gender-neutral but show a clear bias towards the explicit masculine reading (e.g. Demarmels, 2017; Garnham et al., 2012; Gygax et al., 2008; Irmen & Kurovskaja, 2010; Irmen & Linner, 2005; Koch, 2021; Misersky et al., 2019; Stahlberg & Sczesny, 2001; Trutkowski, 2018)
- even though a masculine generic may be used by a speaker with the intention of considering all genders…
- …this intention is not fully translated by the receiver's comprehension system
- instead, a reading favouring male individuals is received

Research Questions

Q1 Do masculine generics show a male bias or is the bias reported in previous research an artefact of behavioural methods?

→ **NDL:** Naïve Discriminative Learning

Q2 Which features of the underlying representations lead to the (dis)similarities of masculine and feminine forms?

→ **LDL:** Linear Discriminative Learning

Part 1

NDL & the male bias

Method: Target Items

- 113 target items were adapted from a study on the influence of stereotypical and grammatical information on the representation of gender in language (Gabriel et al., 2008)
- all target items were role nouns

explicit masculine & generic masculine	translation
Anwalt	‘lawyer’
Bäcker	‘baker’
Dekan	‘dean’
Historiker	‘historian’
Maurer	‘mason’
Professor	‘professor’
Wärter	‘guard’

Method: Target Items

- 113 target items were adapted from a study on the influence of stereotypical and grammatical information on the representation of gender in language (Gabriel et al., 2008)
- all target items were role nouns
- all target items have a common explicit feminine form

explicit masculine & generic masculine	explicit feminine	translation
Anwalt	Anwältin	'lawyer'
Bäcker	Bäckerin	'baker'
Dekan	Dekanin	'dean'
Historiker	Historikerin	'historian'
Maurer	Maurerin	'mason'
Professor	Professorin	'professor'
Wärter	Wärterin	'guard'

Method: Corpus

- 10 million sentences were extracted from the Leipzig Corpora Collection's (Goldhahn et al., 2012) subcorpus "News" → 1 million for each year from 2010 to 2019
- from the 10 million sentences, the following was sampled:
  - 800,000 sentences without any target words
  - 30,000 sentences with target words
- the overall frequency for each target word in our corpus is relative to its overall frequency in the 10 million sentences sample, for example
  - a target with more than 20,000 occurrences is represented by 600 samples
  - a target with less than 200 occurrences is represented by 100 samples

Method: Corpus

- using data from news websites allowed us to strictly control genre
- thus, our results cannot be potential artefacts of ‘genre confusion’, i.e. of chance due to an uncontrolled mix of different styles and genres
- however, this indicates that chances are given that other sources/genres/styles might lead to different results

Method: Annotation

- the 30,000 sentences containing target words were manually annotated by two authors and two assistants, all of which were native speakers of German
- for each target word occurrence, it was annotated whether the form was
  - masculine or feminine; singular or plural; explicit or generic
- the 800,000 sentences without and the 30,000 sentences with target words were then automatically analysed and annotated using the RNNTagger software (Schmid, 1999)
- tagged information consisted of words' base forms and information on inflectional grammar
- the manually compiled annotation and the automatic annotation were finally brought together for sentences with target words

Method: Distributional Semantics

- Distributional Hypothesis (e.g. Harris, 1954)

  difference in meaning ↔ difference in distribution

- difference in meaning is measured via semantic vectors

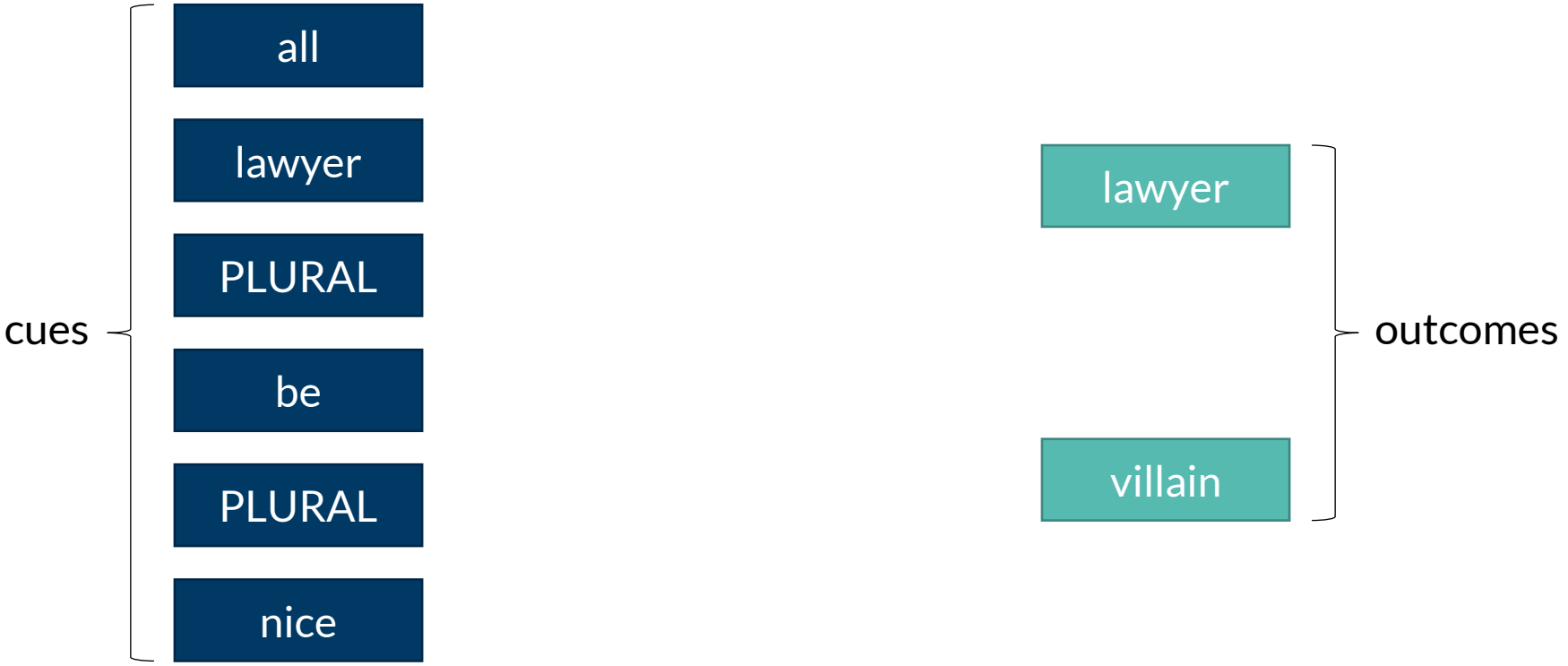
- one way to arrive at a word's semantic vector is Naïve Discriminative Learning (NDL)

  (Baayen & Ramscar, 2015)

Method: Naïve Discriminative Learning

- taking the 830,000 annotated sentence corpus as a starting point, we computed semantic vectors for bases and inflectional functions using NDL
- NDL follows the Rescorla-Wagner rules (Rescorla & Wagner, 1972; Wagner & Rescorla, 1972)
- most importantly, these rules state that
  - outcomes are predicted by cues
  - the associative strength between an outcome and a cue is represented by a single number
- we used each sentence to predict each individual outcome within the sentence by the other bases/inflectional functions in that sentence

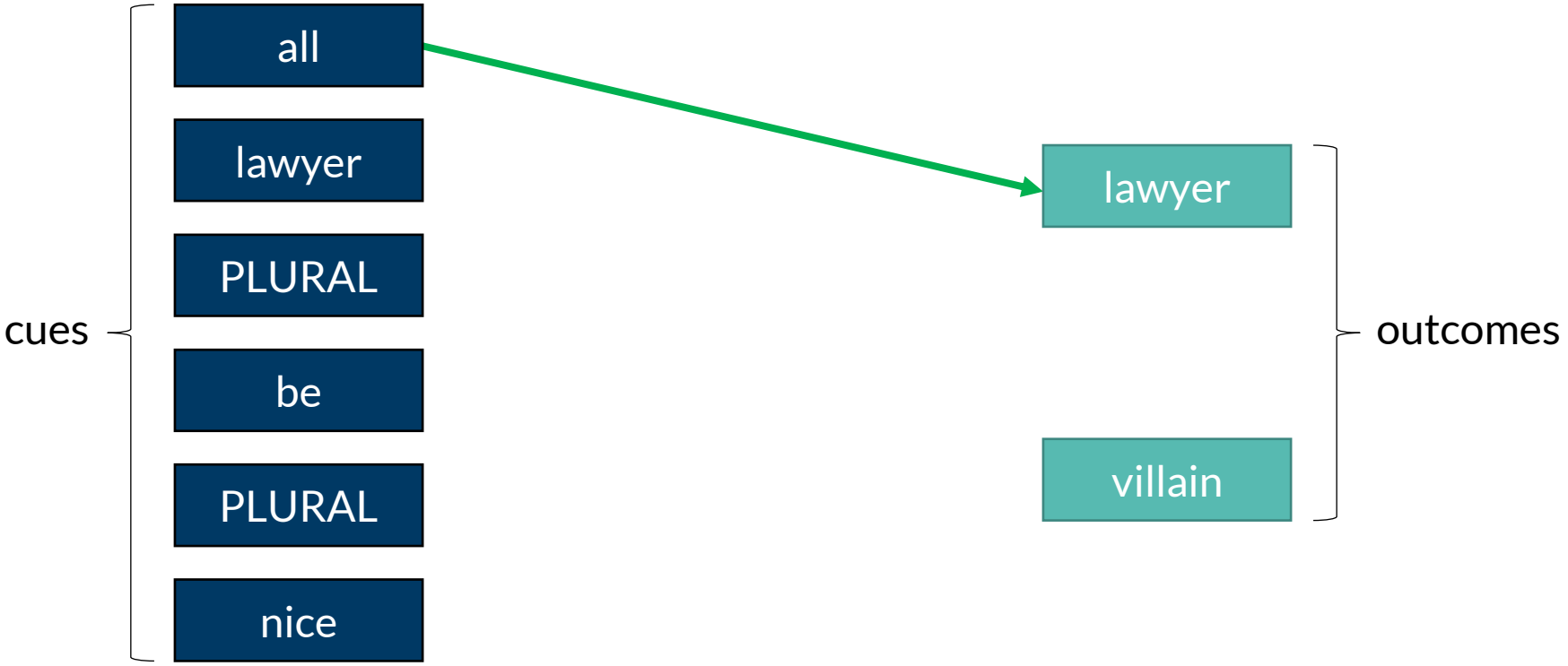
Method: Naïve Discriminative Learning



Example: *All lawyers are nice.*

	all	lawyer	PLURAL	be	nice	villain	evil
lawyer							
villain							

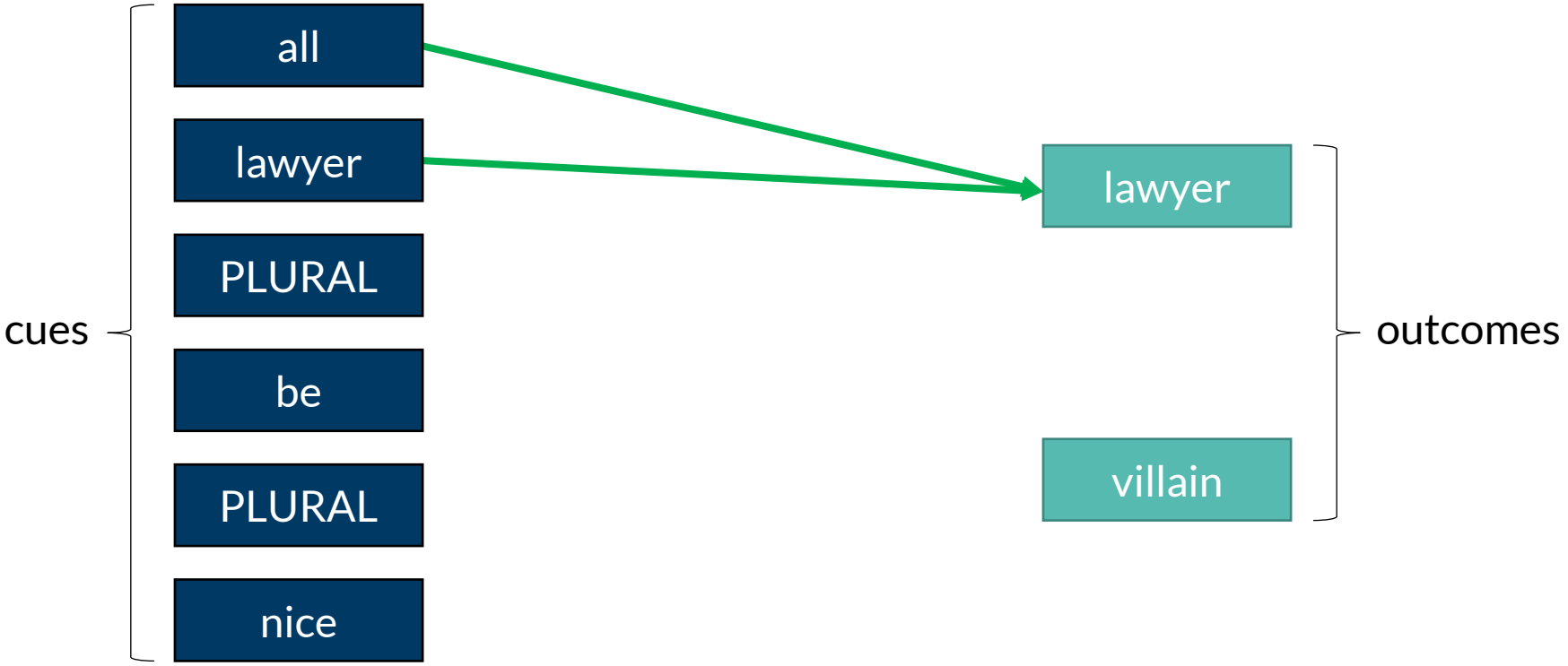
Method: Naïve Discriminative Learning



Example: *All lawyers are nice.*

	all	lawyer	PLURAL	be	nice	villain	evil
lawyer	+						
villain							

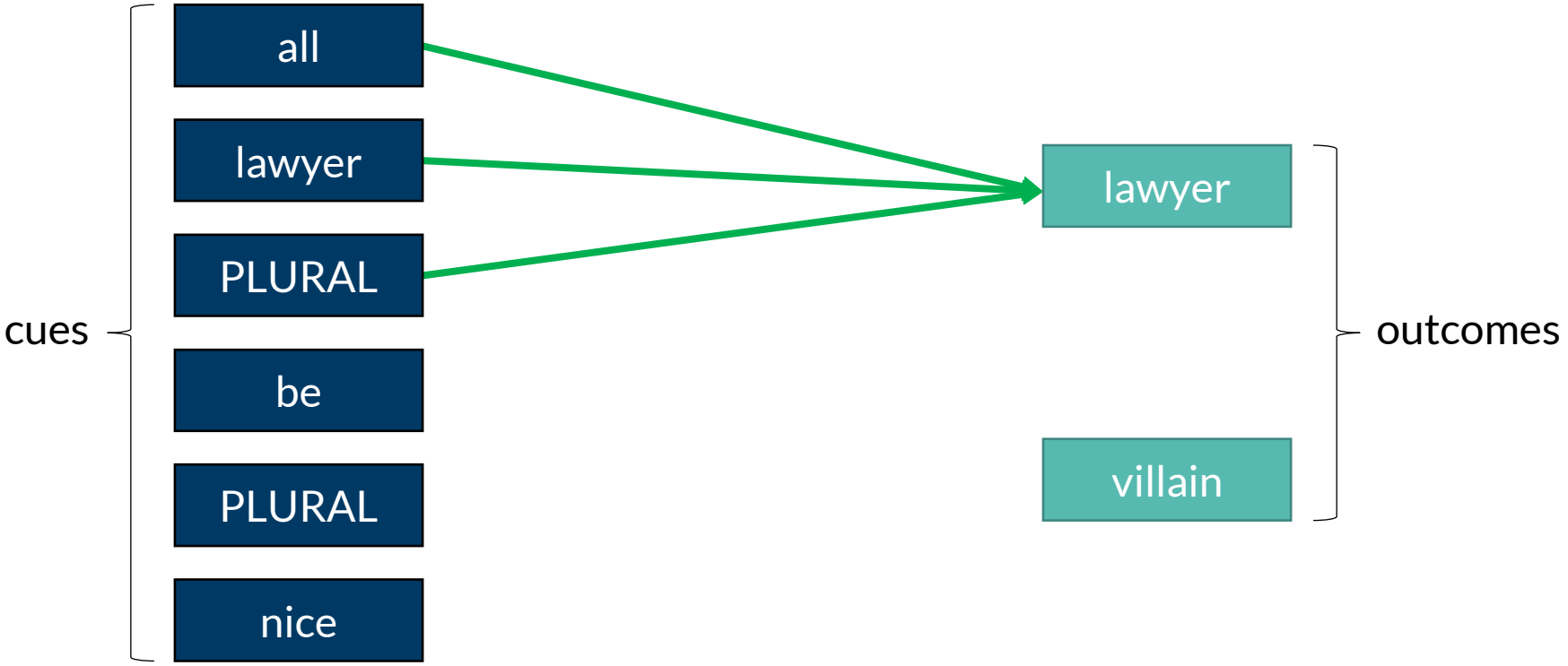
Method: Naïve Discriminative Learning



Example: *All lawyers are nice.*

	all	lawyer	PLURAL	be	nice	villain	evil
lawyer	+	+					
villain							

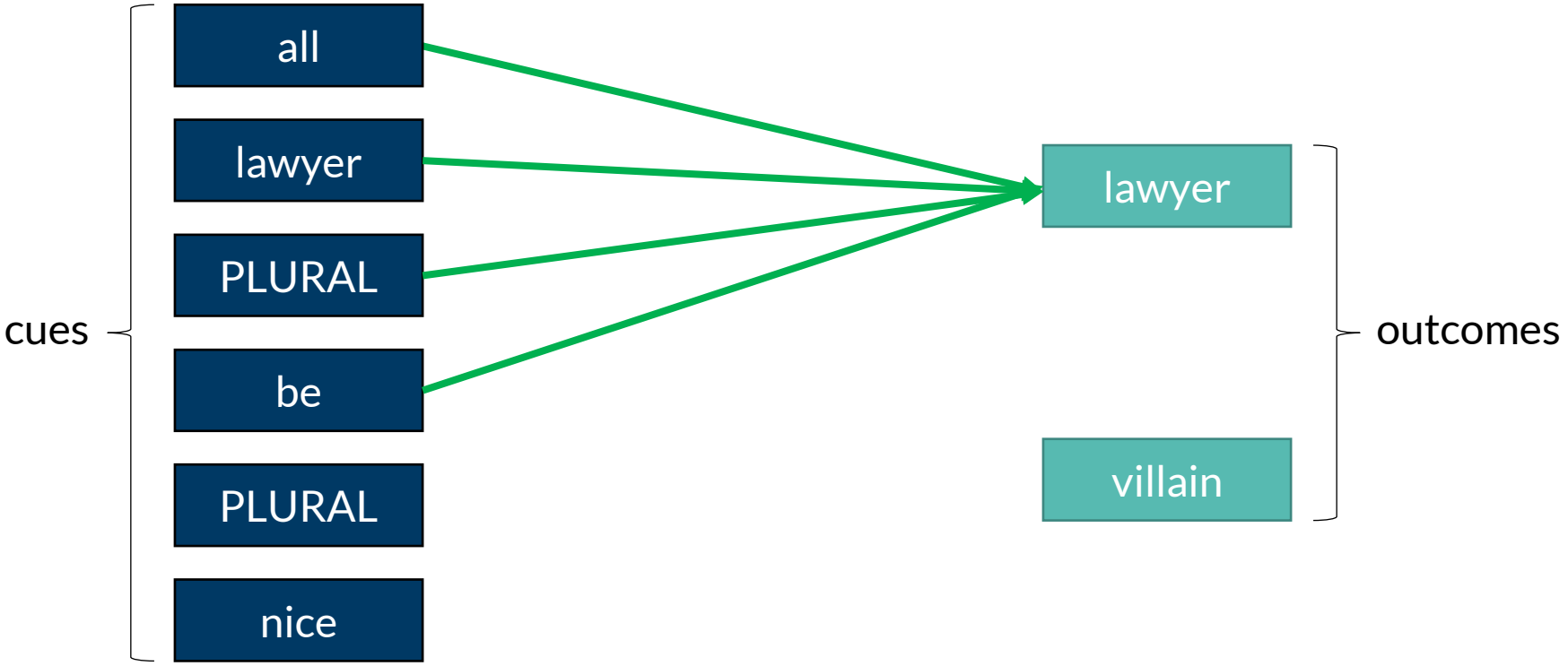
Method: Naïve Discriminative Learning



Example: *All lawyers are nice.*

	all	lawyer	PLURAL	be	nice	villain	evil
lawyer	+	+	+				
villain							

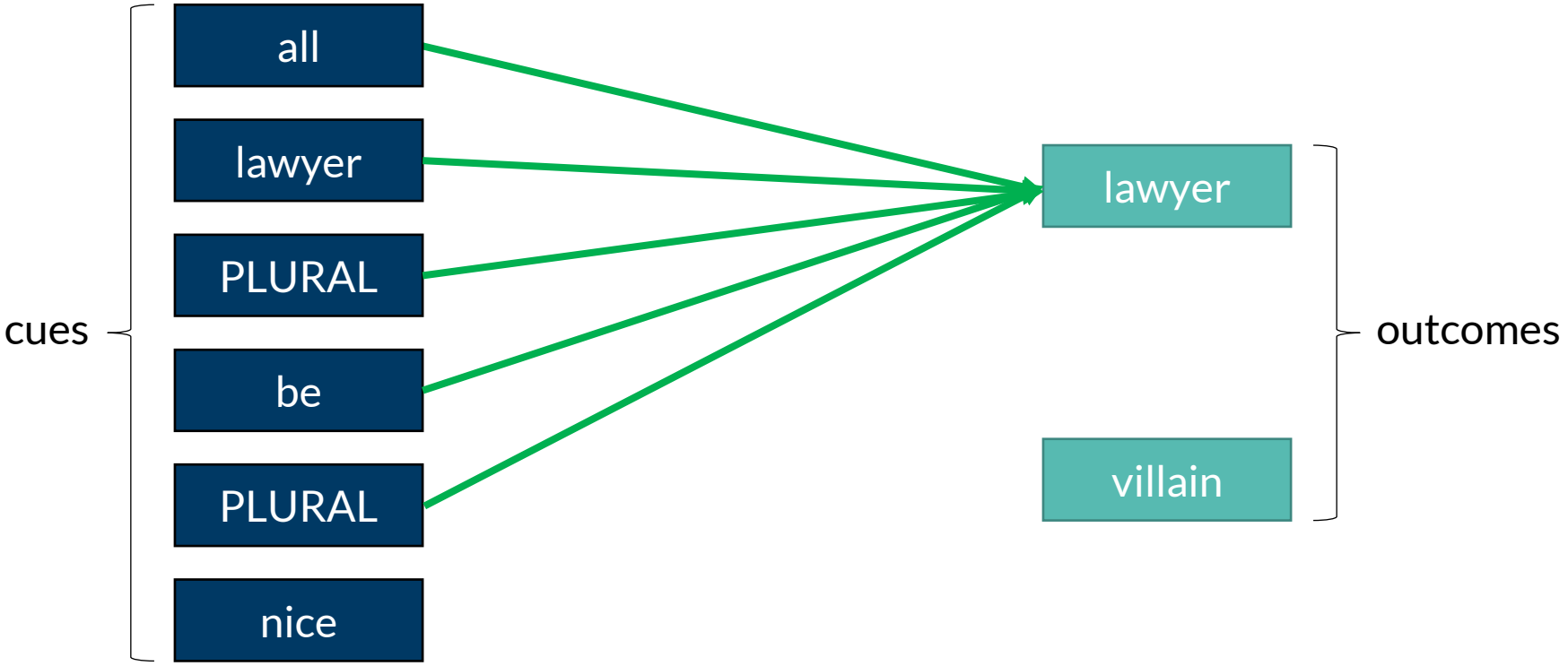
Method: Naïve Discriminative Learning



Example: *All lawyers are nice.*

	all	lawyer	PLURAL	be	nice	villain	evil
lawyer	+	+	+	+			
villain							

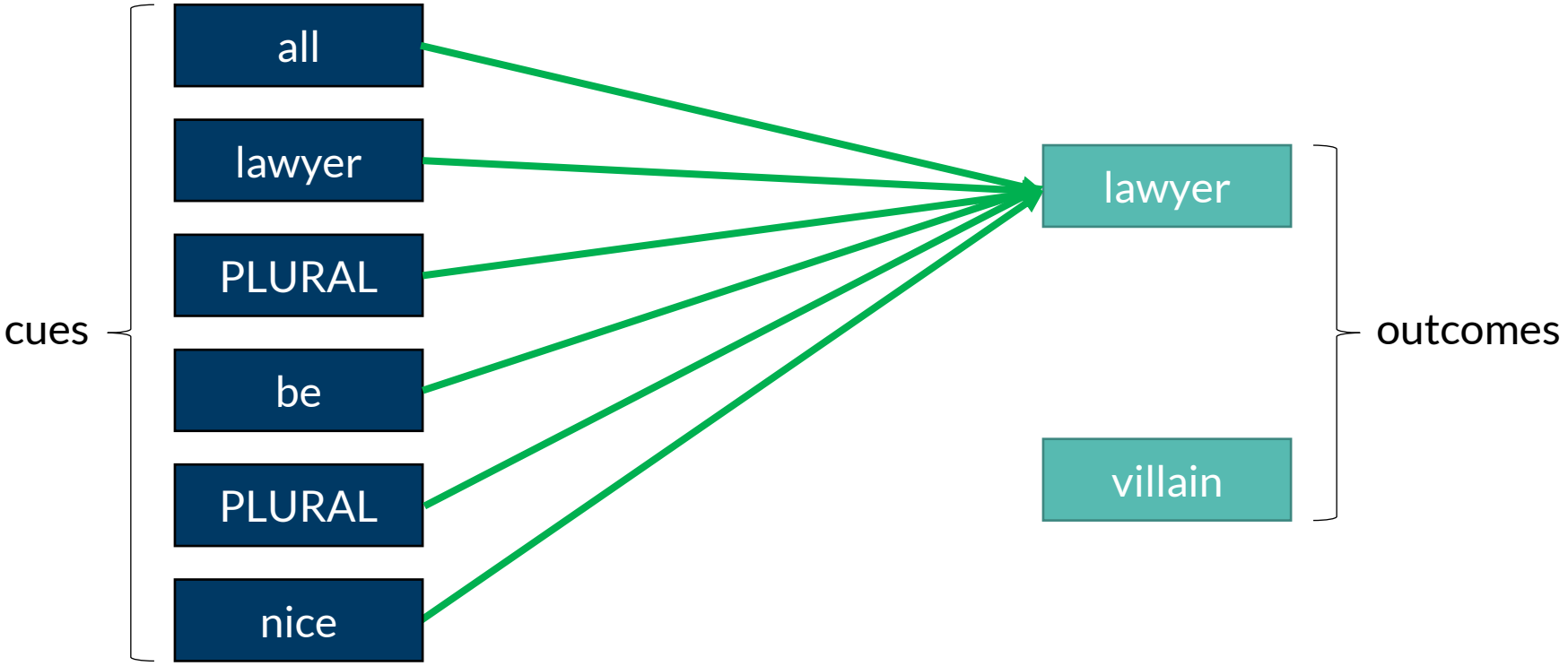
Method: Naïve Discriminative Learning



Example: *All lawyers are nice.*

	all	lawyer	PLURAL	be	nice	villain	evil
lawyer	+	+	++	+			
villain							

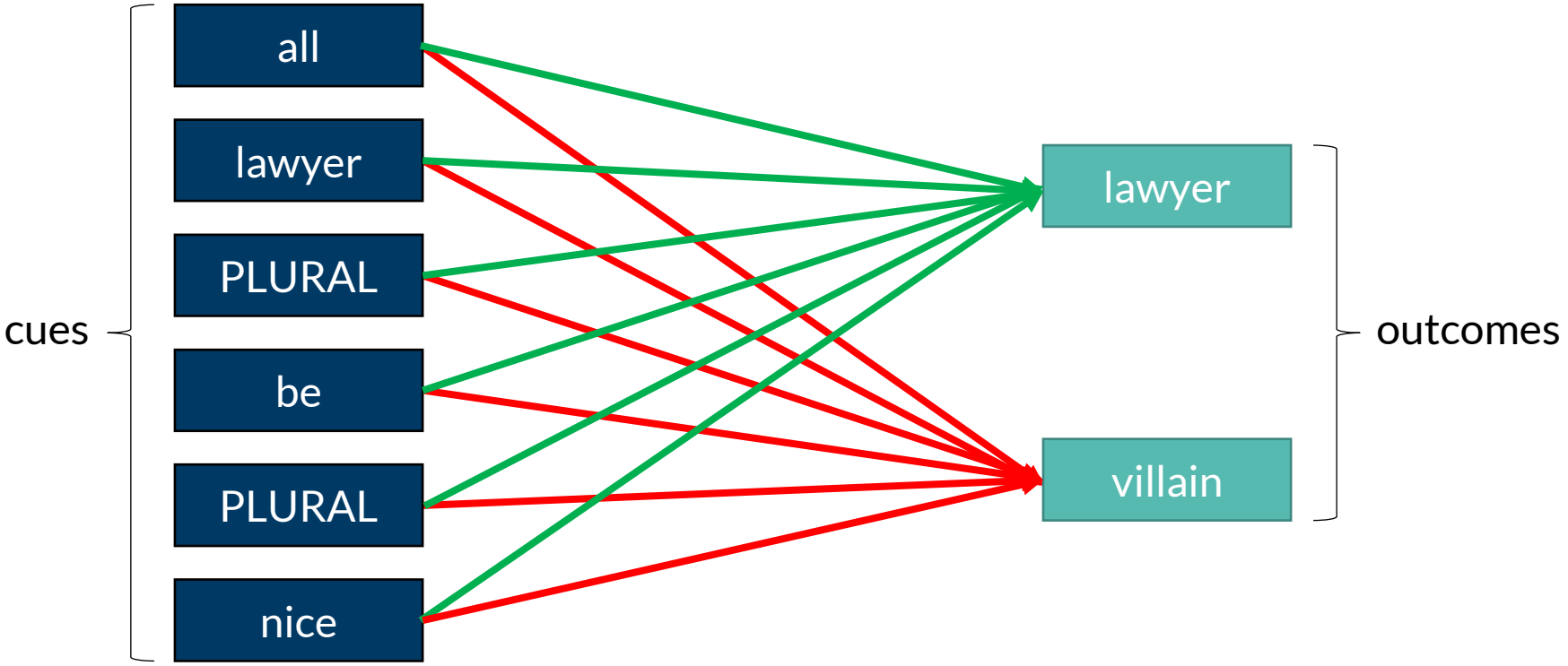
Method: Naïve Discriminative Learning



Example: *All lawyers are nice.*

	all	lawyer	PLURAL	be	nice	villain	evil
lawyer	+	+	++	+	+		
villain							

Method: Naïve Discriminative Learning



Example: *All lawyers are nice.*

	all	lawyer	PLURAL	be	nice	villain	evil
lawyer	+	+	++	+	+	-	-
villain	-	-	-	-	-		

Method: Naïve Discriminative Learning

- repeating this procedure for 830,000 sentences, we obtained association weights for all target word bases, inflectional functions, and a huge number of other bases
- taking these rows of association weights, we obtain semantic vectors of individual bases and inflectional functions (no. of cues: untrimmed 15023; trimmed 7511)
- for example:

	Apfel 'apple'	trinken 'drink'	Gabel 'fork'	Kartoffel 'potato'	Universum 'universe'	Stern 'star'
essen 'eat'	0.3	0.2	0.5	0.4	0.00002	0.000071
Astronomie 'astronomy'	0.0003	0.0015	0.00704	0.0003	0.6	0.8

→ a word's associations with other words and inflectional functions describe the word's semantics

Vectors of Complex Word Forms

- for complex word forms, their vector is the sum of the vectors of their parts,

  e.g. $\overrightarrow{apples} = \overrightarrow{apple} + \overrightarrow{plural}$

- thus, e.g., the semantics of the target word paradigm *Anwalt* 'lawyer' consists of

target form	base		number		gram. gender		type
Anwalt	Anwalt	+	singular	+	masculine	+	generic
Anwalt	Anwalt	+	singular	+	masculine	+	explicit
Anwältin	Anwalt	+	singular	+	feminine	+	explicit

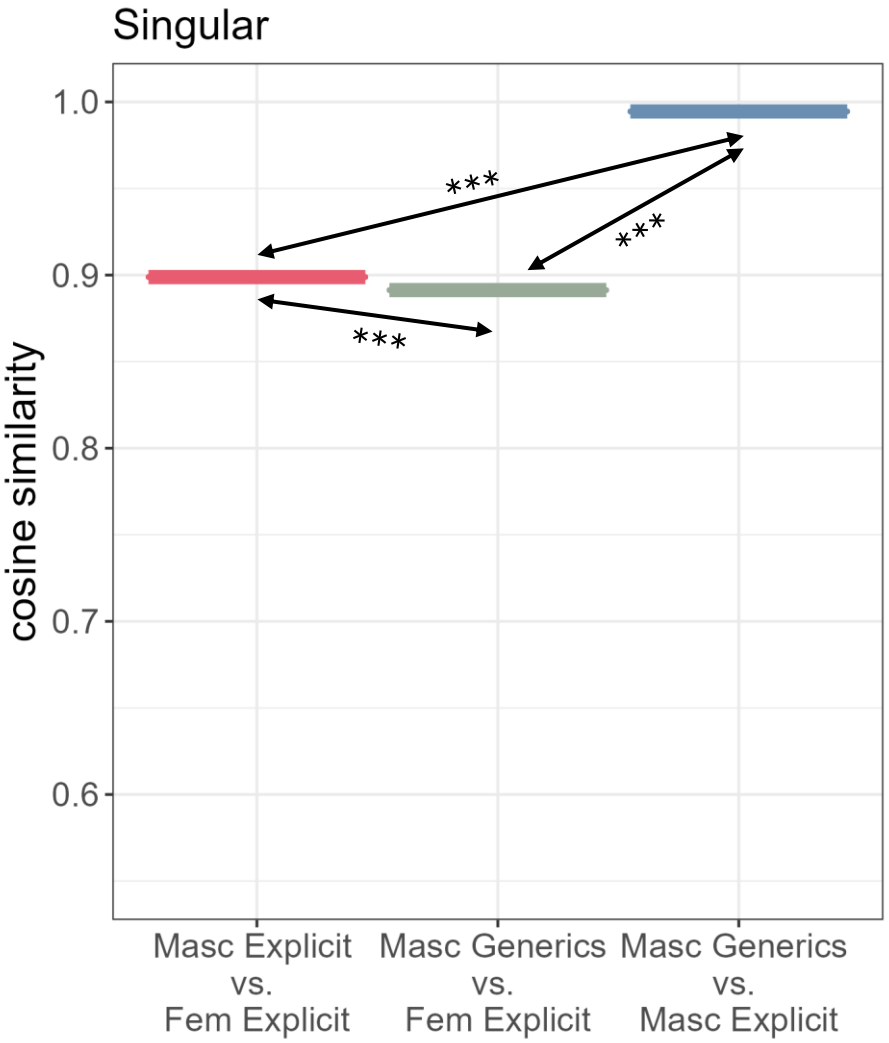
- accordingly, the plural forms are

word form	base		number		gram. gender		type
Anwälte	Anwalt	+	plural	+	masculine	+	generic
Anwälte	Anwalt	+	plural	+	masculine	+	explicit
Anwältinnen	Anwalt	+	plural	+	feminine	+	explicit

Analysis

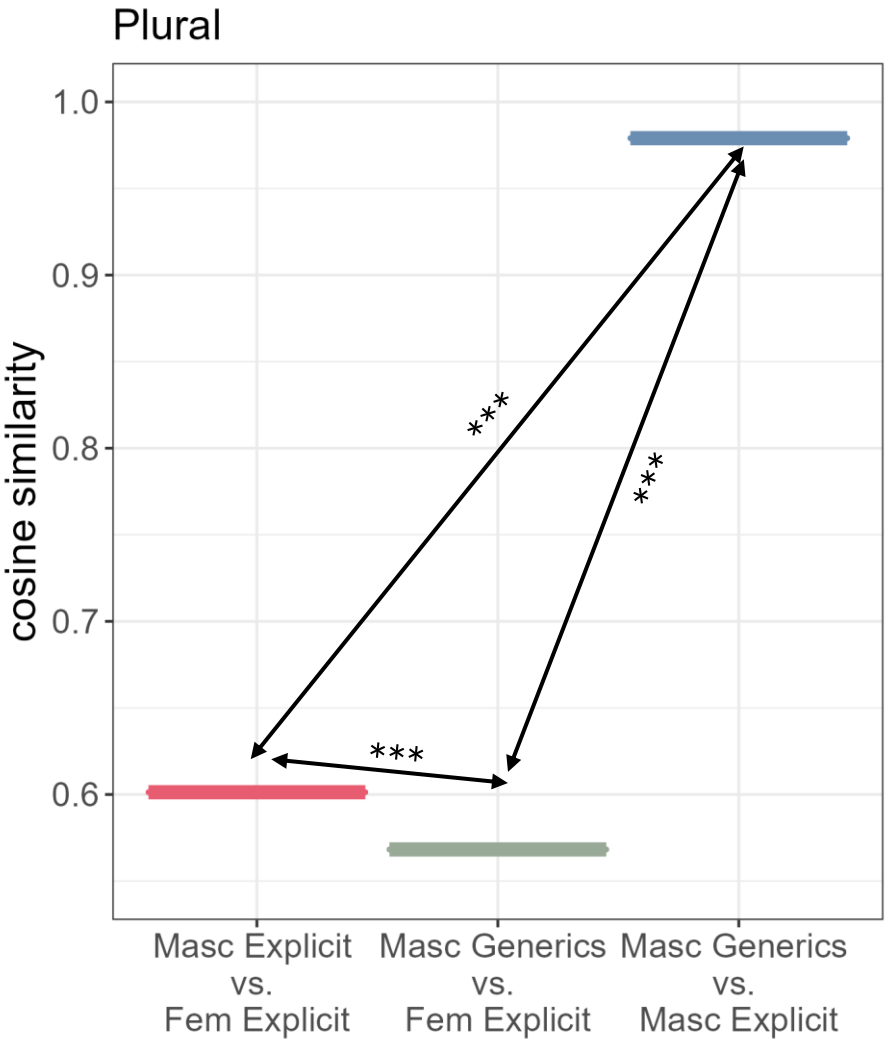
- the resulting semantic vectors of masculine generics, explicit masculines, and explicit feminines can be compared by different statistical means
- we compared their similarity using cosine similarity
- in the present case, cosine similarity values can take values within the interval of $[0, 1]$
- for cosine similarity, a
  - higher value indicates a higher similarity of two vectors
  - lower value indicates a lower similarity of two vectors
- in our case: similarity of vectors reflects similarity of two words' semantics

Results



- masculine generics and the explicit masculine are semantically most similar
- the **explicit feminine** is more similar to the explicit masculine than to masculine generics
- all comparisons are highly significant

Results



- masculine generics and the explicit masculine are semantically most similar
- the **explicit feminine** is more similar to the explicit masculine than to masculine generics
- all comparisons are highly significant
- differences are more pronounced

Interim Summary

Q1 Do masculine generics show a male bias or is the bias reported in previous research an artefact of behavioural methods?

masculine generics do show a male bias

Part 2

LDL & underlying representations

Method: Linear Discriminative Learning

- we simulate an individual's comprehension by implementing a linear discriminative learning network (e.g. Baayen et al., 2019)

Step 1: semantic matrix

	cat	bus	eel
cat	1.0	0.2	0.5
bus	0.4	1.0	0.1
eel	0.2	0.3	1.0

Step 3:
learning comprehension

Step 2: cue matrix

	#k{	k{t	{t#	#bV	bVs	Vs#	#il	il#
cat	1	1	1	0	0	0	0	0
bus	0	0	0	1	1	1	0	0
eel	0	0	0	0	0	0	1	1

Method: Linear Discriminative Learning

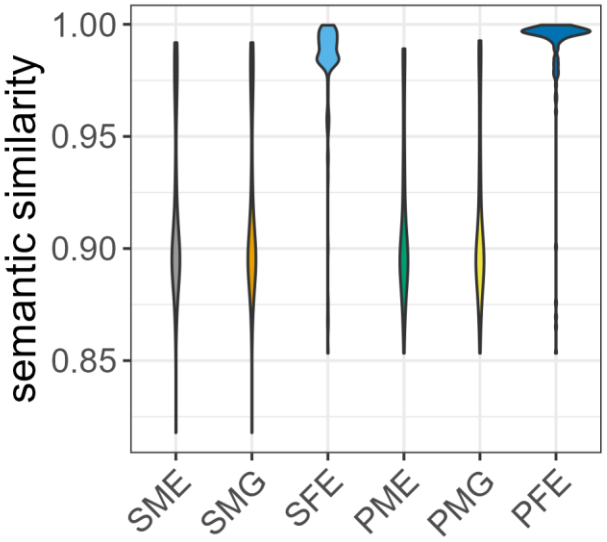
- we simulate an individual's comprehension by implementing a linear discriminative learning network (e.g. Baayen et al., 2019)

LDL Measures

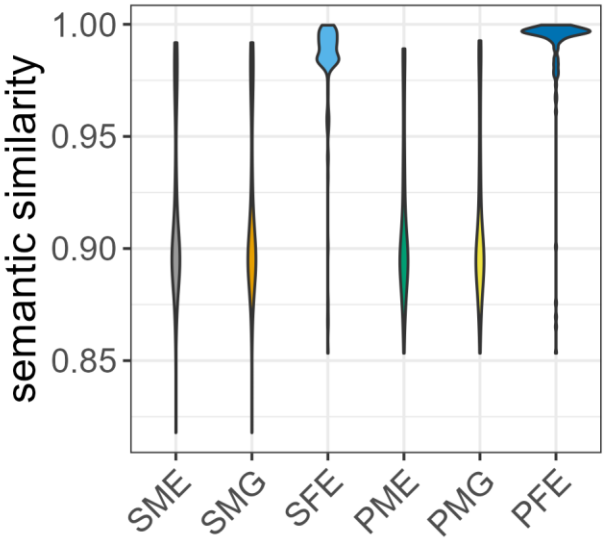
- measures derived from the LDL implementation are
  - total semantic similarity
  - comprehension quality
  - semantic neighbourhood density
  - semantic activation diversity 1
  - semantic activation diversity 2

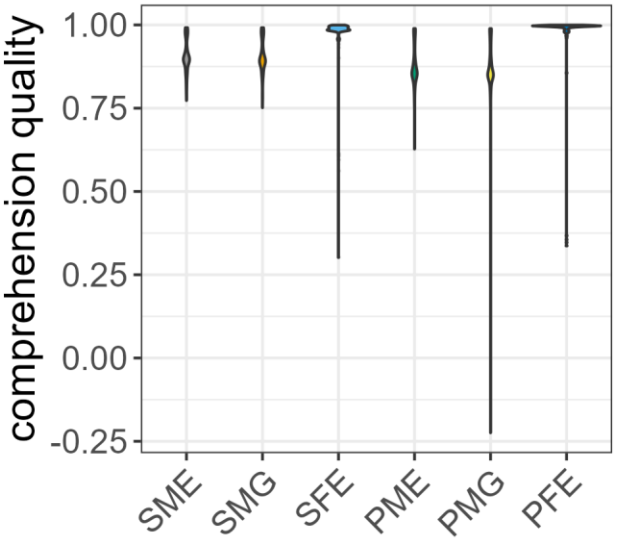
Results 1: LDL Measures

Results 1: LDL Measures

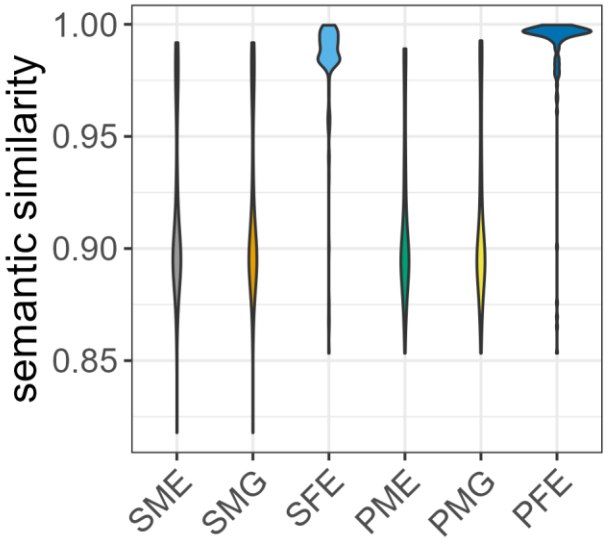


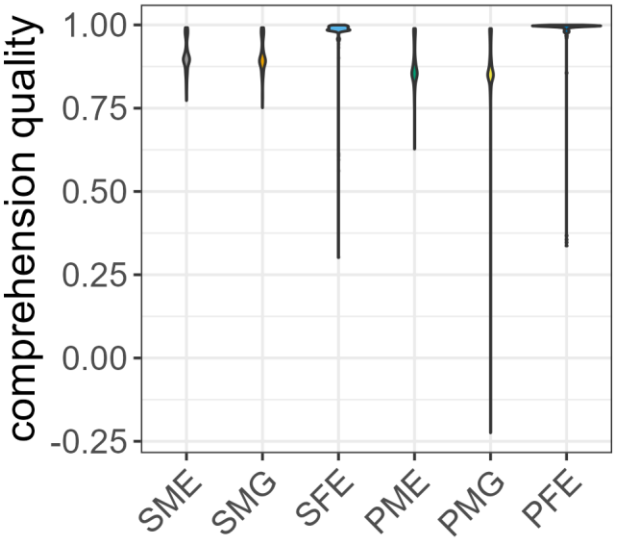
Results 1: LDL Measures

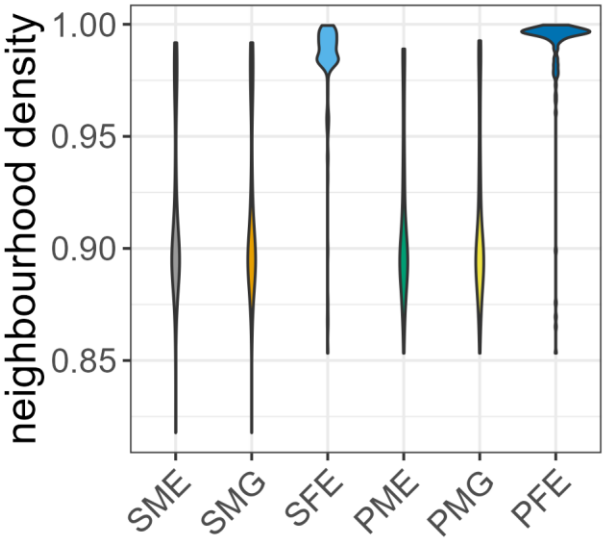




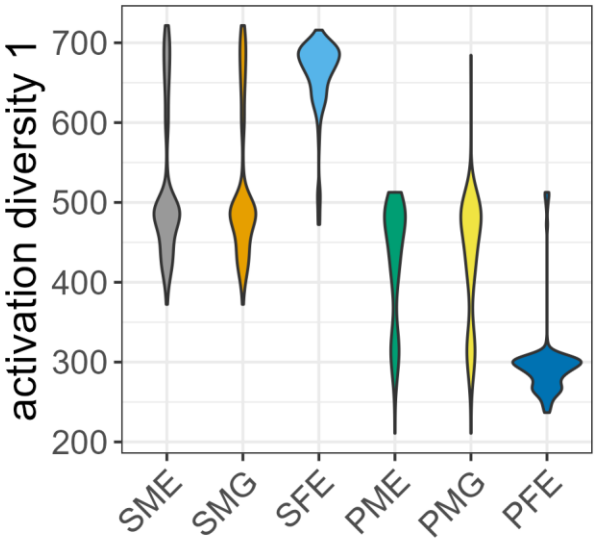
Results 1: LDL Measures

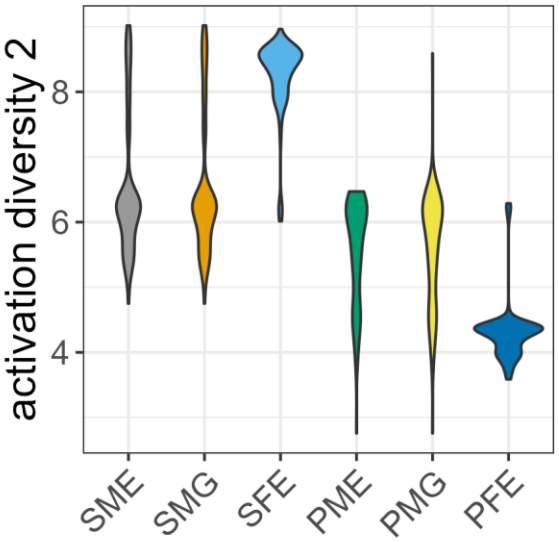




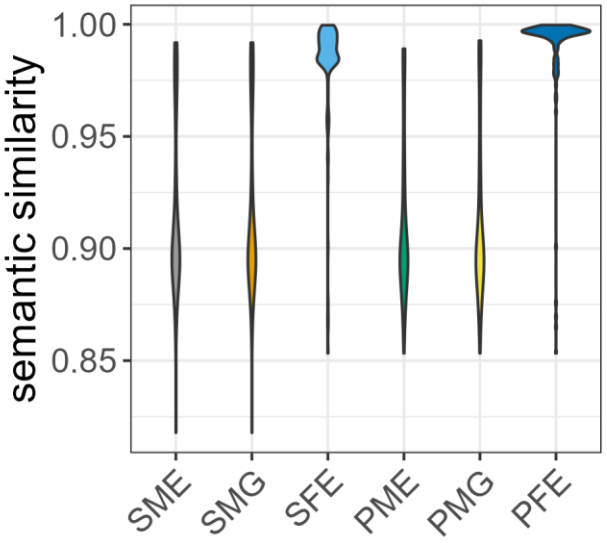


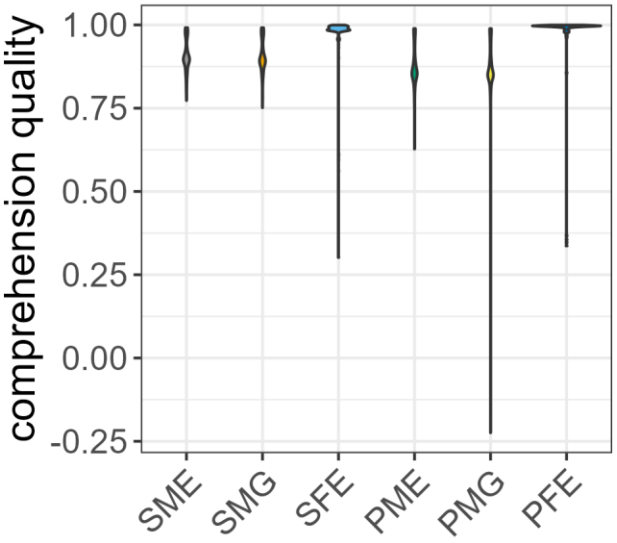
Results 1: LDL Measures

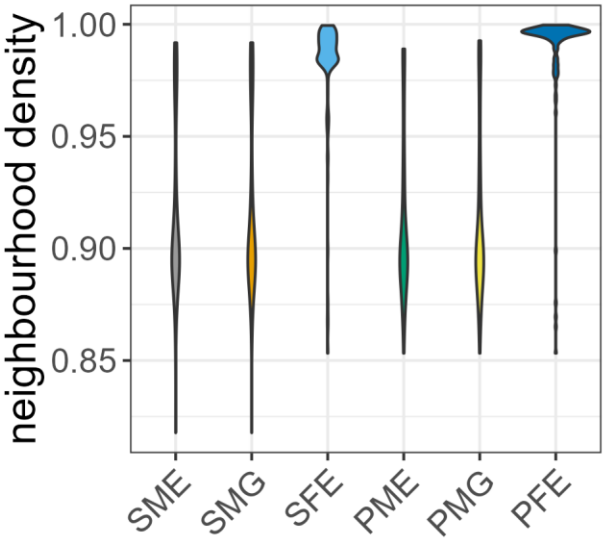


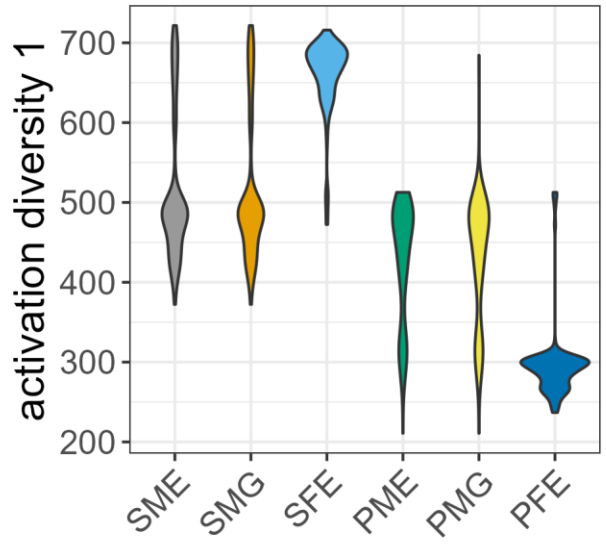


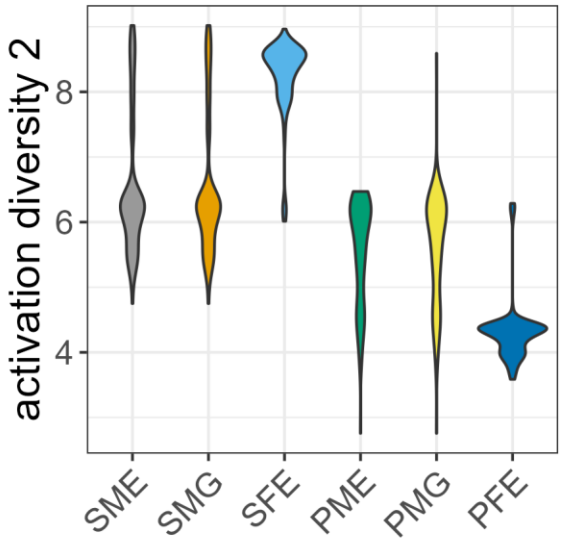
Results 1: LDL Measures











Results 1: LDL Measures

- significant differences mostly found between masculine and feminine forms
- feminine singular and feminine plurals show mostly similar patterns

	singular			plural		
	generic masculine	explicit masculine	explicit feminine	generic masculine	explicit masculine	explicit feminine

Results 1: LDL Measures

- significant differences only found between masculine and feminine forms
- feminine singular and feminine plurals show mostly similar patterns

	singular			plural		
	generic masculine	explicit masculine	explicit feminine	generic masculine	explicit masculine	explicit feminine
total semantic similarity	identical			nearly identical		

Results 1: LDL Measures

- significant differences only found between masculine and feminine forms
- feminine singular and feminine plurals show mostly similar patterns

	singular			plural		
	generic masculine	explicit masculine	explicit feminine	generic masculine	explicit masculine	explicit feminine
total semantic similarity	identical			nearly identical		
comprehension quality	nearly identical			similar		

Results 1: LDL Measures

- significant differences only found between masculine and feminine forms
- feminine singular and feminine plurals show mostly similar patterns

	singular			plural		
	generic masculine	explicit masculine	explicit feminine	generic masculine	explicit masculine	explicit feminine
total semantic similarity	identical			nearly identical		
comprehension quality	nearly identical			similar		
neighbourhood density	identical			nearly identical		

Results 1: LDL Measures

- significant differences only found between masculine and feminine forms
- feminine singular and feminine plurals show mostly similar patterns

	singular			plural		
	generic masculine	explicit masculine	explicit feminine	generic masculine	explicit masculine	explicit feminine
total semantic similarity	identical			nearly identical		
comprehension quality	nearly identical			similar		
neighbourhood density	identical			nearly identical		
activation diversity 1	identical			nearly identical		

Results 1: LDL Measures

- significant differences only found between masculine and feminine forms
- feminine singular and feminine plurals show mostly similar patterns

	singular			plural		
	generic masculine	explicit masculine	explicit feminine	generic masculine	explicit masculine	explicit feminine
total semantic similarity	identical			nearly identical		
comprehension quality	nearly identical			similar		
neighbourhood density	identical			nearly identical		
activation diversity 1	identical			nearly identical		
activation diversity 2	identical			nearly identical		

Results 2: Stereotypicality

- STEREOTYPICALITY ratings of target words (Gabriel et al., 2008) included as predictor for LDL measures
- higher value of STEREOTYPICALITY = more stereotypically male

$$LDL\_measure \sim type + stereotypicality$$

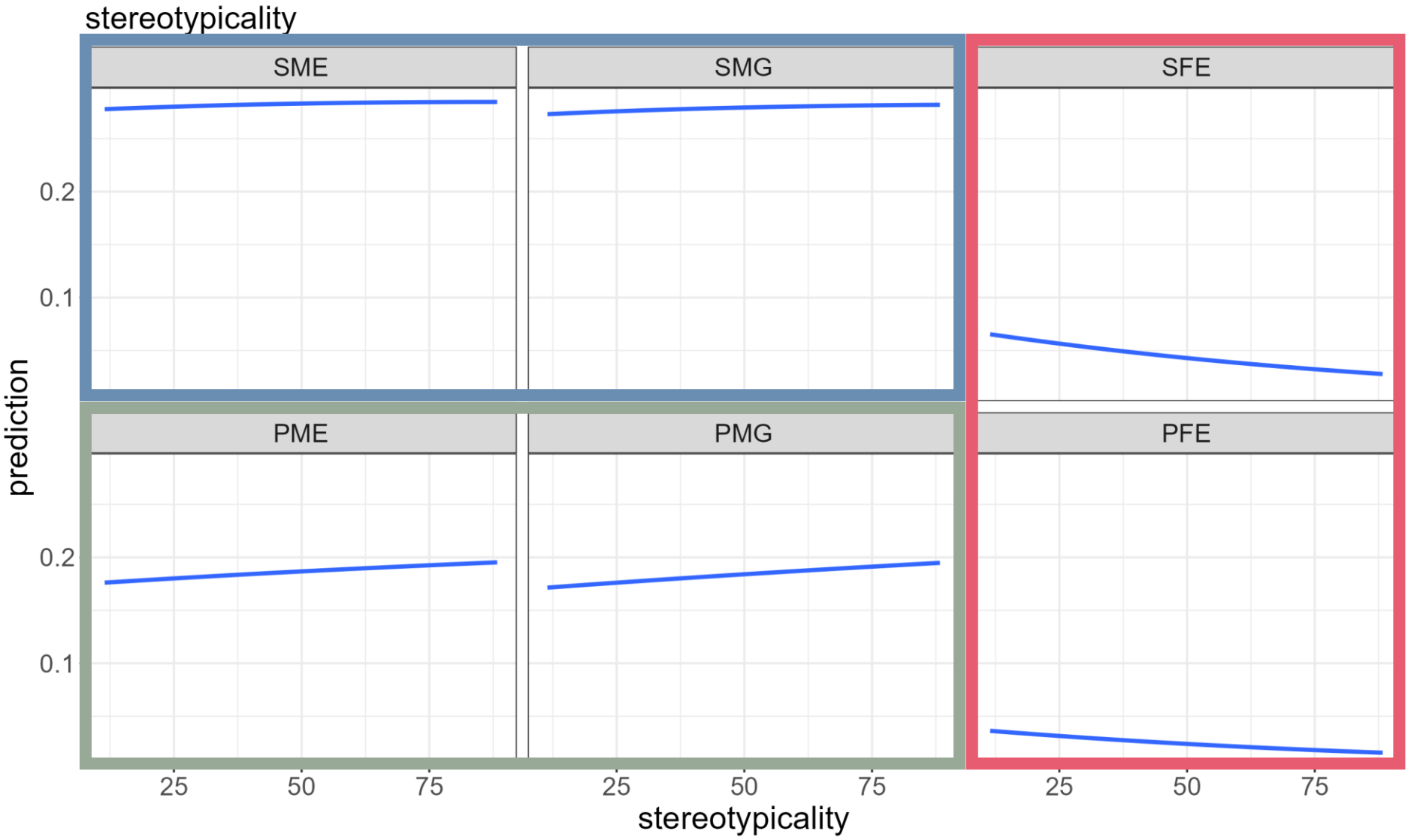
- effects found for TOTAL SEMANTIC SIMILARITY & SEMANTIC NEIGHBOURHOOD DENSITY
- no effects found for COMPREHENSION QUALITY & SEMANTIC ACTIVATION DIVERSITIES
- apparently, some LDL measures are influenced by STEREOTYPICALITY while others are not, **but:** does this play a role in the comprehension of generics?

Results 3: Prediction of Type

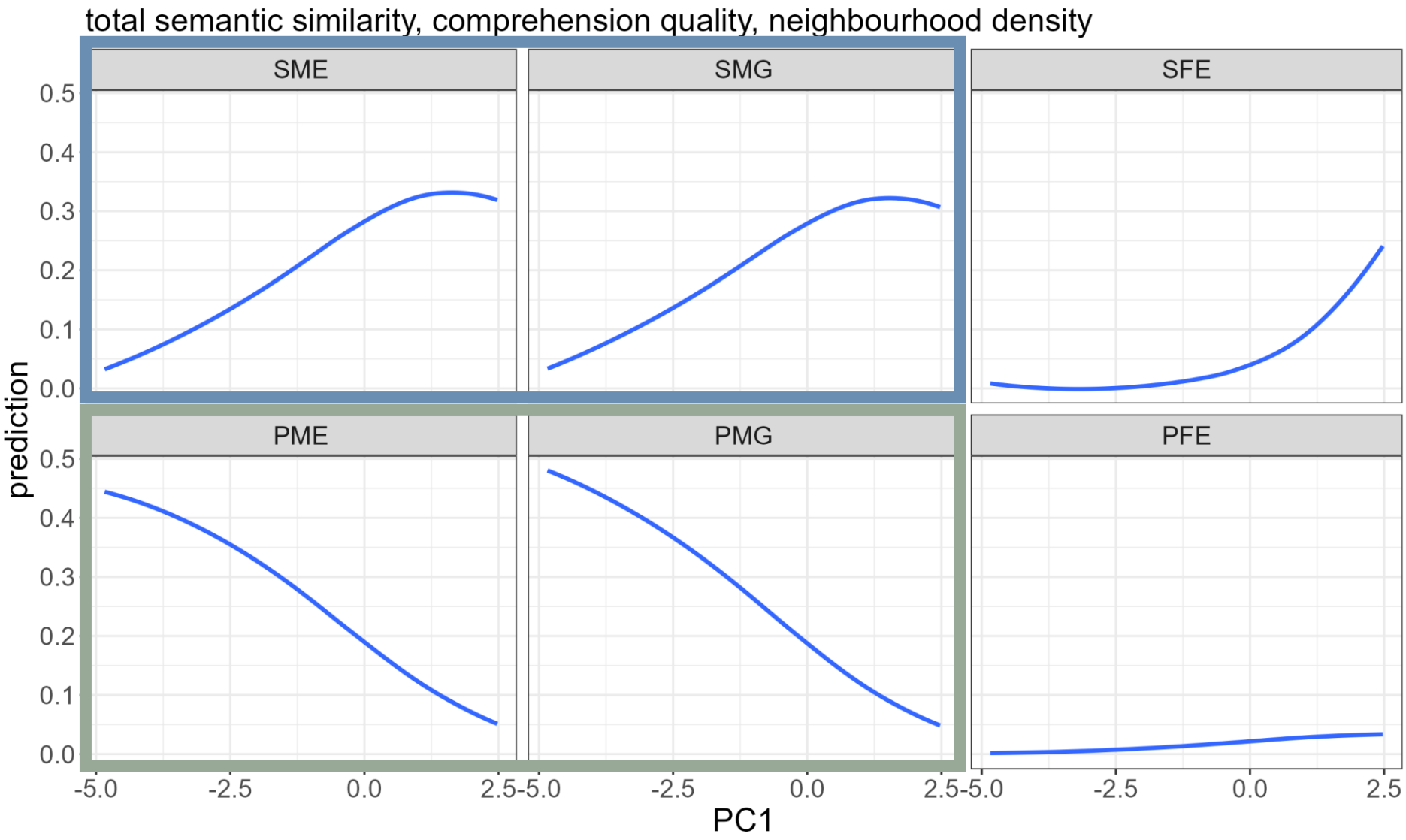
- TYPE predicted by STEREOTYPICALITY ratings of target words (Gabriel et al., 2008) and LDL measures via multinomial logistic regression
- as LDL measures are highly correlated with each other, they are first combined into two principal components
  - **PC1** total semantic similarity, comprehension quality, neighbourhood density
    higher = higher similarity/quality/density
  - **PC2** activation diversity 1 & 2
    higher = lower activation diversity

$$type \sim stereotypicality + PC1 + PC2$$

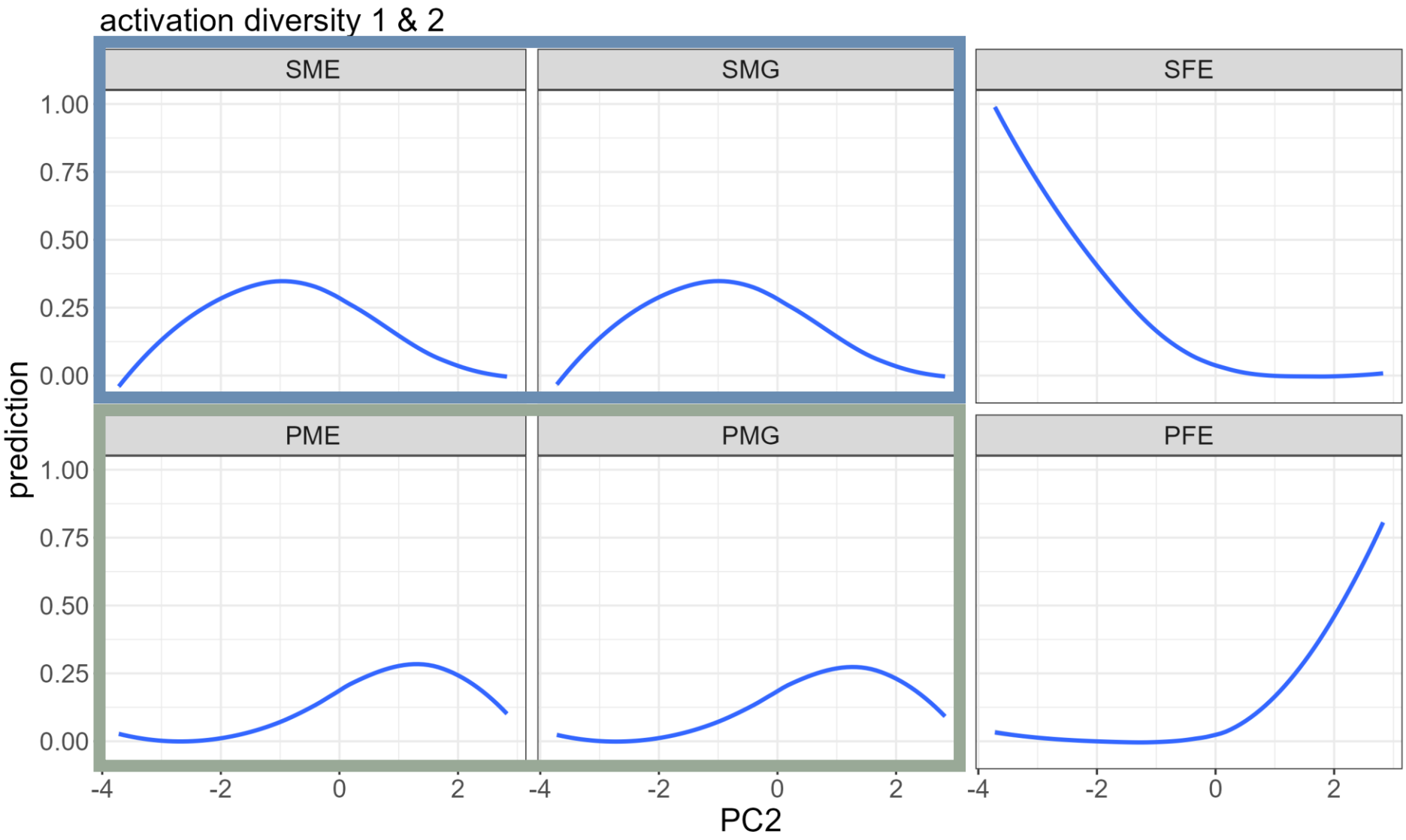
Results 3: Prediction of Type



Results 3: Prediction of Type



Results 3: Prediction of Type



Results 3: Prediction of Type

$$type \sim stereotypicality + PC1 + PC2$$

- no effect of stereotypicality found
- significant effects found for
  - **PC1** opposite patterns for masculines by number; feminines ???
    total semantic similarity, comprehension quality, neighbourhood density
    higher = higher similarity/quality/density
  - **PC2** higher = feminine singular; lower = feminine plural; masculine in-between
    activation diversity 1 & 2
    higher = lower activation diversity

Interim Summary

Q2 Which features of the underlying representations lead to the (dis)similarities of masculine and feminine forms?

YES: all semantic LDL measures

NO: stereotypicality

Discussion

- our findings are in line with assumptions found in previous research
  - Stahlberg et al. (2001)

    masculine gender of generics has a semantic component of “maleness”
  - Irmen & Linner (2005)

    semantic similarity of masculine generics and explicits due to their resonance with the lexicon and each other
  - Gygax et al. (2012) and Gygax et al. (2021)

    masculine generics activate the underlying representations of masculine explicits, leading to a semantic activation of masculine explicits, thus a male bias

Conclusion

- masculine generics show a clear male bias
- the male bias is due to the similar semantic features of the masculine generic and masculine explicit forms
- this leads to a 'male bias' in the language system itself
- thus, our findings confirm the bias found in previous behavioural studies (e.g. Demarmels, 2017; Garnham et al., 2012; Gygax et al., 2008; Irmen & Kurovskaja, 2010; Irmen & Linner, 2005; Koch, 2021; Misersky et al., 2019; Stahlberg & Sczesny, 2001; Trutkowski, 2018)
- future research will show
  - what exact effects this bias has on comprehension and/or production
  - whether the LDL measures computed for our data are predictive of behavioural measures
  - how (new) neutral forms perform (e.g. *Anwält*innen*, *AnwältInnen*)

Thank you!

References

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, 1–39. https://doi.org/10.1155/2019/4895891

Baayen, R. H., & Ramscar, M. (2015). Abstraction, storage and naive discriminative learning. *Handbook of Cognitive Linguistics, 39*, 100–120. https://doi.org/10.1515/9783110292022-006

Demarmels, S. (2017). „Gesucht: Assistentin oder Sekretär der Geschäftsleitung“ – Gendersensitive Formulierungen in Stellenanzeigen aus der Perspektive der Textsorte. In *Stellenanzeigen als Instrument des Employer Branding in Europa*. https://doi.org/10.1007/978-3-658-12719-0_11

Doleschal, U. (2002). Das generische Maskulinum im Deutschen. Ein historischer Spaziergang durch die deutsche Grammatikschreibung von der Renaissance bis zur Postmoderne. *Linguistik Online, 11(2)*. https://doi.org/10.13092/lo.11.915

Garnham, A., Gabriel, U., Sarrasin, O., Gygax, P., & Oakhill, J. (2012). Gender Representation in Different Languages and Grammatical Marking on Pronouns: When Beauticians, Musicians, and Mechanics Remain Men. *Discourse Processes, 49(6)*, 481–500. https://doi.org/10.1080/0163853X.2012.688184

Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. *Proceedings of the 8th International Language Resources and Evaluation (LREC'12)*.

Gygax, P., & Gabriel, U. (2008). Can a group of musicians be composed of women? Generic interpretation of French masculine role names in the absence and presence of feminine forms. *Swiss Journal of Psychology, 67(3)*, 143–151. https://doi.org/10.1024/1421-0185.67.3.143

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes, 23(3)*, 464–485. https://doi.org/10.1080/01690960701702035

Gygax, P., Sato, S., Öttl, A., & Gabriel, U. (2021). The masculine form in grammatically gendered languages and its multiple interpretations: a challenge for our cognitive system. *Language Sciences, 83,* 101328. https://doi.org/10.1016/j.langsci.2020.101328

References

Harris, Z. S. (1954). Distributional Structure. *WORD, 10(2–3)*, 146–162. https://doi.org/10.1080/00437956.1954.11659520

Irmen, L., & Kurovskaja, J. (2010). On the semantic content of grammatical gender and its impact on the representation of human referents. *Experimental Psychology, 57(5)*, 367–375. https://doi.org/10.1027/1618-3169/a000044

Irmen, L., & Linner, U. (2005). Die Repräsentation generisch maskuliner Personenbezeichnungen. *Zeitschrift Für Psychologie / Journal of Psychology, 213(3)*, 167–175. https://doi.org/10.1026/0044-3409.213.3.167

Koch, M. (2021). *Kognitive Effekte des generischen Maskulinums und genderneutraler Alternativen im Deutschen – eine empirische Untersuchung*. Master's Thesis. Technische Universität Braunschweig.

Misersky, J., Majid, A., & Snijders, T. M. (2019). Grammatical Gender in German Influences How Role-Nouns Are Interpreted: Evidence from ERPs. *Discourse Processes, 56(8)*, 643–654. https://doi.org/10.1080/0163853X.2018.1541382

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

Schmid, H. (1999). Improvements in part-of-speech tagging with an application to German. In S. Armstrong, K. Church, P. Isabelle, S. Manzi, E. Tzoukermann, & D. Yarowsky (Eds.), *Natural language processing using very large corpora* (pp. 13–25). Springer. https://doi.org/10.1007/978-94-017-2390-9_2

Stahlberg, D., & Sczesny, S. (2001). Effekte des generischen Maskulinums und alternativer Sprachformen auf den gedanklichen Einbezug von Frauen. *Psychologische Rundschau, 52(3)*, 131–140. https://doi.org/10.1026//0033-3042.52.3.131

Stahlberg, D., Sczesny, S., & Braun, F. (2001). Name Your Favorite Musician. *Journal of Language and Social Psychology, 20(4)*, 464–469. https://doi.org/10.1177/0261927X01020004004

Trutkowski, E. (2018). Wie generisch ist das generische Maskulinum? Über Genus und Sexus im Deutschen. *ZAS Papers in Linguistics, 59*, 83–96. https://doi.org/10.21248/zaspil.59.2018.437

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in pavlovian conditioning: Application of a theory. In R. A. Boakes & M. S. Halliday (Eds.), *Inhibition and learning* (pp. 301–334). Academic Press Inc.