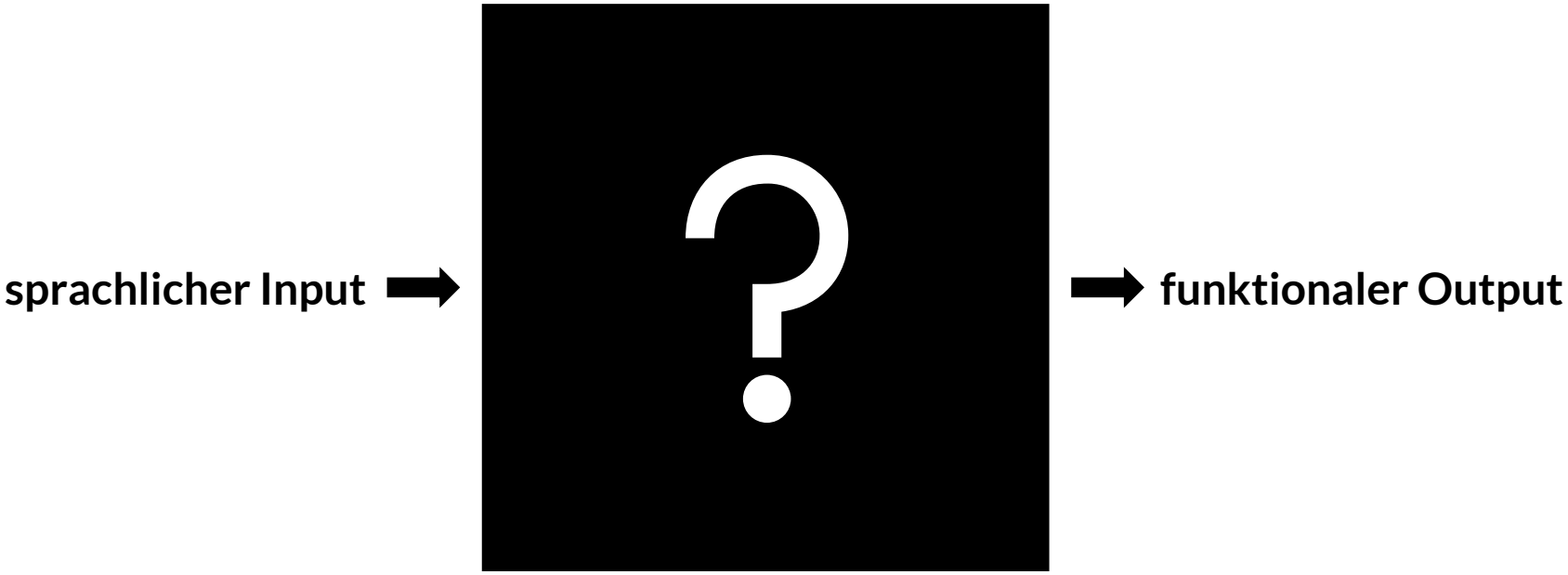
Von Algorithmen zu Geschlechterfragen: Wie computerlinguistische Methoden die Genderlinguistik bereichern

Dominic Schmitz

Heinrich-Heine-Universität Düsseldorf

GAL Research School
Empirische Methoden in der Genderlinguistik
11.-13. Oktober 2023, IDS Mannheim

Computerlinguistische Anwendungen



Beispiele für **Output**: Dialoganwendungen (z. B. Bestellsysteme), Thesauri, Suchen (z. B. Google), Übersetzungssysteme, Silbentrennung, Rechtschreibprüfung, Spracherkennung, Informationsextraktion, Sprach- und Textverarbeitung

Computerlinguistische ~~Anwendungen~~ Methoden

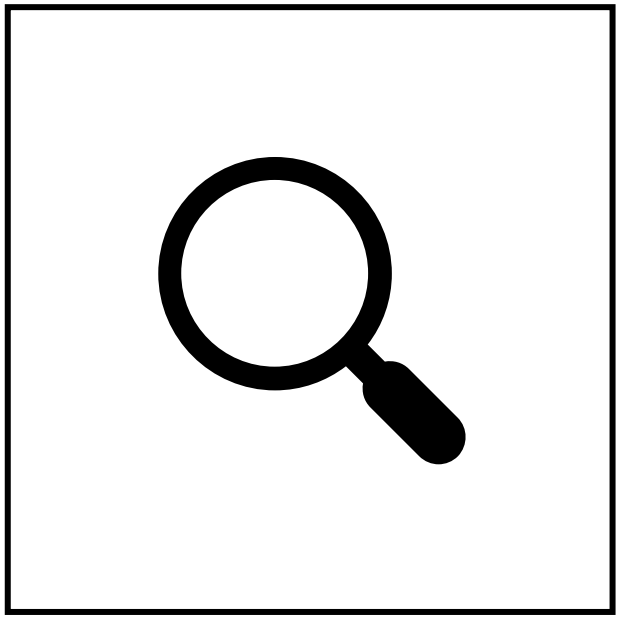
sprachlicher Input →

→ formale berechenbare Modelle natürlicher Sprache bzw. deren Output zur Untersuchung sprachlicher Phänomene

Beispiele für die **Blackbox**: artificial neural networks (z. B. fastText, Word2Vec, GloVe), deep learning (z. B. encoder decoder models), large language models (z. B. ChatGPT)

Computerlinguistische ~~Anwendungen~~ Methoden

sprachlicher Input



verschiedene Vektoren und Maße zur Untersuchung sprachlicher Phänomene

Beispiele für die Whitebox: semantische Vektoren (z. B. Naive Discriminative Learning, Instance-Vektoren), phonologische Vektoren (z. B. Linear Discriminative Learning), Assoziationsmatrizen (z. B. Linear Discriminative Learning)

Computerlinguistische Methoden

- was es bereits seit längerer Zeit gibt:
  computerlinguistische Forschung zu verschiedenen Biases, u.a. auch zu Biases gegenüber Geschlecht
- was es erst seit kurzer Zeit gibt:
  computerlinguistische Methoden in konkret genderlinguistischer Forschung
- wir schauen uns heute die existierenden Studien an, die teils von **Blackbox-** und überwiegend von Whitebox-Methoden Gebrauch machen
- der Fokus wird dabei auf den Methoden und weniger auf den theoretischen Implikationen liegen

Studien

1. Semantik generischer Maskulina im Deutschen (Schmitz, 2023)

   Naive Discriminative Learning

2. Semantik generischer Maskulina im Deutschen (Schmitz, *in Arbeit*)

   fastText, Instance-Vektoren

3. Verständnis generischer Maskulina im Deutschen (Schmitz et al., 2023)

   Naive Discriminative Learning, Linear Discriminative Learning

4. Semantik englischer Personalpronomen (Schmitz, 2023)

   Naive Discriminative Learning, Instance-Vektoren, Linear Discriminative Learning

Studie 1 | Schmitz, 2023

Semantik generischer Maskulina im Deutschen

Naive Discriminative Learning

Studie 1 | Schmitz, 2023

Semantik generischer Maskulina im Deutschen

Naive Discriminative Learning

Hintergrund

Hintergrund

- männliches Bias generischer Maskulina im Deutschen (z. B. 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)
- obwohl generische Maskulina also durchaus geschlechtsneutral intendiert sein können, kommt diese Geschlechtsneutralität im Sprachverständnis nicht (völlig) an
- stattdessen findet eine Verarbeitung statt, aus der eine männliche Überrepräsentation resultiert

Hintergrund

- Probleme bisheriger, überwiegend psycholinguistischer Forschung
  - viele Studien nutzen Daten, die speziell für ihre jeweiligen Untersuchungen erhoben wurden
  - das maskuline Bias könnte ein Resultat der eingesetzten Methoden sein

- Lösungsvorschlag
  - Einsatz computationeller Methoden zur Überprüfung der Existenz des maskulinen Bias generischer Maskulina
  - hier: Naive Discriminative Learning

Methode

Naive Discriminative Learning

- Naive Discriminative Learning (Baayen & Ramscar, 2015) ist ein Algorithmus, der zur Berechnung semantischer Vektoren genutzt werden kann
- die Idee semantischer Vektoren geht auf die **Distributional Hypothesis** zurück (Harris, 1954)
  - ein **semantischer Unterschied** geht immer mit einem **distributionellen Unterschied** einher
  - Beispiel:
    Jacke und ***Hose*** sind semantisch ähnlicher als ***Jacke*** und ***Apfel***, da ***Jacke*** und ***Hose*** regelmäßig in ähnlichen Kontexten vorkommen

Naive Discriminative Learning

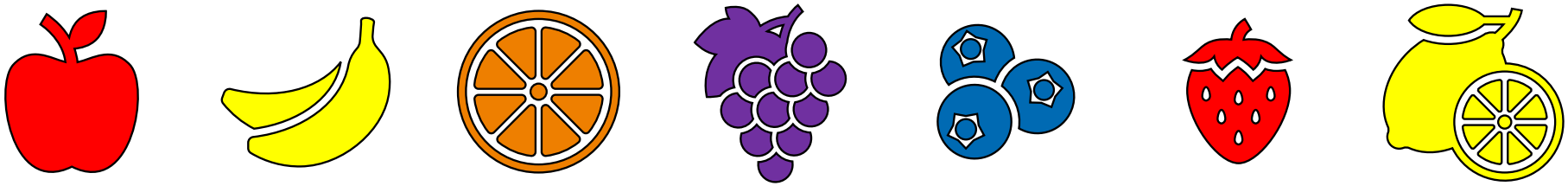
- Naive Discriminative Learning wurde u.a. bereits erfolgreich eingesetzt zur Untersuchung von/vom…
  - Flexion von Pseudowörtern im Maltesischen (Nieder et al., 2022)
  - Pluralsystem des Maltesischen (Nieder & Tomaschek, 2023)
  - Wortakzente im Deutschen (Tomaschek et al., 2023)
  - Wortakzente im Englischen (Arndt-Lappe et al., 2022)
  - akustischer Dauer von wort-finalem /s/ im Englischen (Tomaschek et al., 2019)

Naive Discriminative Learning

- Naive Discriminative Learning beruht auf dem Rescorla-Wagner-Modell, einem etablierten Lern-Modell aus der Kognitionspsychologie (Rescorla & Wagner, 1972)
- Naive Discriminative Learning folgt dabei zwei fundamentalen Regeln
  1. **Outcomes** (z. B. Wortformen/Flexionsmerkmale) werden durch **Cues** (z. B. Wortformen/Flexionsmerkmale) vorhergesagt
  2. Die Assoziationsstärke zwischen einem **Outcome** und einem **Cue** wird stets durch einen einzelnen Wert dargestellt
- alle **Outcomes** werden anhand aller **Cues** im selben Satz/Absatz/Text vorhergesagt

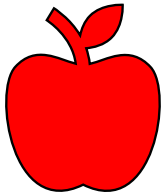
Naive Discriminative Learning

Beispiel: Obst

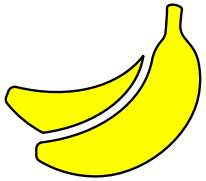


Naive Discriminative Learning

Beispiel: Obst

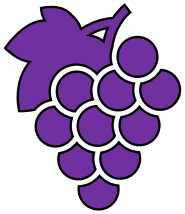


rot
süß
rund

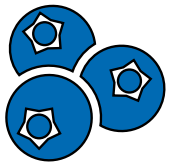


gelb
süß
lang

orange
sauer
rund

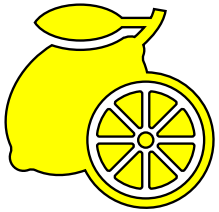


lila
süß
rund



blau
süß
rund

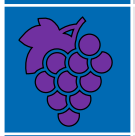
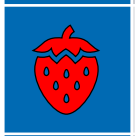
rot
süß
rund
lang



gelb
sauer
rund
lang

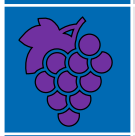
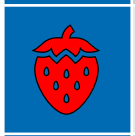
Naive Discriminative Learning

Beispiel: Obst

	rot	gelb	orange	lila	blau	süß	sauer	rund	lang
	1					1		1	
		1				1			1
			1				1	1	
				1		1		1	
					1	1		1	
	1					1			1
		1					1	1	1

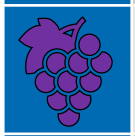
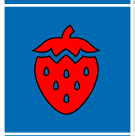
Naive Discriminative Learning

Beispiel: Obst

	rot	gelb	orange	lila	blau	süß	sauer	rund	lang
	30					30		30	
		15				15			15
			18				18	18	
				10		10		10	
					5	5		5	
	45					45		45	45
		20					20	20	20

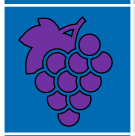
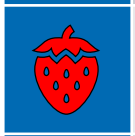
Naive Discriminative Learning

Beispiel: Obst

	rot	gelb	orange	lila	blau	süß	sauer	rund	lang
	29	1				30		30	
		15				15			15
			18				18	18	
				10		10		10	
					5	5		5	
	45					45		45	45
		20					20	20	20

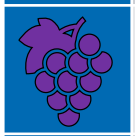
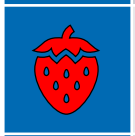
Naive Discriminative Learning

Beispiel: Obst

	rot	gelb	orange	lila	blau	süß	sauer	rund	lang
	29	1				29	1	30	
		15				15			15
			18				18	18	
				10		10		10	
					5	5		5	
	45					45		45	45
		20					20	20	20

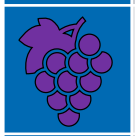
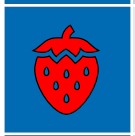
Naive Discriminative Learning

Beispiel: Obst

	rot	gelb	orange	lila	blau	süß	sauer	rund	lang
	29	1	-1	-3	-2	29	1	30	-1
		15				15			15
			18				18	18	
				10		10		10	
					5	5		5	
	45					45		45	45
		20					20	20	20

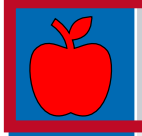
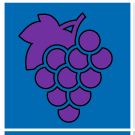
Naive Discriminative Learning

Beispiel: Obst

	rot	gelb	orange	lila	blau	süß	sauer	rund	lang
	29	1	-1	-3	-2	29	1	30	-1
	-10	15	-10	-8	-6	15	-11	-5	15
	-6	-7	18	-14	-15	3	15	18	-2
	-5	-1	-6	10	-9	5	5	10	-7
	-6	-9	-19	2	3	4	1	5	-5
	45	-6	-9	-14	-1	25	20	45	45
	-1	20	-5	-6	-8	-4	20	20	20

Naive Discriminative Learning

Beispiel: Obst

	rot	gelb	orange	lila	blau	süß	sauer	rund	lang
	29	1	-1	-3	-2	29	1	30	-1
	-10	15	-10	-8	-6	15	-11	-5	15
	-6	-7	18	-14	-15	3	15	18	-2
	-5	-1	-6	10	-9	5	5	10	-7
	-6	-9	-19	2	3	4	1	5	-5
	45	-6	-9	-14	-1	25	20	45	45
	-1	20	-5	-6	-8	-4	20	20	20

Naive Discriminative Learning

Beispiel: Obst

	rot	gelb	orange	lila	blau	süß	sauer
Apfel	29	1	-1	-3	-2	29	1
Banane	-10	15	-10	-8	-6	15	-11
Orange	-6	-7	18	-14	-15	3	15
Traube	-5	-1	-6	10	-9	5	5
Blaubeere	-6	-9	-19	2	3	4	1
Erdbeere	45	-6	-9	-14	-1	25	20
Zitrone	-1	20	-5	-6	-8	-4	20

Naive Discriminative Learning

Beispiel: Obst

- möchten wir nun wissen, welches Obst welchem anderen Obst semantisch am ähnlichsten ist, kann man dies über die **Kosinus-Ähnlichkeit** berechnen
  - Vergleich der Winkel zweier Vektoren
  - je höher der Wert, desto ähnlicher die Vektoren
  - da Vektoren hier Semantik repräsentieren, bedeuten hohe Werte eine ähnliche Semantik

	Banane	Orange	Traube	Blaubeere	Erdbeere	Zitrone
Apfel	0.160	-0.019	-0.003	-0.064	0.877	-0.052

Naive Discriminative Learning

- zurück zur eigentlichen Studie…
- NDL-Implementierungen liegen i.d.R. annotierte Korpora zugrunde
- für die vorliegende Studie
  - eigens auf Flexionsmerkmale & Generizität annotiertes Korpus aus Teilen des Subkorpus „News“ der Leipzig Corpora Collection (Goldhahn et al., 2012)
  - ca. 49 Millionen Tokens
  - 30.000 Sätze mit Rollennomen = Zielwörter
  - 800.000 Sätze ohne Rollennomen
- 113 Zielwörter übernommen aus einer vorherigen Studie, in welcher es um die Stereotypizität von Rollenbezeichnungen ging (Gabriel et al., 2008)

Naive Discriminative Learning

- was hat NDL mit dem Korpus konkret gemacht? denken wir an die zwei fundamentalen Regeln zurück:
  1. **Outcomes** (z.B. Wortformen /Flexion) werden durch **Cues** (z.B. Wortformen/Flexion) vorhergesagt
     → **Obstsorten** & **Farbe/Geschmack/Form**
     → **Stamm/Flexionsmerkmale** & **Stamm/Flexionsmerkmale/Generizität**
  2. Die Assoziationsstärke zwischen einem **Outcome** und einem **Cue** wird stets durch einen Wert dargestellt

	rot	gelb	orange	lila	blau	süß	sauer
Apfel	29	1	-1	-3	-2	29	1

Naive Discriminative Learning

- als Resultat erhalten wir semantische Vektoren für Stämme von Inhaltswörtern, Flexionsmerkmale, Funktionswörter, Generizität
- nun müssen wir die Semantik unserer Zielwörter zunächst erstellen

Beispiel: *Ärztin*

$$\overrightarrow{Ärztin} = \overrightarrow{Arzt} + \overrightarrow{Singular} + \overrightarrow{Femininum} + \overrightarrow{Nominativ} + \overrightarrow{Spezifisch}$$

Beispiel: *Arzt* (generisch)

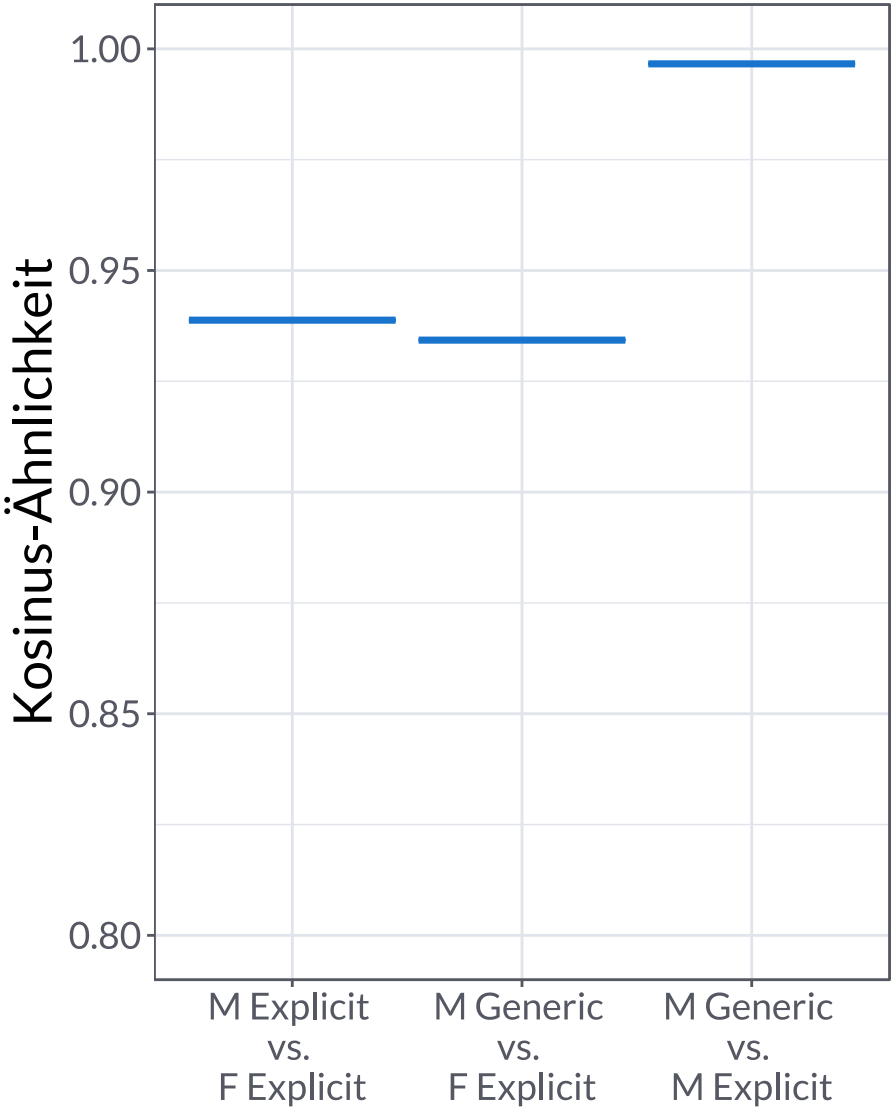
$$\overrightarrow{Arzt} = \overrightarrow{Arzt} + \overrightarrow{Singular} + \overrightarrow{Maskulinum} + \overrightarrow{Nominativ} + \overrightarrow{Generisch}$$

Analyse & Ergebnisse

Analyse

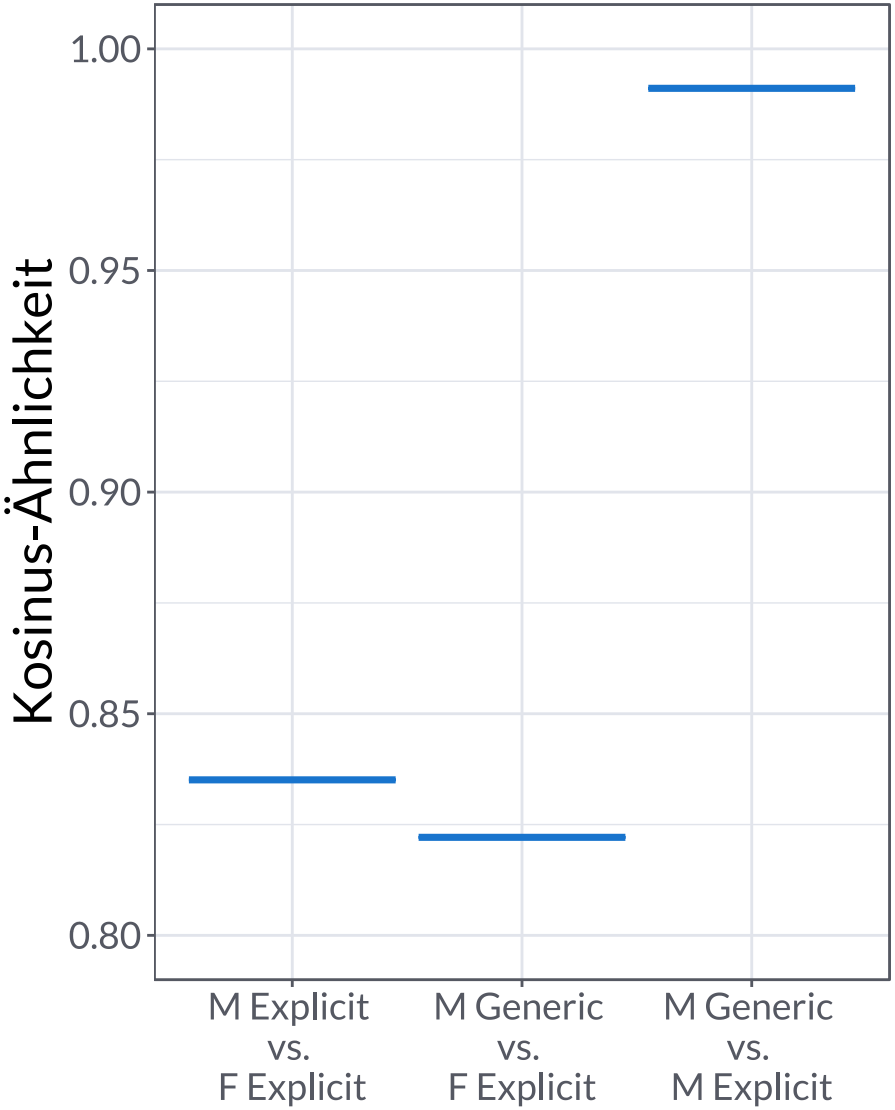
- anschließend können die erstellten Vektoren dann anhand ihrer **Kosinus-Ähnlichkeiten** verglichen werden
- hierzu wird Betaregression als statistisches Mittel der Wahl eingesetzt, um andere potenzielle Einflüsse (z. B. Stereotypizität) zu berücksichtigen
- zur Erinnerung
  - hoher Wert → Vektoren ähnlich → hohe semantische Ähnlichkeit
  - niedriger Wert → Vektoren unähnlich → niedrige semantische Ähnlichkeit

Ergebnisse: Singular



- **generische** und **spezifische Maskulina** sind semantisch am ähnlichsten
- **spezifische Feminina** sind **spezifischen Maskulina** ähnlicher als **generischen Maskulina**

Ergebnisse: Plural



- **generische** und **spezifische Maskulina** sind semantisch am ähnlichsten
- **spezifische Feminina** sind **spezifischen Maskulina** ähnlicher als **generischen Maskulina**
- Unterschiede sind hier noch deutlicher als im Singular

Resümee

- Problem bisheriger, überwiegend psycholinguistischer Forschung
  - viele Studien nutzen Daten, die speziell für ihre jeweiligen Untersuchungen erhoben wurden
  - das maskuline Bias könnte ein Resultat der genutzten Methoden sein
- Lösungsvorschlag
  - Einsatz computationeller Methoden zur Überprüfung der Existenz des maskulinen Bias generischer Maskulina
  - hier: Naive Discriminative Learning
- Ergebnis: das männliche Bias existiert auch in semantischen Vektoren

Studie 2 | Schmitz, *in Arbeit*

Semantik generischer Maskulina im Deutschen

fastText, Instance-Vektoren

Studie 2 | Schmitz, *in Arbeit*

Semantik generischer Maskulina im Deutschen

fastText, Instance-Vektoren

Hintergrund

Hintergrund

- männliches Bias generischer Maskulina im Deutschen (z. B. 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)

- Problem der NDL-Studie
  - Generizität wird als Flexionsmerkmal behandelt, obwohl es keines ist

- Lösungsvorschlag
  - Einsatz alternativer computationeller Methoden zur Überprüfung der Ergebnisse der NDL-Studie
  - hier: **fastText** & Instance-Vektoren

Methode

Instance-Vektoren

- Instance-Vektoren (Lapesa et al., 2018) werden für jede Attestierung eines jeden Zielworts berechnet
- dabei ist jeder Instance-Vektor der Durchschnittsvektor der n vorherigen und folgenden Inhaltswörter
- Voraussetzung zur Berechnung sind daher 2 Dinge
  1. semantische Vektoren für Kontextwörter
  2. eine begründete Größe für n

Semantische Vektoren: Kontextwörter

- semantische Vektoren für Kontextwörter wurden über fastText generiert
- fastText (Bojanowski et al., 2016; Mikolov et al., 2013) ist ein Algorithmus, ein **Artificial Neural Network**, zur Berechnung semantischer Vektoren
- Datengrundlage sind Common Crawl und Wikipedia
- der Algorithmus generiert Vektoren für ganze Wörter, aber auch für n-Gramme
- aus n-Grammen können verlässliche semantische Vektoren für Wörter generiert werden, die nicht Teil der Datengrundlage waren
- fastText-Vektoren sind für 157 Sprachen verfügbar (inkl. Standarddeutsch und, witzigerweise, Pfälzisch)

Instance-Vektoren: n

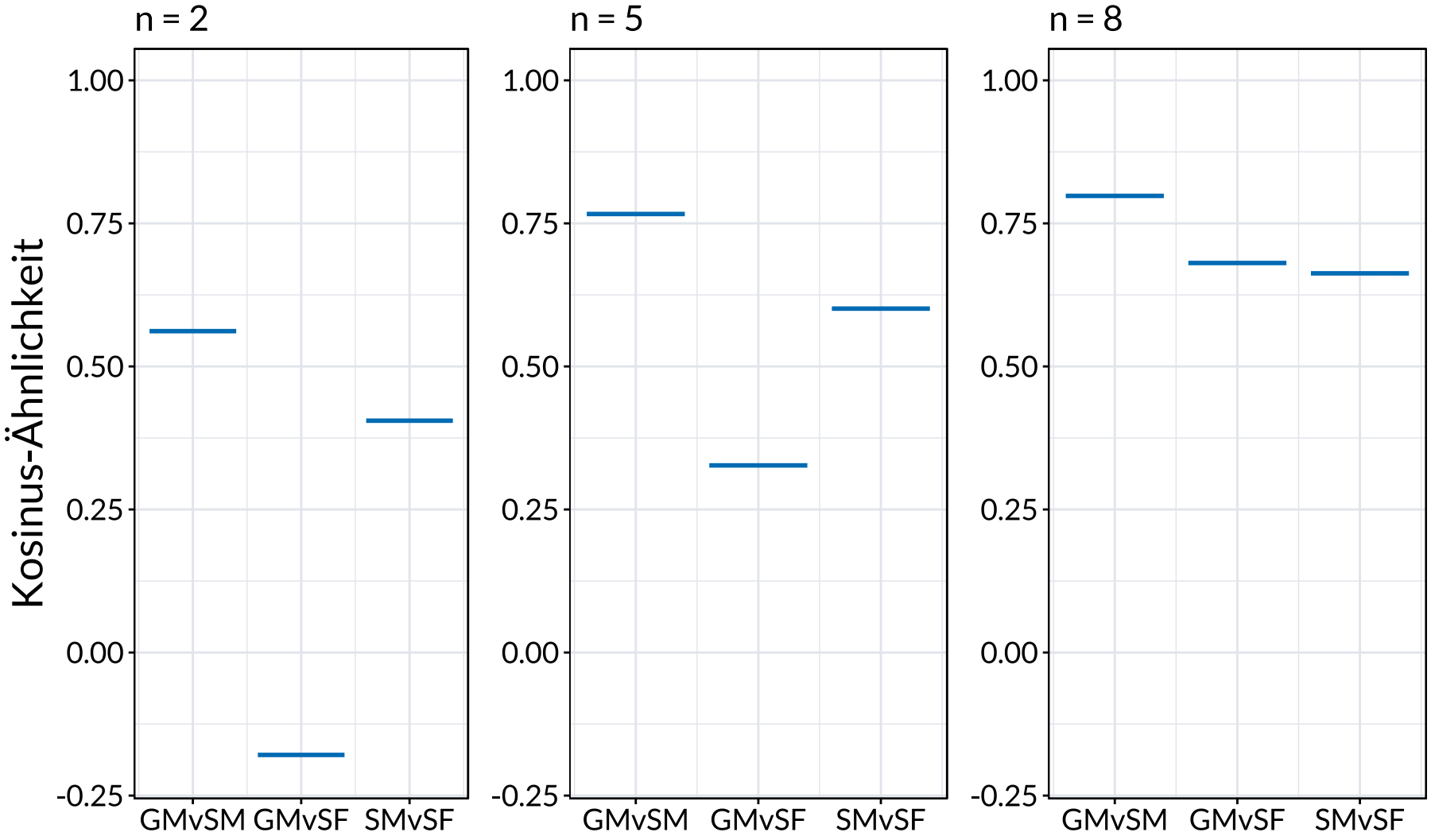
- da es bisher nur eine Publikation gibt, die Instance-Vektoren eingesetzt hat (Lapesa et al., 2018), orientiert sich die Wahl von n stark an dieser
  - $n = 2$ soll Quasi-Synonymie erfassen können
  - $n = 8$ soll thematische Ähnlichkeiten erfassen können
- zusätzlich
  - $n = 5$ als Balance zwischen Quasi-Synonymie und thematischer Ähnlichkeit

Analyse & Ergebnisse

Analyse

- als Resultat erhalten wir einen semantischen Vektor pro Zielwort-Attestierung in unserem Korpus
- anders als in der NDL-Studie müssen wir hier vor einer Analyse keine weiteren Schritte, etwa das Zusammenfügen von Vektoren, vornehmen
- die Instance-Vektoren werden ebenfalls über ihre **Kosinus-Ähnlichkeiten** miteinander verglichen
- hierzu wird erneut Betaregression als statistisches Mittel der Wahl eingesetzt, um andere potenzielle Einflüsse (Stereotypizität, Numerus, Frequenz) zu berücksichtigen

Ergebnisse


n = 2
n = 5
n = 8
Kosinus-Ähnlichkeit
1.00
0.75
0.50
0.25
0.00
-0.25
GMvSM
GMvSF
SMvSF

Resümee

- Problem der NDL-Studie
  - Generizität wird als Flexionsmerkmal behandelt, obwohl es keines ist
- Lösungsvorschlag
  - Einsatz alternativer computationeller Methoden zur Überprüfung der Ergebnisse der NDL-Studie
  - hier: fastText & Instance-Vektoren
- Ergebnis
  - auch die alternative Methode über fastText und Instance-Vektoren bestätigt die Existenz eines männlichen Bias generischer Maskulina

Studie 3 | Schmitz et al., 2023

Verständnis generischer Maskulina im Deutschen

Naive Discriminative Learning, Linear Discriminative Learning

Studie 3 | Schmitz et al., 2023

Verständnis generischer Maskulina im Deutschen

Naive Discriminative Learning, Linear Discriminative Learning

Hintergrund

Hintergrund

- männliches Bias generischer Maskulina im Deutschen (z. B. 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)

- Manko der NDL- und Instance-Vektor-Studien
  - Generizität als Teil der Wortsemantik behandelt, aber isoliert vom mentalen Lexikon betrachtet

- Lösungsvorschlag
  - Implementierung von Linear Discriminative Learning zur Simulation des mentalen Lexikons und der Verständnisvorgänge in diesem

Methode

Linear Discriminative Learning

- Linear Discriminative Learning erzeugt zur Simulation des mentalen Lexikons ein System aus Form-Bedeutungs-Beziehungen (Baayen et al., 2019)
- wurde u.a. bereits erfolgreich eingesetzt zur Untersuchung von/vom…
  - Pluralsystem des Maltesischen (Nieder et al., 2023)
  - Semantik von Pseudowörtern (Chuang et al., 2021; Schmitz et al., 2021)
  - akustischer Dauer von wort-finalem /s/ im Englischen (Schmitz et al., 2021)
  - akustischer Dauer von Stämmen komplexer Wörter im Englischen (Stein & Plag, 2021)
- fundamentale Bestandteile des simulierten Lexikons sind dabei
  1. Semantik, repräsentiert durch semantische Vektoren
  2. Wortformen, repräsentiert durch Form-Vektoren

Linear Discriminative Learning: Semantik

- die für eine Implementierung von LDL benötigten semantischen Vektoren können von verschiedensten Algorithmen stammen
- im vorliegenden Fall sind es die semantischen Vektoren der NDL-Implementierung, die wir bereits betrachtet haben
- zur Verwendung in der LDL-Implementierung werden die semantischen Vektoren für komplexe Wörter wie gehabt zusammengesetzt

Beispiel: *Ärztin*

$$\overrightarrow{Ärztin} = \overrightarrow{Arzt} + \overrightarrow{Singular} + \overrightarrow{Femininum} + \overrightarrow{Nominativ} + \overrightarrow{Spezifisch}$$

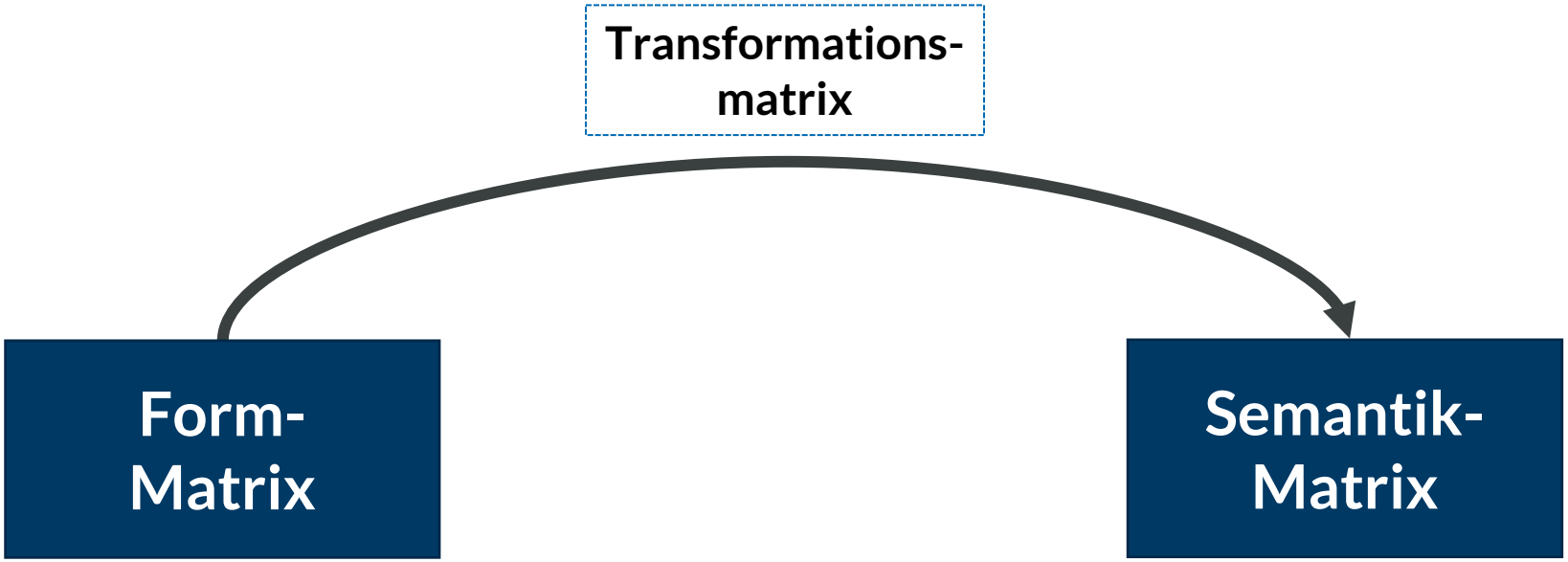
Linear Discriminative Learning: Wortformen

- die Form von Einträgen des mentalen Lexikons kann phonetisch, phonologisch oder orthografisch codiert sein
  - phonetisch: Frequency Band Summaries (z. B. Arnold et al. 2017; Shafaei-Bajestan et al. 2021)
  - phonologisch: n-Phone (z. B. Schmitz et al. 2021)
  - orthografisch: n-Gramme (z. B. Heitmeier et al., 2021)
- in der vorliegenden Studie wurden Triphone eingesetzt

Wortform	#le:	le:r	e:rA	rA#	ArI	rIn	In#
Lehrer	1	1	1	1	0	0	0
Lehrer	1	1	1	1	0	0	0
Lehrerin	1	1	1	0	1	1	1

Linear Discriminative Learning

- Verständnis (engl. *comprehension*) wird durch lineares Abbilden der Matrix der Formen auf die Matrix der semantischen Vektoren simuliert



Linear Discriminative Learning

- Verständnis (engl. *comprehension*) wird durch lineares Abbilden der Matrix der Formen auf die Matrix der semantischen Vektoren simuliert

	D1	D2	D3
#ma	11.33	-2.17	1.33
maU	11.33	-2.17	1.33
aUs	3.17	2.67	-1.83
Us#	3.17	2.67	-1.83
#ha	-8.17	4.83	-3.17
haU	-8.17	4.83	-3.17

	#ma	maU	aUs	Us#	#ha	haU
Maus	1	1	1	1	0	0
Haus	0	0	1	1	1	1

	D1	D2	D3
Maus	29	1	-1
Haus	-10	15	-10

Linear Discriminative Learning

- anders als im Beispiel ist bei hochdimensionalen Matrizen das Ergebnis nicht die „beobachtete“ semantische Matrix, sondern eine Approximation → klingt womöglich negativ, ist aber positiv!
- denn
  - die Approximation spiegelt die Zusammenhänge zwischen Form und Bedeutung im mentalen Lexikon wider
  - die Transformationsmatrix ist der „Verständnisprozess“ der Sprachverarbeitung
  - die approximierte Semantik ist die „verstandene“ Bedeutung

Linear Discriminative Learning

- aus der approximierten Matrix lassen sich semantische Maße berechnen, zum Beispiel
  - **Verständnisqualität**

    Korrelation des ursprünglichen mit dem vorhergesagten Vektor

    höher = besseres Verständnis
  - **Nachbarschaftsdichte**

    durchschnittliche Korrelation des Zielwortes mit seinen 8 nächsten Nachbarn

    höher = engere Nachbarschaft
  - **Aktivierungsvielfalt**

    Euklidische Norm des vorhergesagten Vektors

    höher = höheres Maß an Koaktivierung

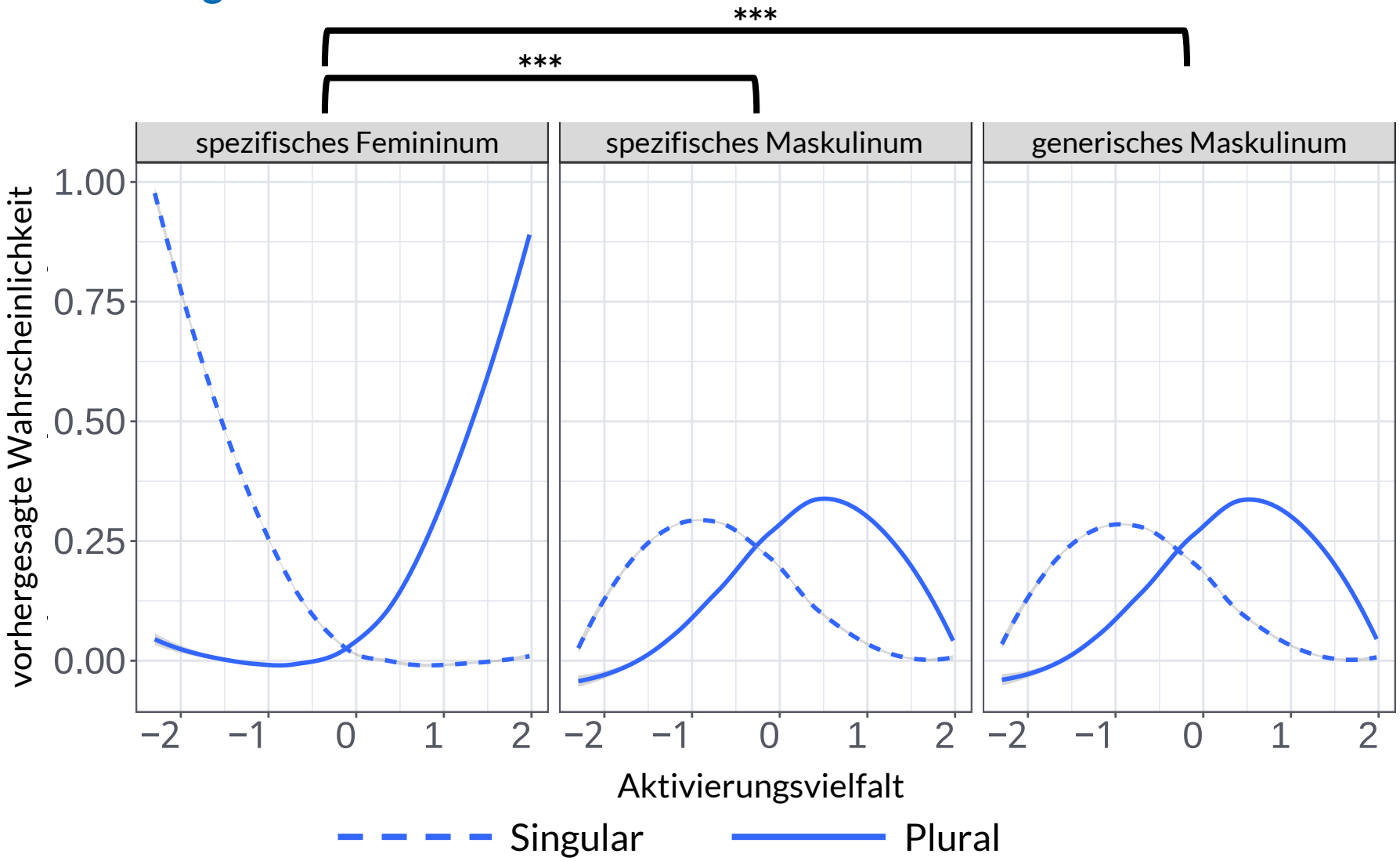
Analyse & Ergebnisse

Analyse

- sollten sich diese Maße für generische Maskulina, spezifische Maskulina und spezifische Feminina signifikant unterscheiden, so lassen sich aus den entsprechenden Unterschieden Schlüsse ziehen
- ob es Unterschiede gibt, zeigt eine multinominale Regressionsanalyse
  - abhängige Variable: Typ (Singular generisches Maskulinum, Singular spezifisches Maskulinum, Singular spezifisches Femininum, Plural generisches Maskulinum, Plural spezifisches Maskulinum, Plural spezifisches Femininum)
  - unabhängige Variablen
    - Aktivierungsvielfalt
    - Principal Component (Verständnisqualität + Nachbarschaftsdichte)

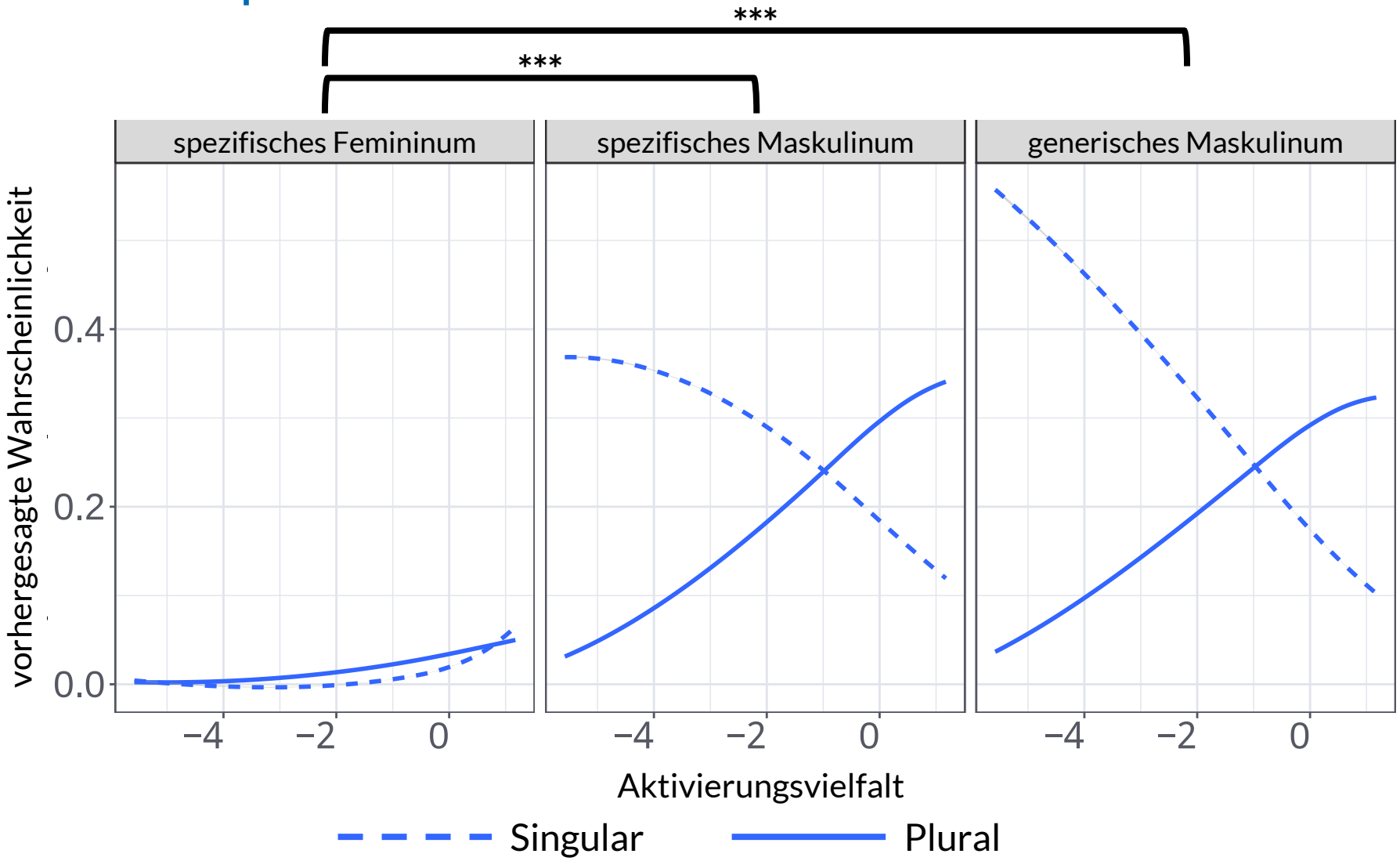
Ergebnisse

Aktivierungsvielfalt



Ergebnisse

Verständnisqualität + Nachbarschaftsdichte



Resümee

- generische Maskulina und spezifische Maskulina zeigen sehr ähnliche, nicht signifikant unterschiedliche Ausprägungen der untersuchten semantischen Maße
- spezifische Feminina hingegen zeigen signifikant andere Ausprägungen der untersuchten Maße
- da die Maße nicht nur auf der Semantik der Formen, sondern darüber hinaus auf simuliertem Sprachverständnis beruhen, scheint der Ursprung des männlichen Bias das Sprachsystem selbst zu sein

Resümee

- Manko der NDL- und Instance-Vektor-Studien
  - Generizität als Teil der Wortsemantik behandelt, aber isoliert vom mentalen Lexikon betrachtet
- Lösungsvorschlag
  - Implementierung von Linear Discriminative Learning zur Simulation des mentalen Lexikons und der Verständnisvorgänge in diesem
- Ergebnis
  - auch unter Berücksichtigung der Zusammenhänge im mentalen Lexikon wird die Existenz eines männlichen Bias generischer Maskulina bestätigt

Studie 4 | Schmitz, 2023

Semantik englischer Personalpronomen

Naive Discriminative Learning, Instance-Vektoren, Linear Discriminative Learning

Studie 4 | Schmitz, 2023

Semantik englischer Personalpronomen

Naive Discriminative Learning, Instance-Vektoren, Linear Discriminative Learning

Hintergrund

Hintergrund

- Pronomen, insbesondere Singular-*they*, im Englischen

- bisherige Studien
  - beschäftigen sich mit soziolinguistischen und syntaktischen Fragen (z. B. Conrod, 2020; Han & Moulton, 2022; Konnelly et al., 2020)
  - nicht mit der Semantik von Pronomen

- Lösungsvorschlag
  - semantische Vektoren durch eine NDL-Implementierung berechnen
  - Instance-Vektoren anhand der NDL-Vektoren berechnen
  - mit LDL das Sprachverständnis von Pronomen analysieren

Hintergrund

- heute kennt das Englische mindestens vier Arten des Singular-*they*

1. **generic indefinite**

   *Someone ran out of the classroom, but **they** forgot their backpack.*

2. **generic definite**

   *The ideal student completes the homework, but not if **they** have an emergency.*

3. **specific definite ungendered**

   *The math teacher is talented, but **they** hand back grades late.*

4. **specific definite gendered**

   *James is great at laundry, but **they** never wash their dishes.*

- außerdem: *he*, *she*, Plural-*they*

Methode

Naive Discriminative Learning

- relativ kleines Korpus, basierend auf COCA (Davies, 2008-)
  - 17.805 Tokens
  - 1.000 Sätze
  - 30 + Attestierungen pro Pronomen

    he, *she*, Plural-*they*, Generic-*they*
- da alle Pronomen spezifisch und generisch genutzt werden können, wurden alle Attestierungen manuell auf ihre Generizität überprüft und entsprechend annotiert
- Implementierung von NDL zur Berechnung semantischer Vektoren für Stämme und Flexionsmerkmale – nicht für Generizität

Instance-Vektoren

- als Resultat der NDL-Implementierung erhalten wir auch semantische Vektoren für die vier Pronomen
- **Problem**: die Pronomen haben, je nach Kontext, sehr wahrscheinlich sehr unterschiedliche Referenten → sehr unterschiedliche Semantik
- hierfür haben wir bereits eine **Lösung**, die den Kontext stark einbezieht, kennengelernt: **Instance-Vektoren**!
- daher berechnen wir für jede Pronomen-Attestierung einen eigenen Instance-Vektor

Linear Discriminative Learning

- nun können wir das mentale Lexikon mit LDL simulieren
  - Semantik: Instance-Vektoren für Pronomen, NDL-Vektoren für andere Wörter
  - Form: Trigramme
- aber: 30 oder mehr Vektoren – und damit 30 oder mehr Einträge – pro Pronomen entsprechen nicht dem, was wir vom mentalen Lexikon annehmen
- daher berechnen wir 30 LDL-Implementierungen
- pro Implementierung benutzen wir einen zufälligen Instance-Vektor für jedes Pronomen

Linear Discriminative Learning

- aus den approximierten Matrizen lassen sich semantische Maße berechnen, zum Beispiel
  - **Nachbarschaftsdichte**

    durchschnittliche Korrelation des Zielwortes mit seinen 8 nächsten Nachbarn

    höher = engere Nachbarschaft
  - **Aktivierungsvielfalt**

    Euklidische Norm des vorhergesagten Vektors

    höher = höheres Maß an Koaktivierung
- außerdem können die „verstandenen“ semantischen Vektoren anhand ihrer Kosinus-Ähnlichkeit verglichen werden

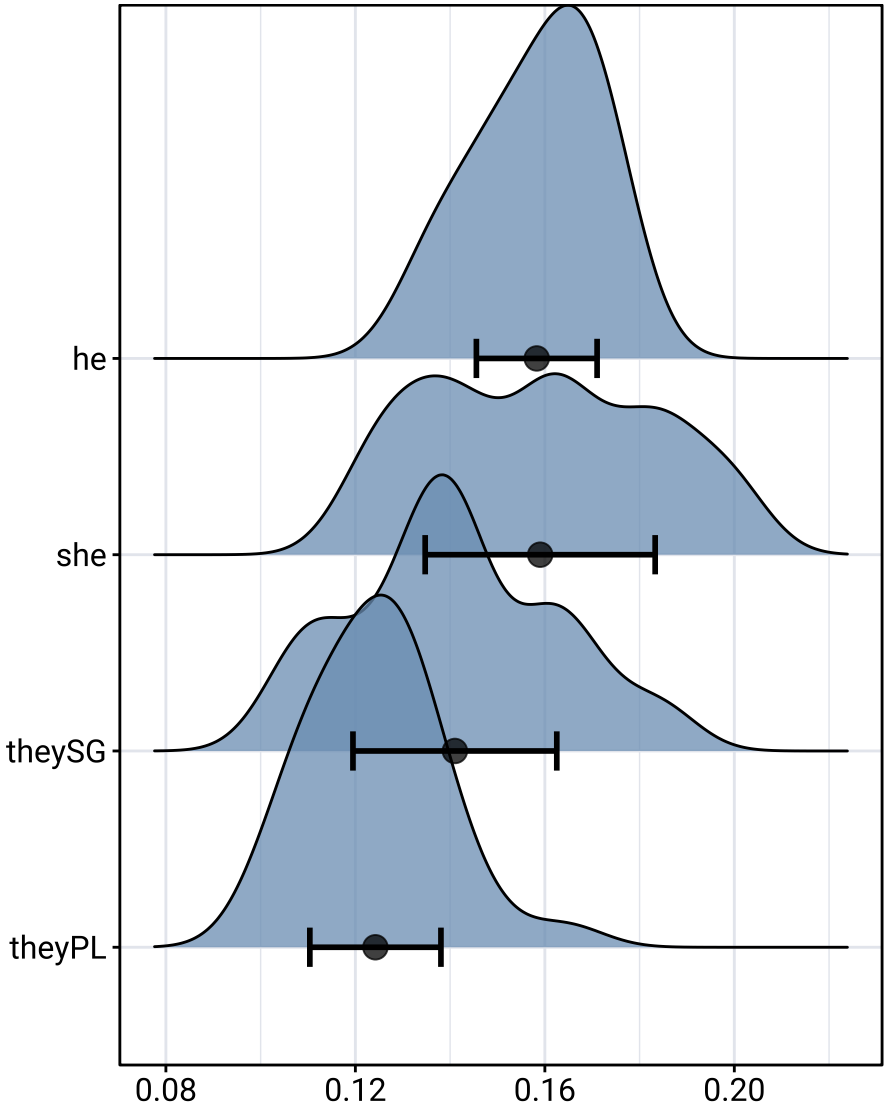
Analyse & Ergebnisse

Ergebnisse

Aktivierungsvielfalt

- höhere Werte = mehr Koaktivierung im mental Lexikon

he, she
↓
singular *they*
↓
plural *they*

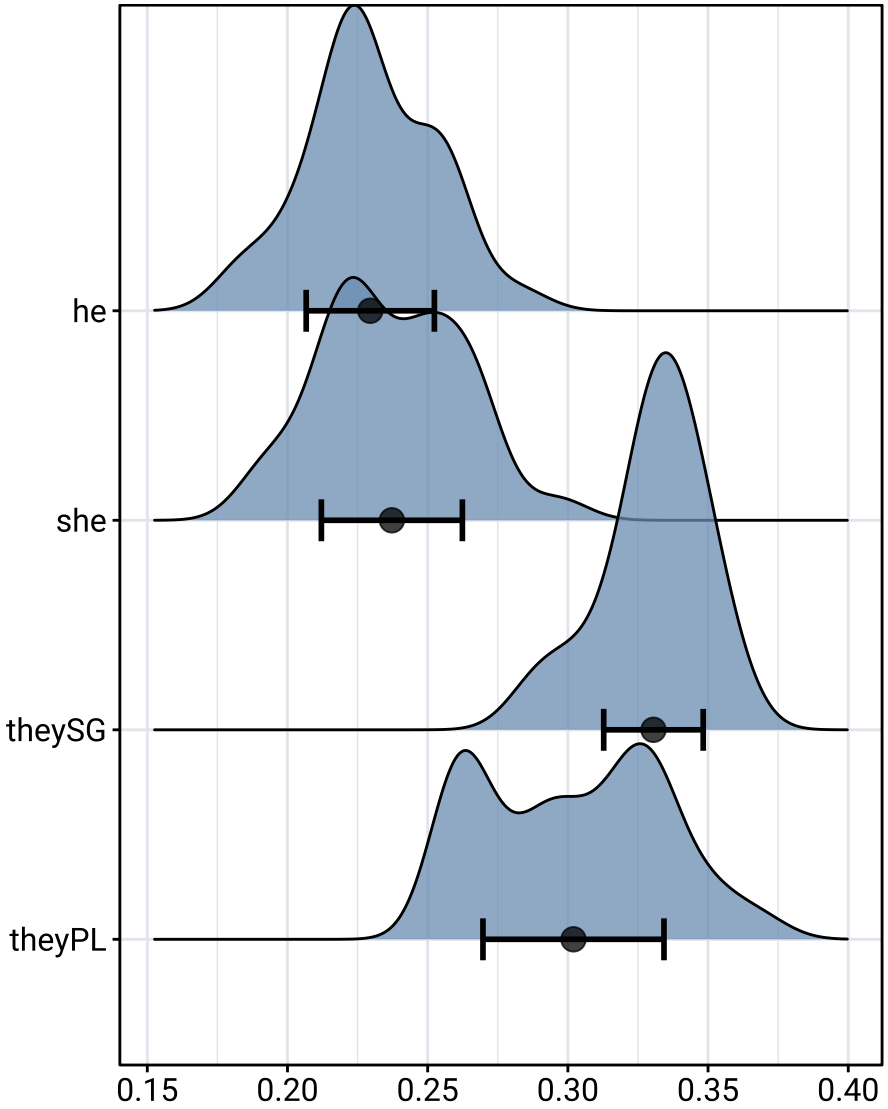


Ergebnisse

Nachbarschaftsdichte

- höhere Werte = mehr semantisch ähnliche Wörter im mentalen Lexikon

singular *they*
↓
plural *they*
↓
he, she



Ergebnisse

Kosinus-Ähnlichkeit

- Generic-*they* ist *he* und *she* ähnlicher als Plural-*they*

Resümee

- bisherige Studien
  - beschäftigen sich mit soziolinguistischen und syntaktischen Fragen (z. B. Conrod, 2020; Han & Moulton, 2022; Konnelly et al., 2020)
  - nicht mir der Semantik von Pronomen

- Lösungsvorschlag
  - semantische Vektoren durch eine NDL-Implementierung berechnen
  - Instance Vectors anhand der NDL-Vektoren berechnen
  - mit LDL das Sprachverständnis von Pronomen analysieren

Resümee

- Ergebnis
  - die Semantik von Generic-*they* ist anders als die von *he* und *she*, aber sie entspricht nicht der von Plural-*they*
  - Generic-*they* scheint Semantik jenseits von *he*, *she* und Plural-*they* zu besitzen → womöglich genau die genderneutrale Semantik, die es besitzen soll

- aus methodologischer & theoretischer Sicht spannend: scheinbar haben Pronomen eine eigene Semantik und diese kann computationell dargestellt und analysiert werden

Bereichern computerlinguistische Methoden die Genderlinguistik?

Zusammenfassung

- wir haben gesehen, dass computerlinguistische Methoden anwendbar sind zur Erforschung von…
  - semantischen Biases bzw. semantischer Ähnlichkeit bei Rollennomen
  - semantischer Ähnlichkeit bei Pronomen
  - Pronomensematik
  - Verarbeitung/Verständnis von generischen und nicht generischen Formen sowie von Pronomen im mentalen Lexikon

➡ computerlinguistische Methoden scheinen die Forschung in der Genderlinguistik also durchaus zu bereichern

Und jetzt?

- NDL, LDL, fastText, Instance-Vektoren und Co. bieten noch viele weitere Anwendungsmöglichkeiten
- so könnte man beispielsweise untersuchen…
  - wie genderneutrale Formen im Lexikon verarbeitet/verstanden werden
  - ob die Verarbeitung generischer Maskulina und genderneutraler Formen im Lexikon verschiedene andere Maße (z. B. Reaktionszeiten) erklären kann
  - wie die Semantik (anderer) Pronomen aussieht & wie diese zueinander stehen
  - wie (z. B. deutsche) Pronomen im Lexikon verarbeitet/verstanden werden
  - wie Semantik und Verarbeitung/Verständnis von Neo-Pronomen aussehen
  - …

Quellen, Tipps & Co.

- eine Auflistung zu nützlichen Quellen rund um die Methoden, die heute besprochen wurden, sowie diese Folien findet ihr unter

linktr.ee/dmncschmtz

Vielen Dank

für eure

Aufmerksamkeit!

Literatur 1/4

Arndt-Lappe, S., Schrecklinger, R., & Tomaschek, F. (2022). Stratification effects without morphological strata, syllable counting effects without counts – modelling English stress assignment with Naive Discriminative Learning. *Morphology*, 1–25. https://doi.org/10.1007/S11525-022-09399-9/TABLES/4

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. https://doi.org/10.1371/journal.pone.0174623

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

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. https://doi.org/10.48550/arxiv.1607.04606

Chuang, Y.-Y., Vollmer, M. L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., & Baayen, R. H. (2021). The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using linear discriminative learning. *Behavior Research Methods*, *53*(3), 945–976. https://doi.org/10.3758/s13428-020-01356-w

Conrod, K. (2020). Pronouns and gender in language. In *The Oxford Handbook of Language and Sexuality*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190212926.013.63

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/

Demarmels, S. (2017). „Gesucht: Assistentin oder Sekretär der Geschäftsleitung“ – Gendersensitive Formulierungen in Stellenanzeigen aus der Perspektive der Textsorte. In M. Nielsen, K. Luttermann, & M. Lévy-Tödter (Eds.), *Stellenanzeigen als Instrument des Employer Branding in Europa: Interdisziplinäre und kontrastive Perspektiven* (pp. 249–270). Springer. https://doi.org/10.1007/978-3-658-12719-0_11

Literatur 2/4

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. https://doi.org/10.3758/BRM.40.1.206

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., 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

Han, C. H., & Moulton, K. (2022). Processing bound-variable singular they. Canadian Journal of Linguistics/Revue Canadienne de Linguistique, 67(3), 267–301. https://doi.org/10.1017/CNJ.2022.30

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

Heitmeier, M., Chuang, Y. Y., & Baayen, R. H. (2021). Modeling morphology with linear discriminative learning: Considerations and design choices. Frontiers in Psychology, 12, 4929. https://doi.org/10.3389/FPSYG.2021.720713/BIBTEX

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

Literatur 3/4

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

Konnelly, L., Cowper, E., Konnelly, L., & Cowper, E. (2020). Gender diversity and morphosyntax: An account of singular they. *Glossa: A Journal of General Linguistics*, *5*(1). https://doi.org/10.5334/GJGL.1000

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. https://doi.org/10.3366/word.2018.0131

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. https://doi.org/10.48550/arxiv.1301.3781

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

Nieder, J., Chuang, Y. Y., van de Vijver, R., & Baayen, H. (2023). A discriminative lexicon approach to word comprehension, production, and processing: Maltese plurals. *Language*, *99*(2), 242–274. https://doi.org/10.1353/LAN.2023.A900087

Nieder, J., & Tomaschek, F. (2023). Maltese as a merger of two worlds: A cross-language approach to phonotactic classification. *PLOS ONE*, *18*(4), e0284534. https://doi.org/10.1371/JOURNAL.PONE.0284534

Nieder, J., van de Vijver, R., & Tomaschek, F. (2023). "All mimsy were the borogoves" – a discriminative learning model of morphological knowledge in pseudo-word inflection. *Language, Cognition and Neuroscience*, *38*(5), 693–710. https://doi.org/10.1080/23273798.2022.2127805

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.

Literatur 4/4

Schmitz, D. (2023). He, she, they, they: A first discriminative analysis of third-person pronoun semantics. Linguistic Intersections of Language and Gender, Heinrich-Heine-Universität Düsseldorf, Germany. 20-21 July. doi: 10.13140/RG.2.2.12243.96801

Schmitz, D. (2023). In German, all professors are male. In J. Pfeifer, S. Arndt-Lappe, H. Dorgeloh, G. Kunter, & C. Uffmann (Eds.), *INGO 6.0. The Proceedings. New empirical Insights about laNguage, presented on a Great day Out in September*. Preprint.

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*. https://doi.org/10.3389/fpsyg.2021.680889

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).

Shafaei-Bajestan, E., Moradipour-Tari, M., Uhrig, P., & Baayen, R. H. (2021). LDL-AURIS: a computational model, grounded in error-driven learning, for the comprehension of single spoken words. *Language, Cognition and Neuroscience*, 1–28. https://doi.org/10.1080/23273798.2021.1954207

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

Stein, S. D., & Plag, I. (2021). Morpho-phonetic effects in speech production: Modeling the acoustic duration of English derived words with linear discriminative learning. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.678712

Tomaschek, F., Domahs, U., Domahs, F., Tomaschek, F., Domahs, U., & Domahs, F. (2023). Modelling German Word Stress. *Glossa: A Journal of General Linguistics*, *8*(1). https://doi.org/10.16995/GLOSSA.9015

Tomaschek, F., Plag, I., Ernestus, M., & Baayen, R. H. (2019). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics*, *2019*, 1–39. https://doi.org/10.1017/S0022226719000203