A Bayesian approach to modelling German personal and demonstrative pronouns
Clare Patterson, University of Cologne; Johannes Hagen, University of Cologne; Andrew Kehler, University of California, San Diego; Petra B. Schumacher, University of Cologne
cpatters@uni-koeln.de

The Bayesian Model for pronouns (Kehler et al 2008; Kehler & Rohde 2013) predicts that pronoun production and comprehension are related by Bayesian principles (1).

(1) P(referent | pronoun) ∝ P(pronoun | referent)P(referent)

P(referent | pronoun) represents the comprehension bias: the probability that a particular referent is being referred to by a pronoun. The likelihood term *P(pronoun | referent)* represents the production bias: the hearer's estimate of the probability that speaker will elect to use a pronoun to refer to a particular referent. The prior term, *P(referent)*, represents the next-mention bias: the probability that a particular referent will get mentioned next, regardless of the type of referring expression used. Values for the prior and likelihood terms are typically estimated from passage completion experiments with free prompt conditions, yielding a predicted comprehension bias that can then be compared to the actual comprehension bias measured using pronoun-prompt conditions with the same contexts. Applying the Bayesian model to German pronouns presents an interesting challenge because both personal and demonstrative pronouns can refer to human entities. Additionally, the referential biases for the demonstrative *dieser* are not well understood nor have been tested within a Bayesian framework (see Bader & Portele 2019 for demonstrative *der*).

Two passage completion studies were conducted to compare the behavior of German personal and demonstrative pronouns against the predictions of the Bayesian model and two competing models (Rohde & Kehler 2014): a Mirror Model (the normalized probability of entity pronominalization) and an Expectancy Model (the normalized probability of entity next mention; Arnold 2001). In both experiments, items consisted of a context sentence followed by one of three prompt types: personal pronoun (*er*), demonstrative pronoun (*dieser*), and free prompt (a blank line). We further explored the contribution of the antecedents' syntactic function and thematic role, testing verbs where subject and agent/experiencer are aligned (2/4) or not (3/5). Experiment 1 (N=48) compared contexts with active-accusative verbs (2) and dative-experiencer verbs (3). Experiment 2 (N=40) compared contexts with experiencer–stimulus verbs (4) and stimulus–experiencer verbs (5).

Model performance is evaluated using R^2, MSE and average cross entropy (ACE) by comparing item-by-item predictions against item-by-item interpretations. For now, we compare the raw outcomes in Tables 1-4. The measured comprehension biases for personal pronouns in the pronoun prompt conditions better fit the predictions of the Bayesian model than the competing ones in three of the four context conditions (Tables 1 and 2). In the remaining condition (dative-experiencer), both the Bayesian and Expectancy Models outperform the Mirror model, but the Expectancy Model provides a slightly better fit.

For demonstratives (Tables 3 and 4), the predictions of the Bayesian and Mirror Models aligned closely with the measured comprehension biases in three conditions, with Mirror edging out Bayes in two. (The Expectancy Model is not included since demonstratives are not predicted to refer to highly expected referents; e.g. Bosch & Hinterwimmer 2016.) The results of the remaining condition – stimulus–experiencer – are puzzling for both models; whereas demonstratives were seldom interpreted as referring to the preceding subject in the other three conditions, an unusually high number of such references (28%) occurred here.

As an ensemble, the results for German personal pronouns support the predictions of the Bayesian Model, according to which comprehenders reverse engineer the speaker's referential intentions using Bayesian principles. The results for *dieser*, however, do not clearly differentiate the predictions of the Bayesian and Mirror models, and in one condition were problematic for both models. Further, it seems that the demonstrative is resolved towards the less agentive thematic role (patient or stimulus), at least in those cases where the canonical argument order places proto-agents before proto-patients (Dat and E–S verbs).

(2) Vorletzte Nacht hat der Hund den Papagei geärgert. Er/Dieser/__/_________
The night before last the dog (nom.masc.) annoyed the parrot (acc.masc.).
(3) Gestern ist dem Feuerwehrmann der Polizist aufgefallen. Er/Dieser/__/_________
Yesterday the firefighter (dat.masc.) noticed the police officer (nom.masc.).
(4) Der Dieb fürchtete den Polizisten. Er/Dieser/__/_________
The thief (nom.masc.) feared the police officer (acc.masc.).
(5) Der Fußballer erstaunte den Manager. Er/Dieser/__/_________
The footballer (nom.masc.) astonished the manager (acc.masc.).

Table 1. Experiment 1 biases and model outcomes, personal pronoun

	Acc /NP1	Acc/NP2	Dat/NP1	Dat/NP2
Production bias P(pronoun | referent) – estimated from free prompt condition	.87	.28	.93	.69
Prior term P(referent) – estimated from free prompt condition	.42	.58	.32	.68
Bayes Model prediction	.69	.31	.39	.61
Mirror Model prediction	.76	.24	.57	.43
Expectancy Model prediction	.42	.58	.32	.68
Empirical (actual) value – estimated from the pronoun prompt condition	.66	.34	.34	.66

Table 2. Experiment 2 biases and model outcomes, personal pronoun

	E–S/NP1	E–S/NP2	S–E/NP1	S–E/NP2
Production bias P(pronoun | referent)	1	.50	.95	.35
Prior term P(referent)	.39	.61	.72	.28
Bayes Model prediction	.56	.44	.87	.13
Mirror Model prediction	.67	.33	.73	.27
Expectancy Model prediction	.39	.61	.72	.28
Empirical (actual) value	.59	.41	.91	.09

Table 3. Experiment 1 biases and model outcomes, demonstrative pronoun

	Acc/NP1	Acc/NP2	Dat/NP1	Dat/NP2
Production bias P(pronoun | referent)	.03	.45	.03	.28
Prior term P(referent)	.42	.58	.32	.68
Bayes Model prediction	.05	.95	.05	.95
Mirror Model prediction	.06	.94	.10	.90
Empirical (actual) value	.08	.92	.11	.89

Table 4. Experiment 2 biases and model outcomes, demonstrative pronoun

	E–S/NP1	E–S/NP2	S–E/NP1	S–E/NP2
Production bias P(pronoun | referent)	0	.37	.02	.55
Prior term P(referent)	.39	.61	.72	.28
Bayes Model prediction	0	1	.09	.91
Mirror Model prediction	0	1	.04	.96
Empirical (actual) value	0	1	.28	.72

Arnold, *Discourse Processes*, 2001.
Bader & Portele, *Zeitschrift für Sprachwissenschaft*, 2019.
Bosch & Hinterwimmer, In Holler & Suckow (Eds), *Empirical Perspectives on Anaphora Resolution*, 2016.
Kehler et al., *Journal of Semantics,* 2008.
Kehler & Rohde, *Theoretical Linguistics*, 2013.
Rohde & Kehler, *Language, Cognition and Neuroscience*, 2014.

Model 1. The Bayesian Model

$$P(referent \mid pronoun) \propto P(pronoun \mid referent)P(referent)$$

- Comprehenders interpret a pronoun by reverse engineering the speaker's referential intentions in accordance with Bayesian principles
- Interpretation thus has top-down and bottom-up components
- Interpretation and production do not mirror each other

Model 2. The Mirror Model

$$P(referent \mid pronoun) \propto P(pronoun \mid referent)$$

- The interpreter arrives at an interpretation based on reasoning about the speaker's referential intentions — what referent is the speaker most likely to pronominalize a mention of?
- That means the interpreter's biases will be proportional to (their estimates of) the speaker's production biases

Model 3. The Expectancy Model

$$P(referent \mid pronoun) \propto P(referent)$$

- According to Arnold's Expectancy Hypothesis (Arnold 2001), comprehenders will interpret a pronoun to refer to whatever referent they expect to be mentioned next
- Note that is model is more 'top down': the prior is assumed to be precomputed as component of predictive processing