

# RAILS

## Rational Approaches in Language Science

24th-26th October 2019





# Contents

Organization . . . . .	1
Welcome . . . . .	2
Schedule . . . . .	3
Invited talks . . . . .	7
Jaeger . . . . .	7
Kuperberg . . . . .	8
Rohde . . . . .	9
Turnbull . . . . .	10
Talks . . . . .	11
Anderson and Dillon . . . . .	11
Aurnhammer and Frank . . . . .	12
Degaetano-Ortlieb, Bizzoni, Fankhauser and Teich . . . . .	13
Dotlačil and Haan . . . . .	14
Häuser and Kray . . . . .	15
Heller and Stevenson . . . . .	16
Hoek, Kehler and Rohde . . . . .	17
Hörberg . . . . .	18
Jágrová, Mosbach, Hedderich, Avgustinova and Klakow . . . . .	19
Lemke, Schäfer, Drenhaus and Reich . . . . .	20
Levshina . . . . .	21
Muljadi, Ankener, Sikos and Staudte . . . . .	22
Ness and Meltzer-Asscher . . . . .	23
Radulescu, Giannoupoulou, Avrutin and Wijnen . . . . .	24
Ryzhova and Demberg . . . . .	25
Schäfer, Lemke, Drenhaus and Reich . . . . .	26
Seminck and Amsili . . . . .	27
Shirtz and Verkerk . . . . .	28
Sikos, Venhuizen, Drenhaus and Crocker . . . . .	29
Thothathiri . . . . .	30
Posters . . . . .	31
Bhandari, Demberg and Kray . . . . .	31
Bizzoni, Teich, Juzek, Menzel and Przybyl . . . . .	32
Brandt, Möbius and Andreeva . . . . .	33
Hong and Yung . . . . .	34
Naumann and Petersen . . . . .	35
Saget . . . . .	36
Schlotterbeck . . . . .	37
Tellings . . . . .	38
Voigtmann . . . . .	39
Zarrieß . . . . .	40
Getting around . . . . .	41
Conference dinner . . . . .	42
Internet and WiFi . . . . .	42

## **Organizing Committee**

Matthew Crocker (chair)  
Bistra Andreeva  
Patricia Borrull Enguix  
Stefania Degaetano-Ortlieb  
Vera Demberg  
Robin Lemke  
Noortje Venhuizen

## **Program Committee**

Bistra Andreeva  
Harm Brouwer  
Matthew Crocker  
Stefania Degaetano-Ortlieb  
Vera Demberg  
Heiner Drenhaus  
Gerhard Jäger  
Dietrich Klakow  
Gina Kuperberg  
Robin Lemke  
Bernd Möbius  
Ingo Reich  
Hannah Rohde  
Augustin Speyer  
Rory Turnbull  
Noortje Venhuizen

## **Steering Committee**

Elke Teich  
Bernd Möbius  
Ingo Reich  
Maria Staudte  
Jutta Kray  
Marie-Ann Kühne

## Welcome

The motivation for RAILS – despite a recent proliferation of conferences and workshops in the various sub-disciplines of language science – was to bring together researchers from a wide range of disciplines, who have something else in common. This “something else” is the idea that language and its use can be better understood by considering rational explanations. Typically such explanations identify the goal of a linguistic system and the environment in which it operates in order to reason about how this goal might be accomplished optimally, or rationally. In order to operationalize the notion of rationality, probabilistic approaches – such as Bayesian, Information Theoretic, and Game Theoretic frameworks – are often adopted. The goal of this conference is to bring together speech and language researchers that have benefited from, and share, such rational explanations, and whose scientific contributions reflect the full diversity of disciplines and methodologies – from speech to discourse, on-line processing to corpus-based investigation, through to language change and evolution.

Scientific and financial support for this conference comes from **SFB1102** “Information Density and Linguistic Encoding”, a Collaborative Research Center funded by the German Science Foundation (DFG), which applies information theoretic and Bayesian approaches to a variety of linguistic encoding phenomena.

We are grateful for your contribution to, and participation in, our event. We look forward to fruitful exchanges and insights from all across the language sciences. And we wish you a very warm welcome to Saarbrücken!

## Schedule

### Thursday October 24

10:00 – Registration + Coffee

11:30

11:00 – Opening

11:05

11:05 – **Gerhard Jäger (invited talk)**

12:05 Bayesian typology

12:05 – **Robin Lemke, Lisa Schäfer, Heiner Drenhaus and Ingo Reich**

12:30 Predictable words are more likely to be omitted in fragments – Evidence from production data

12:30 – Lunch

13:30

13:30 – **Thomas Hörberg**

13:55 The processing of grammatical functions in Swedish is expectation-based

13:55 – **Klára Jágrová, Marius Mosbach, Michael A. Hedderich, Tania**

14:20 **Avgustinova and Dietrich Klakow**

On the correlation of context-aware language models with the intelligibility of Polish target Words to Czech readers

14:20 – **Shahar Shirtz and Annemarie Verkerk**

14:45 The Negative Existential Cycle in Indo-European: Is a rational approach enlightening?

14:45 – Coffee

15:15

15:15 – **Margarita Ryzhova and Vera Demberg**

15:40 Are pragmatic inferences triggered by informationally redundant utterances effortless?

15:40 – Poster Slam

16:00

16:00 – Poster Session

17:00

17:00 – Drinks

18:00

**Friday October 25**

09:00 –	<b>Hannah Rohde (invited talk)</b>
10:00	Why are you telling me this: Comprehension as a process of reverse engineering
10:00 –	<b>Jet Hoek, Andrew Kehler and Hannah Rohde</b>
10:25	Modeling coreference in contexts with three referents
10:25 –	Coffee
11:00	
11:00 –	<b>Daphna Heller and Suzanne Stevenson</b>
11:25	Speakers are not maximally rational when choosing referring expressions
11:25 –	<b>Olga Seminck and Pascal Amsili</b>
11:50	A cost metric for pronoun resolution: uncertainty increases processing cost
11:50 –	<b>Les Sikos, Noortje Venhuizen, Heiner Drenhaus and Matthew Crocker</b>
12:15	Reevaluating pragmatic reasoning in web-based language games
12:15 –	Lunch
13:30	
13:30 –	<b>Lisa Schäfer, Robin Lemke, Heiner Drenhaus and Ingo Reich</b>
13:55	Speakers use verb phrase ellipsis to satisfy UID: Psycholinguistic evidence from length and context effects
13:55 –	<b>Tal Ness and Aya Meltzer-Asscher</b>
14:20	Rational adaptation in lexical prediction: The influence of prediction strength
14:20 –	<b>Christine Muljadi, Christine Ankener, Les Sikos and Maria Staudte</b>
14:45	Verb Surprisal in the Visual World
14:45 –	Coffee
15:15	
15:15 –	<b>Katja Häuser and Jutta Kray</b>
15:40	“As you like it”: Individual differences in predictive processing relate to successful episodic memory retrieval
15:40 –	<b>Gina Kuperberg (invited talk)</b>
16:40	How rational is the brain? – A probabilistic generative framework of language comprehension
18:30 –	Dinner
late	

### Saturday October 26

- 09:00 – **Natalia Levshina**  
09:25 Information-theoretic approaches to communicative efficiency: some controversial issues and open questions
- 09:25 – **Stefania Degaetano-Ortlieb, Yuri Bizzoni, Peter Fankhauser and Elke Teich**  
09:50 A communicative perspective on conventionalization in diachronic language change
- 09:50 – **Carolyn Anderson and Brian Dillon**  
10:15 Taking other perspectives into account: an RSA model of perspectival reasoning
- 10:15 – Coffee  
10:45
- 10:45 – **Silvia Radulescu, Efi Giannoupoulou, Sergey Avrutin and Frank Wijnen**  
11:10 Item bound vs category-based generalization. An entropy model
- 11:10 – **Christoph Aurnhammer and Stefan Frank**  
11:35 Evaluating information-theoretic measures of word prediction in naturalistic sentence reading
- 11:35 – **Jakub Dotlačil and Puck de Haan**  
12:00 Data-driven parsing in an adaptive cognitive architecture
- 12:00 – Lunch  
13:00
- 13:00 – **Malathi Thothathiri**  
13:25 The effect of cue validity during input on sentence choice during output
- 13:25 – **Rory Turnbull (invited talk)**  
14:25 Phonetic reduction, natural selection and bounded rationality
- 14:25 – Closing  
14:30



# Abstracts

### **Bayesian Typology**

*Gerhard Jäger (University of Tübingen)*

gerhard.jaeger@uni-tuebingen.de

In a landmark paper, Maslova (2000) argued that the synchronous frequencies of a typological variable do not reveal distributional universals. As there is no guarantee that the underlying dynamic process has reached equilibrium, observed frequencies may reflect properties of ancestor languages rather than functional tendencies. As a remedy, Maslova proposes to estimate the transition rates between types from diachronic data and to compute the equilibrium distribution analytically instead.

Probably due to the sparsity of diachronic typological evidence, this program has not been realized so far. Techniques from the \*phylogenetic comparative method\* (cf. Nunn, 2011) in computational biology, however, paired with the newly available electronic typological data sources and Bayesian inference, afford an alternative way to realized Maslova's goal.

Once a typological variable and a collection of languages has been fixed, the workflow is as follows:

1. Infer a (distribution of) phylogeny(ies) from lexical data.
2. Estimate the transition matrix between the values of the variable.
3. Calculate the equilibrium distribution of this Markov process.

In the talk, I will present the method as well as several case studies pertaining to word order and case marking typology.

**How rational is the brain?**  
**A probabilistic generative framework of language comprehension**  
**Gina Kuperberg MD PhD**

A large body of evidence suggests that comprehenders are able to use the preceding context, in combination with their stored linguistic and non-linguistic knowledge, to probabilistically predict upcoming inputs (Kuperberg & Jaeger, 2016). So long as our probabilistic knowledge mirrors the statistics of the linguistic input, this anticipatory processing provides a rational way of ensuring that comprehension is both fast and accurate. But what exactly does rational mean, given that our brains have limited resources, and that we sometimes encounter inputs that are completely unexpected?

One prominent theory of rational comprehension — surprisal theory — argues that we use the prior context to pre-activate even very low probability continuations (Hale, 2001; Levy, 2008; Smith & Levy, 2013). According to this theory, a word's processing is predicted by its negative log probability, given the context. It therefore assumes that we devote more resources to pre-activating information that is less likely to appear in the bottom-up input. It also assumes that the difficulty of lexical access and the difficulty of building a message-level representation can be collapsed into a single processing stage. I will present evidence that challenges these two assumptions.

First, I will present data from a series of controlled behavioral and event-related potential (ERP) experimental studies, which show that the relationship between the probability of an incoming word and its processing is, in fact, linear rather than logarithmic (Brothers & Kuperberg, 2019). This contradicts the corpus-based findings reported by Smith & Levy (2013). This linear relationship is also seen in a meta-analysis of existing eye tracking studies. These findings suggest that we probabilistically predict upcoming words in proportion to their likelihood of actually appearing. Second, I will present neural data (ERP, MEG and fMRI) showing that, in some situations, low probability words in discourse contexts produce additional later responses that are not seen to higher probability words (e.g. Kuperberg, Brothers, & Wlotko, in press). These data suggest that lexical access and message-level integration are not always mediated by the same neural mechanism.

I will argue that these findings can be understood within a *hierarchical generative framework* of language comprehension (Kuperberg & Jaeger, 2016; Kuperberg, Brothers, & Wlotko, in press). Within this framework, the comprehender draws upon a hierarchical generative model that she believes mirrors the statistical structure of her current communicative environment. Probabilistic predictions are passed down from higher to lower levels of the hierarchy in proportion to the certainty of her high-level beliefs. Bottom-up information that matches these predictions is 'explained away', leading to facilitated processing. Unpredicted/unexplained information that cannot be explained at lower levels drives belief updating at higher levels of the hierarchy, or, if it conflicts with the structure of the current generative model, triggers reanalysis and adaptation. This architecture can be understood within a bounded rational framework that assumes that language evolved to support communication across brains with limited metabolic resources, in communicative environments that vary systematically across situations.

Brothers, T., & Kuperberg, G. R. (2019). Linear (not logarithmic) effects of lexical predictability: A challenge to surprisal theory. Presented at the 32nd Annual CUNY Conference on Human Sentence Processing, Boulder, CO.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. Presented at the Proceedings of the North American Chapter of the Association for Computational Linguistics on Language technologies (NAACL '01), Pittsburgh, PA.

Kuperberg, G. R., Brothers, T., & Wlotko, E. (in press). A Tale of Two Positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32-59. doi:10.1080/23273798.2015.1102299

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177. doi:10.1016/j.cognition.2007.05.006

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302-319. doi:10.1016/j.cognition.2013.02.013

## Why are you telling me this: Comprehension as a process of reverse engineering

*Hannah Rohde (University of Edinburgh)*

*hannah.rohde@ed.ac.uk*

Communication depends in part on the hope that people will tell us things we're interested in hearing. In this sense, language can be understood as a channel by which speakers can convey, among other things, newsworthy and informative messages (i.e., content that is otherwise unpredictable to the comprehender). Despite this understanding, research on language comprehension often emphasizes a preference for the opposite (i.e., for content that is real-world predictable). Indeed, several decades of research has shown that comprehenders can deploy knowledge about situation plausibility to generate fine-grained context-driven predictions about upcoming words. This talk asks whether such knowledge is always deployed directly in favor of real-world plausible content, or whether an additional measure of situation surprisal is also relevant, one that introduces a bias in favor of real-world unpredictable content because its mention would be informative. Reverse engineering a speaker's meaning thus requires a comprehender to not only guess what situations a speaker would probably encounter but to combine that guess with one about what situations the speaker would be likely to mention.

The first study considers this distinction from the perspective of color modification. Prior work has found that speakers produce color modifiers at different rates for different objects. For example, although bananas are typically yellow in the real world, speakers don't typically mention their yellowness in their descriptions; in contrast, objects with more variable color are more often described with color modifiers. In a study testing comprehenders' anticipation of upcoming words, we show that comprehenders make use of this variable modification across objects: In a scene with a yellow banana and a yellow t-shirt, they anticipate that an instruction that starts "click on the yellow..." is more likely to be followed by the t-shirt rather than the (prototypically yellow) banana. This suggests that comprehenders are not simply consulting their real-world knowledge about bananas and t-shirts, but rather they are using their knowledge about what speakers tend to \*say\* about bananas and t-shirts.

More generally, one can ask whether comprehenders distinguish between the probability that a proposition is true and the likelihood that a speaker would choose to formulate an utterance to convey that proposition. The second study tests whether comprehenders' expectations about upcoming content depend on whether the content is portrayed as someone's belief or someone's spoken utterance. Indeed, estimates about the content of beliefs are found to correspond more closely with estimates of real-world situations, whereas estimates about the content of spoken utterances are found to diverge from real-world estimates.

The final set of studies use self-paced reading to test whether sentences about newsworthy unpredictable situations can indeed be easy to process. We manipulate semantic context (unusual protagonists), syntactic construction (wh- clefts), and the communicative environment (text messages). Together, the results highlight the need for models that distinguish comprehenders' knowledge about situation plausibility from their knowledge about what people actually choose to talk about.

**Phonetic reduction, natural selection and bounded rationality***Rory Turnbull (University of Hawai'i at Mānoa)*

rory.turnbull@hawaii.edu

Phonetic reduction is pervasive in natural speech. Previous research has found robust relationships between phonetic reduction and linguistic predictability. In this talk I examine this relationship through the lens of natural selection and bounded rationality. Particularly, I ask whether such processes are the result of a rational optimizing process, and if so, whether there are limits on this rationality. Secondly, I explore the extent to which phonetic reduction can be reduced to consequences of natural selection over speech exemplars. Taken together, these insights provide a new perspective on the relationship between speech and predictability.

## Taking other perspectives into account: an RSA model of perspectival reasoning

Carolyn Jane Anderson and Brian Dillon, *University of Massachusetts, Amherst*  
carolynander@umass.edu

We propose a Rational Speech Acts (RSA) model of perspectival expressions and provide experimental support for its key insight: listeners consider multiple perspectives simultaneously. Perspectival expressions like English "come", which describes motion relative to a perspective-holder, pose an interpretative problem because there are multiple possible perspective-holders: the speaker, listener, or attitude-holder (Fillmore 1997). Consequently, (1) can mean that Thelma is traveling to either Seattle or London.

1. Context: Sam, in Seattle, says to Lucy, in London: "Thelma is coming."

Because listeners must reason both about the speaker's adopted perspective and their message, the interpretation of perspectival items can be modeled as a joint reasoning process.

We model perspectival reasoning in the RSA framework, in which listeners use Bayesian inference to calculate probabilities for worlds representing possible meanings (Bergen et al., 2012; Frank and Goodman, 2012). In our perspectival version, the listener jointly infers the probability of a world and perspective according to their model of how the speaker selects an utterance-perspective pair (the Literal Speaker).

Literal Listener:  $p(w|m,a) \propto \text{denotation}(m,a,w) p(w)$

Literal Speaker:  $p(m,a|w) \propto \text{softmax} (p(w|m,a) \sum_w \text{denotation}(m,a,w) p(a) - \text{Cost}(a))$

Pragmatic Listener:  $p(w,a|m) \propto p(m,a|w) p(w)$

where  $a$  = perspective,  $w$  = world, and  $m$  = utterance

A critical component of the model is that listeners consult multiple perspectives simultaneously, in contrast with the speaker-default proposed in prior work (Harris 2012, Barlew 2017). The multiple perspectives approach predicts that given (1), the marginal posterior probability should be highest for worlds where both the speaker and listener are at the destination.

We tested this prediction in a comprehension task. Participants read a sentence with a perspectival verb or manner-of-motion verb and then saw a scene depicting both the speaker and listener at the destination; just the speaker; just the listener; or neither. Participants indicated whether the scene and sentence matched.

For scenes showing both perspective-holders at the destination, there was no difference in reaction times in the come condition relative to the control condition (walk/come difference = -19 (+/-119) 95%CI), but in all others, RTs were slower for come (speaker: 138 (+/-126); listener: 408 (+/-152); none: 138 (+/-137)). A mixed effects regression revealed a significant interaction between the both-scene and come condition. Participants are faster to recognize the both-scene for perspectival expressions, in support of the multiple-perspectives account.

Barlew, J. (2017). The semantics and pragmatics of perspectival expressions in English and Bulu: The case of deictic motion verbs. Dissertation, Ohio State University.

Bergen, L., Goodman, N., & Levy, R. (2012). "That's what she (could have) said: how alternative utterances affect language use." *Proceedings of the Annual Conference of the Cognitive Science Society* 34.

Fillmore, C. (1997). *Lectures on Deixis*.

Frank, M. & Goodman, N. (2012). "Predicting pragmatic reasoning in language games." *Science* 336:998-998.

Epley, N., Keysar, B., Van Boven, L. & Gilovich T. (2004). "Perspective taking as egocentric anchoring and adjustment." *Journal of Personality and Social Psychology* 87(3).

Harris, J. (2012). *Processing Perspectives*. Dissertation, University of Massachusetts Amherst.

## Evaluating information-theoretic measures of word prediction in naturalistic sentence reading

Christoph Aurnhammer, Department of Language Science and Technology, Saarland University, Saarbrücken, Germany;

Stefan Frank, Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

aurhammer@coli.uni-saarland.de

In this study, we review information-theoretic measures of cognitive load during sentence processing that have been used to quantify word prediction effort. Two such measures, *surprisal* (Hale, 2001) and *next-word entropy*, suffer from shortcomings when employed for a predictive processing view. We propose a novel metric, *lookahead information gain* (LIG), that can overcome these short-comings. Unlike surprisal, LIG is strictly forward looking, and effects of LIG are indicative of prediction. Unlike next-word entropy, LIG is theoretically derived from assumptions about the language comprehension process.

We revisit the formal derivation of surprisal in Levy (2008), which demonstrated that surprisal expresses the effort involved in shifting probability mass from the language processor's predictions over what word comes next to a distribution in which the (then known) next word has a probability of 1. We then apply this derivation to next-word entropy and find that it forces the inadequate assumption that during prediction all words have a uniform prior probability. Our new metric, LIG, improves upon this by assuming either a word frequency prior or a prior that is itself conditional on the words observed so far.

Using probabilistic language models, we compute all three measures (surprisal, next-word entropy, LIG). Subsequently, we put them to the test by analysing how well the estimated measures predict human processing effort in three data sets of naturalistic sentence reading. The data had been collected using self-paced reading, eye-tracking, and electroencephalography.

Our results replicate the well-known effect of surprisal on word reading effort, but do not indicate a role of next-word entropy or LIG. Unexpectedly, the LIG values increased with improved language model training, suggesting that, in a predictive processing system, the cost of predicting may outweigh the gains. This idea poses a potential limit to the value of a predictive mechanism for the processing of language.

The results illustrate three unresolved problems of finding estimations of word-by-word prediction: First, they need to be truly independent of perceptual processing of the to-be-predicted words, second, they need to be statistically reliable predictors of experimental data, and third, they must be derived from more general assumptions about the cognitive processes involved.

## References

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.

Hale, J. (2001, June). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 1-8). Association for Computational Linguistics.



**A communicative perspective on conventionalization in diachronic language change**

*Stefania Degaetano-Ortlieb, Yuri Bizzoni, Peter Fankhauser and Elke Teich*

*University of Saarland*

*stefania-deg@gmx.de*

Language use varies according to a number of factors, from pragmatic and cognitive to social. For on-line production it has been shown that specific forms of variation directly serve rational communicative goals by offering ways to modulate information density (Jaeger/Levy 2007); and for comprehension, there is ample evidence that particular linguistic choices are associated with specific levels of surprisal (Levy 2008; Schulz et al. 2016; Delogu et al. 2017; Sikos et al. 2017). What is much less clear is what the communicative effects of interaction are, if any, i.e. of particular linguistic choices recurring across interactants, interaction instances and contexts. Spontaneously occurring linguistic co-adaptation or convergence among interactants in on-line situations is a widely studied phenomenon (see e.g. Coles-Harris (2017) for the phonetic level) but communicative effects have so far only rarely been discussed (see e.g. Hume/Mailhot (2013) for phonological effects). Similarly, *conventionalization*, i.e. the longer-term linguistic effects of repeated interaction, has hitherto hardly been considered from a communicative perspective. Widely acknowledged as a relevant process in language change (Schmid 2015), conventionalization provides a prerequisite for innovation (de Smet 2016) and leads to persistent change in the language system overall (Bybee/Hopper 2001) and, when pertaining to particular socio-cultural contexts, to the formation of varieties (registers, dialects (Ure 1982; Trudgill 2008).

We suggest here that conventionalization is a cornerstone in communication since it comes with significant surprisal and entropy-reducing effects (Harris 1991, 2002). To show this, we pursue an exploratory, corpus-based approach, focusing on scientific writing (Degaetano-Ortlieb/Teich 2018, Degaetano-Ortlieb/Teich forthcoming), a well-studied and fairly controlled domain, and its evolution across 300 years from the mid-17<sup>th</sup> century onwards. The data we use are the Proceedings/Transactions of the Royal Society of London. To capture lexical and syntactic aspects of linguistic change leading to conventionalization, we employ probabilistic language models (word and part-of-speech based n-gram models, topic models, word embeddings); and to evaluate the observed effects, we apply various measures of information content (surprisal, entropy, relative entropy). We find for instance that diachronically, within the scientific domain, relative entropy on n-gram models overall decreases, and topic-document entropy and entropy over word embedding clusters also go down, thus indicating conventionalization at both lexical and syntactic levels. For qualitative interpretation, we inspect the linguistic features that significantly contribute to these trends, resulting in stable average surprisal and low entropy on those linguistic choices that become characteristic of scientific language over time.



## Data-driven parsing in an adaptive cognitive architecture

*Jakub Dotlačil, Puck de Haan, University of Amsterdam*

*j.dotlacil@gmail.com*

In psycholinguistics two strategies are popular in the development of parsing models that predict reading times (RTs): a data-driven parser is transported from computational linguistics or constructed using similar techniques (Demberg et al., 2013) and a linking hypothesis is formulated to connect the model to RTs (Boston et al., 2011, Hale, 2014); alternatively, a parser is developed in a cognitively informed framework by hand-coding rules and no (independent) linking hypothesis is needed (Lewis and Vasishth, 2005).

I present a combination of the two strategies: developing a data-driven parser in a cognitively informed framework. Using the cognitive architecture Adaptive Control of Thought-Rational (ACT-R, Anderson, 2007) I construct a deterministic phrase-structure classifier-based parser (cf. Sagae and Lavie, 2004). In this parser, only single parse into constituents is present at a time and whenever multiple choices are available, the parser lets the classifier select a unique action. Using ACT-R, the classification task becomes a case of retrieval of a chunk from declarative memory, such that the retrieved chunk is in turn a parsing action done in the past that partially matches the current local context and has the highest activation. This approach has several consequences: (i) given that declarative memory and retrieval from declarative memory in ACT-R is shaped by environment and adapted to it (Anderson, 1990), this approach constitutes a rational approach to parsing; (ii) since ACT-R links memory structures and activations to latencies and recall probabilities, the approach immediately makes predictions for reading times, (iii) since cognitive architectures are strong in bringing together various pieces of cognition (memory, reasoning, visual encoding...), it is possible to let the parser affect and react to other components of human mind.

The plausibility of this approach is tested on one implementation, a bottom-up parser that was trained on Penn Treebank (PTB) data (S1-S21). The precision and recall was evaluated on S23 of PTB (F1: 74.8). More importantly, its processing predictions were evaluated on three sets: data from Grodner & Gibson (2005), eye-tracking corpus from Frank et al (2013) and from Kennedy et al. (2005). The parser improves the fit of the model beyond lexical (frequency)/visual (length) information and, after Bayesian estimation of free ACT-R parameters, makes largely correct predictions for early RTs, e.g., first-pass RTs (details in the presentation).

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale: Lawrence Erlbaum Associates.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Boston, M. F., Hale, J. T., Vasishth, S., and Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, 26(3), 301–349.

Demberg, V., Keller, F., & Koller, A. (2013). Incremental, predictive parsing with psycholinguistically motivated tree-adjointing grammar. *Computational Linguistics*, 39(4), 1025-1066.

Frank, S. L., Monsalve, I. F., Thompson, R. L., and Vigliocco, G. (2013). Reading time data for evaluating broad-coverage models of English sentence processing. *Behavior Research Methods*, 45(4):1182–1190.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive science*, 29(2), 261-290.

Hale, John T (2014). *Automaton theories of human sentence comprehension*. CSLI Publications.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, 45.

Lewis, R. L., and S. Vasishth (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science* 29(3), pp. 375-419.

Sagae, K., & Lavie, A. (2005). A classifier-based parser with linear run-time complexity. In *Proceedings of the Ninth International Workshop on Parsing Technology*, 125-132.

**‘As you like it’: Early context use during encoding  
relates to successful retrieval from episodic memory**

Katja Häuser & Jutta Kray  
Universität des Saarlandes  
katja.haeuser@mail.mcgill.ca

According to the traditional view of psycholinguistics, language processing is primarily integrative in nature (*integration* view), but recent years have seen a heated debate about whether language users actively predict upcoming words and their features (*prediction* view), including morpho-syntactic aspects such as gender. Here, we used the gender marking of German nouns to investigate how individual differences in the use of predictive processing during language encoding relate to subsequent recall of words from memory.

We used a self-paced reading task and a surprise (i.e., previously unannounced) word recognition test. During self-paced reading, participants read sentences such as *Nachdem Paul seinen Führerschein erhalten hatte, fuhr er ständig mit DEM AUTO / DER GRUPPE von Freunden*, where the gender of the definite article preceding the noun (*dem*, *der*) could be used as an early contextual cue to indicate whether the most highly expected noun (*Auto*) would appear later on or not. To account for spill-over from the definite article, we inserted modifiers and adverbs (e.g., .... *mit dem / der alten, aber zuverlässigen Auto / Gruppe*), which were not gender-marked. In the subsequent word recognition task, participants were presented with target nouns from the reading task and new nouns, and had to indicate whether they remembered reading these words in the prior part of the study.

Two key findings emerged. First, during encoding, participants showed elevated reading times on the spill-over region after the unexpected article. Thus, readers seemed to actively anticipate upcoming nouns (including their gender), and then showed early effects of having their predictions disconfirmed after reading the unexpected article. Second however, this pattern of early prediction, though significant for the whole sample, was only present in high recallers, in other words, in participants who showed higher recall performance in the word recognition task. Of note, these same people also showed earlier effects of integration during reading, in the form of prolonged reading times at the level of the unexpected noun. Low recallers, in contrast, showed no signs of early prediction or of early integration, but instead experienced very late-stage integration difficulties, in the form of a small and delayed spill-over effect on the first two words after the noun.

In sum, our findings indicate that early use of context during encoding relates to enhanced subsequent memory retrieval: Whereas high recallers showed earlier and stronger signs of having their expectations disconfirmed (in the form of an early prediction effect on the article and an early integration effect on the noun), low recallers only showed evidence of very late-stage integration difficulty (only after encountering the noun). These results corroborate recent research suggesting that not all language users readily use prediction during language comprehension, and that prediction might be contingent on faster and more efficient use of context in some individuals.

## Speakers are not maximally rational when choosing referring expressions

Daphna Heller and Suzanne Stevenson, University of Toronto

daphna.heller@utoronto.ca

The goal of using a referring expression is for the addressee to pick out the entity intended by the speaker. Thus, it seems reasonable that speakers should choose referring expressions based on the consideration of referential success. Indeed, speakers have generally been shown to consider what the addressee knows when tailoring referential forms.

Here we discuss two cases that show the limits of this consideration; both cases involves situations where the speaker sees more objects than their addressee. We model the experimental results using the multiple-perspectives model (Mozuraitis et al., 2018).

First, Wardlow Lane and Ferreira (2008, Exp. 1) examined reference in two situations: only mutually-visible objects could be referred to (“3”), or all four objects were potential referent (“4”), including an object not visible to the addressee. Modelling results reveal that, in the first round, speakers in “3” generally ignored the hidden object in choosing how to label objects, whereas speakers in “4” did not. In the second round, participants switched roles, and also condition. Here the difference between the situations disappeared, and speakers in both situations ignored the hidden object. Why would speakers in “4” ignore the hidden object, a behavior that can lead the addressee not to be able to choose it as the referent? We propose that those speakers are guided by their earlier experience as addressee in “3”. This is despite the fact that this experience is no longer relevant, because the set of referents has now changed.

Second, Vanlangendonck et al. (2016) examined two different situations: where the knowledge mismatch is such that information from the speaker’s perspective would threaten referential success, and where doing so would only lead to the inclusion of unnecessary information. Experimental results showed that in both cases speakers adapted to the addressee, but not fully. Interestingly, even with the more fine-grained modelling, there is no evidence that speakers adapt to their addressee more when not doing so can lead to referential failure.

These modelling results point to the (surprising) conclusion that speakers do not choose referring expressions based on a case-by-case consideration of whether the referring expressions used would allow the addressee to identify the correct object. We thus demonstrate that speakers are not maximally rational in tailoring referential forms. Instead, we propose that speakers are guided by more general consideration of situational cues.

### References

- Mozuraitis, M., Stevenson, S. & Heller D. (2018). Modelling reference production as the probabilistic combination of multiple perspectives. *Cognitive Science*, 42, 974-1008.
- Vanlangendonck, F., Willems, R. M., Menenti, L. & Hagoort P. (2016). An early influence of common ground during speech planning. *Language, Cognition and Neuroscience*, 31(6), 741-750.
- Wardlow Lane, L., Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1466-1481.

### Modeling Coreference in Contexts with Three Referents

Jet Hoek (*The University of Edinburgh*), Andrew Kehler (*University of California, San Diego*), & Hannah Rohde (*The University of Edinburgh*)  
jet.hoek@ed.ac.uk

Coreference provides a measure of speakers' inferences and expectations about relationships that hold across sentences. Different approaches place different emphasis on the roles of meaning (Winograd 1972; Hobbs 1979) and form (Grosz et al. 1995)—two components which combine in the Bayesian Model put forward by Kehler et al. (2008). The Bayesian Model, in its strong form, posits the independence of a referent's predictability for re-mention and its likelihood of being mentioned with a pronoun. However, evidence regarding this independence is mixed. Here, we use a new context type to test (i) whether predictability influences pronominalization and (ii) whether Bayes Rule captures the relationship between pronoun interpretation and production.

A story continuation experiment (N=83) varied prompt type (pronoun vs. full-stop), to test participants' pronoun interpretations (1a), re-mention preferences (1b), and pronominalization rates (1b). We counterbalanced which referents were gender-matched (NP1&NP2, NP1&NP3, NP2&NP3).

- |   |                            |
|---|----------------------------|
| (1a) Adam scolded Diana for Russell. He _____ | [pronoun-prompt condition] |
| (1b) Adam scolded Diana for Russell. _____    | [full-stop condition]      |

We replicate two known patterns. First, the pronoun prompt yields more NP1 continuations than the full stop prompt ( $\beta=1.52$ ,  $p<.001$ ). Second, grammatical role influences pronominalization: the subject referent is preferentially re-mentioned with a pronoun. For question (i) on predictability~pronominalization independence, we compare referents' re-mention rates to the rates with which they are pronominalized. The re-mention rates of NP1 and NP2 do not differ ( $\beta=0.22$ ,  $p=.53$ ) but their pronominalization rates do ( $\beta=-3.26$ ,  $p<.001$ ); conversely, the re-mention rates of NP2 and NP3 differ ( $\beta=1.12$ ,  $p<.001$ ) but their pronominalization rates do not ( $\beta=0.19$ ,  $p=.42$ ). We thus find no evidence of any dependence between predictability and pronominalization.

For question (ii) on capturing the observed pronoun interpretation behavior, we follow Rohde and Kehler's (2014) methodology for computing interpretation estimates from the Bayesian Model and two alternative models. In contrast to prior work, the Bayesian Model is not the best fit for the observed pronoun interpretations. It is outperformed by the Mirror model, which posits that a speaker is licensed to use a pronoun to refer to a topical referent because the listener will interpret it to refer to that same topical referent. We are planning two follow-up studies to determine whether the difference between our result and previous work has more to do with the construction type or with the number of referents.

#### References

- Grosz, B.J., Joshi, A.K. & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21, 203–225.
- Hobbs, J. R. 1979. Coherence and coreference. *Cognitive Science* 3. 67–90.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and Coreference revisited. *Journal of Semantics* 25(1), 1–44.
- Rohde, H. & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience* 29(8), 1–16.
- Winograd, T. 1972. *Natural language understanding*. New York: Academic Press.



## The Processing of Grammatical Functions in Swedish is Expectation-based

Thomas Hörberg, Department of Linguistics, Stockholm University  
thomas\_h@ling.su.se

Language comprehension is expectation-based (e.g. Venhuizen et al. 2019). Statistical regularities in the linguistic input set up expectations that are utilized during incremental interpretation. A central part of language comprehension involves assigning grammatical functions (GFs) to NPs, thereby determining how participants are related to events or states. In many languages, speakers have many ways to encode GFs morphosyntactically (e.g. word order, case), and their encoding preferences depend on an interplay between NP properties (e.g. animacy) and verb semantic properties (e.g., volitionality) (Hörberg 2016). This creates complex statistical patterns in the distribution of these *GF information types* that can be utilized during on-line GF processing. In this talk, I present evidence indicating that GF assignment in transitive sentences in written Swedish is expectation-based, drawing upon such statistical patterns. I will present a corpus-based probabilistic model of incremental GF assignment in Swedish transitive sentences, together with results from a self-paced reading experiment, showing that the model's strongest predictions are confirmed by human processing preferences.

The model is based upon 16552 transitive sentences, extracted from a corpus of written Swedish, that were annotated for word order (SVO vs. OVS), GF information (e.g., animacy, definiteness, case), and verb semantic properties (e.g. volitionality, sentience). Based on the distribution of these features, estimates of the probability for SVO vs. OVS GF assignment at each sentence region (NP1, verb, NP2) were calculated, using logistic mixed effects regression modeling. In the model, these estimates are used to predict incremental processing costs related to the change in the expectation for a GF assignment at each sentence region. This is done in terms of *Bayesian surprise* - the relative entropy over the two possible GF assignments before and after seeing the constituent at hand (Kuperberg & Jaeger 2016). Bayesian surprise (over syntactic trees) has also been argued to underlie the correlation between word surprisal and both processing times (Smith & Levy 2013) and certain neural responses (e.g., the N400 effect, Frank et al. 2015).

In the self-paced reading experiment, 45 participants read 64 transitive sentences that varied with respect to word order (SVO vs. OVS), NP1 animacy (animate vs. inanimate) and verb class (volitional vs. experiencer). By-region reading times on NP1, the verb, and NP2 were well-described by the region-specific Bayesian surprise predicted by the model. For example, reading times in the NP2 region observed in locally ambiguous, object-initial sentences were mitigated when the animacy of NP1 and its interaction with the verb class bias towards an object-initial word order.

These findings indicate that on-line GF assignment draws upon statistical regularities in the previous language input, as predicted by expectation-based accounts.

- Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140, 1–11.
- Hörberg, T. (2016). Probabilistic and Prominence-driven Incremental Argument Interpretation in Swedish (PhD thesis, Stockholm University).
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59.
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302–319.
- Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019). Expectation-based Comprehension: Modeling the Interaction of World Knowledge and Linguistic Experience. *Discourse Processes*, 56(3), 229–255.

**On the Correlation of Context-Aware Language Models with the Intelligibility of Polish Target Words to Czech Readers**

Klára Jágrová, Marius Mosbach, Michael Hedderich, Tania Avgustinova, Dietrich Klakow  
Saarland University, Saarbrücken, Germany  
kjagrova@coli.uni-saarland.de

This contribution seeks to provide a rational probabilistic explanation for the intelligibility of words in a genetically related language that is unknown to the reader – a phenomenon referred to as intercomprehension. In this research domain, the intelligibility of stimuli was, among other factors, traditionally explained by linguistic distance and neighbourhood density of the stimulus towards a language in the reader's linguistic repertoire (e.g., Heeringa et al. 2013).

Jágrová & Avgustinova (2019) showed that predictability in context contributes to the intelligibility of the target items. They gathered data from web-based cloze translation experiments for 149 Polish sentences (Block and Baldwin, 2010). These were presented to Czech readers who were asked to translate the highly predictable target words in sentence final position. The majority of the items were more comprehensible within the sentences than if presented without context to another group of Czech respondents. However, for some target words the situation was reversed: the target word intelligibility in context decreased if compared to the condition without context. An error analysis revealed systematic patterns, such as L1/Ln interferences or perceived morphological mismatches. Most of them were in combination with the readers' priming by a dominant concept in the sentence.

Jágrová & Avgustinova (2019) correlated the intelligibility scores of the target words with surprisal values from 3-gram language models (LMs). Since 3-gram surprisal can explain predictability effects only using the two words preceding the target word, the overall correlations with surprisal are low. Interestingly, surprisal correlates stronger with intelligibility of target words that are non-cognates and false friends.

In this contribution we hypothesize that intelligibility of these highly predictable words will correlate better with surprisal values obtained from LMs which incorporate information from the entire sentence. We evaluate two context-aware LM architectures: LSTMs that can take long distance dependencies into account and Transformer based LMs which are able to access the whole input sequence at the same time. We investigate how their use of context affects surprisal and its correlation with intelligibility.

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, 42(3), 665–670. <https://doi.org/10.3758/BRM.42.3.665>

Heeringa, W., Golubovic, J., Gooskens, C., Schüppert, A., Swarte, F., & Voigt, S. (2013). Lexical and orthographic distances between Germanic, Romance and Slavic languages and their relationship to geographic distance. In C. Gooskens, & R. van Bezooijen (Eds.), *Phonetics in Europe: Perception and Production* (pp. 99-137). Frankfurt a. M.: P.I.E. - Peter Lang.

Jágrová, K., & Avgustinova, T. (2019). Intelligibility of highly predictable Polish target words in sentences presented to Czech readers. To appear in *Proceedings of CICLing: International Conference on Intelligent Text Processing and Computational Linguistics*. [http://www.coli.uni-saarland.de/~tania/ta-pub/CICLing\\_preprint\\_Jagrova\\_Avgustinova\\_2019.pdf](http://www.coli.uni-saarland.de/~tania/ta-pub/CICLing_preprint_Jagrova_Avgustinova_2019.pdf)

**Predictable words are more likely to be omitted in fragments –  
Evidence from production data**

*Robin Lemke, Lisa Schäfer, Heiner Drenhaus & Ingo Reich (Saarland University)*  
robin.lemke@uni-saarland.de

Uniform Information Density (UID, Levy & Jaeger 2007) has been shown to constrain the omission of optional function words. Our study extends this finding to the choice between fragments (Morgan 1973) (1a) and full sentences (1b).

- (1) [Passenger to taxi driver:]  
a. To the university, please.  
b. Drive to the university, please.

According to UID, speakers tend to distribute surprisal, defined as  $-\log_2 p(\text{word}|\text{context})$ , uniformly across their utterance. Undesirable surprisal minima caused by uninformative words are avoided by omitting these words. For instance, surprisal is distributed more uniformly in the fragment (1a) than in the corresponding sentence (1b) because the predictable “drive” is omitted. Investigating UID effects on omissions requires first a set of linguistic data containing the relevant omissions and, second, a method for estimating the surprisal of both the omitted and realized words in that data set. We collected such a data set with a production task. Subjects read 24 script-based stories like (2) based on event chains extracted from the DeScript corpus (Wanzare et al. 2016). They produced the utterance they considered most likely to be said by a specific character in this situation. For each of the stories we collected 100 responses.

- (2) Annika and Jenny want to cook pasta. Annika put a pot with water on the stove. Then she turned the stove on. After a few minutes, the water started to boil. Now Annika says to Jenny:

Responses were preprocessed by removing all non-content words, by lemmatizing verbs and nouns and by pooling synonyms (3). We estimated the surprisal of each word in the simplified representations with unigram language models trained on the data for each story separately. Models were trained on an enriched data set, where we inserted all content words that had been omitted in the original data despite being required in grammatical sentences (verbs and their arguments). Otherwise, predictable words would be rare just because they are frequently omitted.

- (3) Put the pasta into the pot → put pasta pot

We analyzed the data with a mixed effects logistic regression (lme4, R) that reveals a significant effect of surprisal on omission ( $z=2.7$ ,  $p<.01$ ): As predicted by our hypothesis, more likely words are more often omitted. We also discuss similar results obtained with a different surprisal metric based on Hale (2001), that additionally takes linguistic context into account. Taken together, our study extends evidence for UID from function words to more diverse omissions in fragments.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In Proceedings of NAACL (Vol. 2), pages 159-166. • Levy, R. P. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In Schölkopf, B., Platt, J., and Hoffman, T., editors, Advances in Neural Information Processing Systems, pages 849-856. MIT Press. • Wanzare, L. D. A., Zarccone, A., Thater, S., and Pinkal, M. (2016). DeScript: A crowdsourced corpus for the acquisition of high-quality script knowledge. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 16), 3494–3501.

**Information-theoretic approaches to communicative efficiency:  
some controversial issues and open questions**

Natalia Levshina, Leipzig University  
natalevs@gmail.com

This paper provides a critical perspective on the recent information-theoretic (IT) approaches to linguistic reduction and enhancement of linguistic forms (see an overview in Gibson et al. 2019), which are based on the assumption that language users act rationally, trying to maximize the benefit-to-cost ratio of their linguistic behaviour. In this paper I will argue that this approach, albeit very promising and illuminating, has several caveats and problematic issues, which will be illustrated by cross-linguistic data, such as the following:

- the problem of measuring the costs and benefits of linguistic communication;
- the existing emphasis on the immediate linguistic co-text (e.g. n-grams) in measuring predictability, disregarding the larger context and encyclopaedic knowledge;
- the scarcity of studies that could help us understand which linguistic phenomena can be explained by the language users' tendency to be communicatively efficient, and which should be attributed to other factors, e.g. automatization of articulation routines (Bybee 2006) or a mechanism that coordinates the speaker's planning of articulation (e.g. Bell et al. 2009);
- insufficient recognition of the role of intersubjective rationality in communication, where the speaker and the hearer rely on the mutual assumption of each other's efficient behaviour (Levshina 2018);
- a lack of consensus about the diachronic mechanisms that explain how efficient language structures emerge (cf. Cristofaro 2019).

In the second part, I address a more specific claim that highlights the importance of average contextual predictability (informativity) for explanation of different word lengths (Piantadosi et al. 2011; Gibson et al. 2019), while at the same time questioning the role of context-free frequency (Zipf 1935). Using large-scale English and Russian corpus data and a range of regression techniques, such as Poisson positive regression and ridge regression, I will show that a) both average predictability and frequency different types of probabilistic information play a role in predicting the length of a word; b) for particular word classes, the relative prominence of frequency is higher than that of contextual predictability.

## References

- Bell, A., Brenier, J., Gregory, M., Girand, C. & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92-111.
- Bybee, J. L. (2006). From usage to grammar: the mind's response to repetition. *Language*, 82(4), 711-733.
- Cristofaro, S. (2019). Taking diachronic evidence seriously: Result-oriented vs. source-oriented explanations of typological universals. In K. Schmidtke-Bode, N. Levshina, S.M. Michaelis & I. A. Seržant (Eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 25-46. Berlin: Language Science Press.
- Gibson, E., Futrell, R., Piantadosi, S., Dautriche, I., Mahowald, K., Bergen, L. & Levy, R. (2019). How Efficiency Shapes Human Language. *Trends in Cognitive Science*, 23(5), 389-407.
- Levshina, N. (2018). *Towards a Theory of Communicative Efficiency in Human Languages*. Habilitation thesis, Leipzig University. <http://doi.org/10.5281/zenodo.1542857>
- Piantadosi, S., Tily, H. & Gibson, E. (2011). Word lengths are optimized for efficient communication, *PNAS*, 108(9), 3526.
- Zipf, G. (1935). *The Psychobiology of Language: An Introduction to Dynamic Philology*. Cambridge, Mass.: M.I.T. Press.



## Verb Surprisal in the Visual World

Christine Muljadi, Christine Ankener, Les Sikos, Maria Staudte  
Saarland University  
masta@coli.uni-saarland.de

The effort for processing a word in a given linguistic context is correlated with its surprisal and predictability (Hale, 2001). Previous work has extended classical notions of surprisal to visually-situated contexts (Ankener, Sekicki & Staudte, 2018) in a visual world eye-tracking study which examined whether manipulations of referential entropy (i.e., the number of potential referents in the visual display) modulated the processing of the target noun. Participants heard an utterance (e.g., “The woman spills now the water.”) along with a varying number of objects in the display that were consistent with the verb. Eye movements during the verb indicated that listeners anticipated (only) suitable objects: Uncertainty about the upcoming referent was reduced by exploiting the verb constraint. This parametrically decreased processing effort (as measured by the pupillometric Index of Cognitive Activity (ICA); see also Demberg & Sayeed, 2016) at the sentence-final noun—but it did not affect processing effort at the verb. One explanation may be that the effect of referential entropy reduction is only manifest at nouns, because nouns serve as direct pointers to objects in the world, while verb selectional restrictions simply constrain expectations to a smaller set of possibilities.

The current study directly investigates this question using a setup similar to Ankener et al., (2018) and relative sentences (in German) because they allow the noun to appear before the verb (e.g. “Tell me if the \*rose\* that is \*watered\* by the figure is located at the top.”). Experimental displays varied the number of verb options (4, 3, 1, or 0) by depicting actions rather than objects. ICA-results revealed that processing effort at the target verb was reliably modulated by the number of objects in the display that were consistent with the mentioned noun, e.g. when only 1 action involving a rose was displayed, mean ICA values were significantly lower ( $m=42.9$ ,  $SD=13.1$ ) than when 3 rose-actions were shown ( $m=47.1$ ,  $SD=11.2$ ;  $p<0.01$ ). Typical anticipatory eye movements during the noun to likely upcoming actions/verbs were again observed, but as in Ankener et al. (2018), this did not appear to modulate processing effort at the noun.

These results indicate that visual context can similarly affect the predictability and surprisal of both verbs and nouns. We also replicate the lack of an effect on processing effort for the word that provides the constraining information. Thus, regardless of word class, processing effort seems to correlate with situated surprisal but not with referential entropy reduction.

Ankener, C., Sekicki, M., & Staudte, M. (2018). The influence of visual uncertainty on word surprisal and processing effort. *Frontiers in psychology*, 9, 2387.

Demberg & Sayeed (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PloS one*, 11(1), e0146194.

Hale, J. (2001). A probabilistic Early parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 1-8).

## Rational adaptation in lexical prediction: The influence of prediction strength

Tal Ness & Aya Meltzer-Asscher, Tel Aviv University  
talness@mail.tau.ac.il

Recent studies indicate that the processing of an unexpected word entails additional costs when the initial prediction was strong (e.g. Federmeier et al., 2007). This was suggested to stem from a commitment made to the initial (strong) prediction, requiring its inhibition in order to integrate the actual input (e.g. Ness & Meltzer-Asscher, 2018). Additional studies show that comprehenders rationally adapt their predictions in different situations (e.g. Delaney-Busch et al., 2019).

In the current study we hypothesized that since the disconfirmation of strong predictions incurs processing costs, it would also trigger adaptation mechanisms. We tested whether repeated disconfirmation of strong predictions throughout the experiment results in lesser commitment to predictions in later trials, reflected in reduced costs for unexpected words. The experiment (N=120) included two-word phrases in which the first word was either highly constraining (e.g. 'global' strongly predicts 'warming') or not (e.g. 'green' does not have any highly probable completion). The second word was unexpected (i.e. low cloze) in both cases, e.g. 'global epidemic' (HL - high constraint, low cloze), 'green pepper' (LL). Filler trials were manipulated between participants; half of the participants encountered a high proportion of HL trials, and half - a high proportion of LL trials. Participants had to respond whether the phrase was anomalous as quickly as possible after seeing the second word.

Results showed a main effect of constraint such that reaction times were higher in the HL trials relative to LL, demonstrating costs of disconfirming a strong prediction. Additionally, there was an interaction such that a high proportion of disconfirmed strong predictions reduced the processing costs incurred by HL trials, indicating that participants adjusted the strength of their predictions when strong prediction was discouraged.

We formulated a Bayesian adaptation model whereby inhibition cost was modeled as  $\mu * PE$ .  $\mu$  is the mean of a beta distribution representing the participant's belief (updated on each trial) about the likelihood of encountering the expected word (i.e. her current estimation of the predictive validity). The initial prior was  $\text{beta}(1, 1)$ , and updating occurs whenever the participant encounters a high-low trial:  $\text{beta}(1, 1 + \text{number of HL})$ . PE is the prediction error (the difference between the cloze probability of the most probable word, namely constraint, and the cloze probability of the presented word). We show that this model accounts for the trial-by-trial data, indicating that participants adapt by using their belief about predictive validity to weigh the strength of their subsequent predictions.

Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. R. (2019). Neural evidence for bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*, 187, 10-20.  
Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain research*, 1146, 75-84.  
Ness, T., & Meltzer-Asscher, A. (2018). Lexical inhibition due to failed prediction: Behavioral evidence and ERP correlates. *Journal of experimental psychology: learning, memory and cognition*, 44(8), 1269-1285.

## Item-bound vs Category-based Generalizations. An Entropy Model

Silvia Radulescu (Utrecht University), Efi Giannoupoulou (Utrecht University), Sergey Avrutin (Utrecht University), & Frank Wijnen (Utrecht University)  
S.Radulescu@uu.nl

What triggers the inductive leap from memorizing items and statistical regularities to inferring abstract rules? We propose an innovative information-theoretic model for both learning statistical regularities and generalizing to new input. Our entropy model hypothesizes that rule induction is an encoding mechanism triggered by the interaction between input entropy and the finite encoding power of the human brain (channel capacity).

While traditional cognitive psychology claimed that rule learning relies on encoding linguistic items as abstract categories [1], as opposed to learning statistical regularities between specific items [2], recent views converge on the hypothesis that it is one mechanism – statistical learning – that underlies both item-bound learning and abstract rule learning [3,4,5]. However, it is unclear how a single mechanism outputs two qualitatively different forms of encoding – item-bound and category-based generalizations, and what triggers the transition.

In our model, less input entropy facilitates finding regularities between specific items, i.e. item-bound generalization, while a higher entropy exceeding channel capacity drives category-based generalization. Rule learning is a phased mechanism that starts out by memorizing specific items and finding regularities between them (item-bound generalizations) and it gradually moves to an abstract category-based encoding, as a function of increasing input entropy.

In two artificial grammar experiments, we exposed adults to a 3-syllable XXY artificial grammar to probe the effect of input entropy on rule induction. We designed six language versions with different input entropy (from 2.8 to 4.8 bits). Participants gave grammaticality judgements on correct familiar XXY strings, correct new-syllable XXY, ungrammatical familiar-syllable X1X2Y, and ungrammatical new-syllable X1X2Y. Results showed that when input entropy increases, the tendency to infer abstract rules increases gradually. Also, in the lower entropy conditions participants correctly accepted familiar XXY, and correctly rejected familiar-syllable X1X2Y.

Since low entropy allows easy memorization, acceptance of familiar XXY might be supported by memory of the specific strings, not necessarily by item-bound generalization. To further test the hypothesis that low entropy facilitates item-bound generalization, we ran another experiment. One group was exposed to the lowest entropy (2.8 bits), and another group to a medium entropy condition (4.25 bits). But instead of familiar XXY strings, we tested familiar-syllable YYX. As expected, participants accepted familiar-syllable YYX strings in the low and medium entropy conditions, based on the *same-same-different* rule, but in the low entropy condition they accepted new-syllable XXY less than in medium entropy. The results support our model that low entropy facilitates item-bound generalization.

[1] Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77–80.

[2] Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical Learning by 8-Month-Old Infants. *Science* 274 (5294), 1926–1928.

[3] Aslin, R.N., and Newport, E. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21, 170–176.

[4] Aslin, R.N., and Newport, E. (2014). Distributional Language Learning: Mechanisms and Models of Category Formation. *Language Learning*, 64, 86–105.

[5] Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, 147, 70–74.

## Are pragmatic inferences triggered by informationally redundant utterances effortless?

Margarita Ryzhova (Saarland University), Vera Demberg (Saarland University)  
mryzhova@coli.uni-saarland.de

Current models of language comprehension typically associate processing difficulty with word predictability, or memory-related measures such as the length of dependency links or word familiarity. In the pragmatics community, it has long been debated whether pragmatic inferences are seamless, or whether the pragmatic inferencing process, especially when concerns particularized implicatures, is related to additional processing costs. The processing cost of particularized implicatures is currently under-researched – most results have been obtained for scalar implicatures, for which some studies report no cost [3,5], while others do report increased processing costs [1,2, 6].

The core idea of our experimental study is as follows: if pragmatic inferences are indeed difficult, fewer inferences or less strong ones, should be observed when cognitive resources are reduced [1,2, 6]. In a dual-task study including mouse-tracking and language comprehension, we test for pragmatic implicatures related to informationally redundant utterances [4].

The experimental design and English stimuli from [4] were translated to German for this study. Stories establish a particular topic, thus making some topic-related activities a priori highly predictable. For example, given “*going to the swimming pool*” scenario, “*bringing a swimsuit*” activity is anticipated from world knowledge. In stories, we manipulate the presence or absence of the informationally redundant (IR) utterance which describes topic-related activity (“*Lisa brought her swimsuit!*”), and ask participants to rate how strongly they would assume that person mentioned in the story usually performs the IR activity. In the high load condition, participants perform a mouse tracking task while listening to a story. In the low load condition, they perform only listening.

Data analysis of ninety-eight German-native speakers showed a main effect of IR utterance ( $\beta = -21.97$ ,  $t = -6.14$ ,  $p < .001$ ). Thus, habituality estimates are significantly lower in the with-IR condition, showing that participants accommodated the presence of an IR utterance by altering their prior beliefs about activity typicality (i.e., they inferred that Lisa frequently forgets her swimsuit). This finding replicates results on single task in English by [4]. Contrary to our expectations, participants did draw pragmatic inferences under high load condition too. Moreover, their ratings were significantly stronger than under low load ( $\beta = -8.07$ ,  $t = -2.08$ ,  $p < 0.05$ ). Thus, we find no evidence for cost. Although, these findings, while not supporting the need for a cost function related to these implicatures, do bring up questions for how to model the increased size of such pragmatic inferences under load in rational models of language comprehension.

1. Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, 51(3), 437-457. 2. Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: Further evidence that scalar implicatures are effortful. *The Quarterly Journal of Experimental Psychology*, 64(12), 2352-2367. 3. Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). “Some,” and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, 116(1), 42-55. 4. Kravtchenko, E., & Demberg, V. (2015). Semantically underinformative utterances trigger pragmatic inferences. In *CogSci*. 5. Marty, P., & Chemla, E. (2013). Scalar implicatures: working memory and a comparison with only. *Frontiers in psychology*, 4, 403. 6. De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental psychology*, 54(2), 128-133.

**Acknowledgments** This research was funded by the German Research Foundation (DFG) as part of SFB 1102 ‘Information Density and Linguistic Encoding’.



**Speakers use verb phrase ellipsis to satisfy UID:  
Psycholinguistic evidence from length and context effects**

Lisa Schäfer, Robin Lemke, Heiner Drenhaus & Ingo Reich (Saarland University)  
lisa.schaefer@uni-saarland.de

We investigate the hypothesis that speakers use verb phrase ellipsis (VPE) (1a) instead of full forms (1b) to avoid redundancy by omitting predictable material from utterances.

- (1) Sam played football *in the backyard of the house* and ...  
 a. ... Dean did, too. (ellipsis)  
 b. ... Dean played football *in the backyard of the house*, too. (full form)

**Account.** According to the Uniform Information Density (UID) hypothesis (Levy & Jaeger 2007), speakers distribute information (Shannon 1948) as uniformly as possible across utterances. Specifically, they use ellipsis to minimize troughs in the ID profile which are caused by low, i.e. redundant information. As information indexes processing effort (Levy 2008), such troughs would hamper the listener's comprehension. UID predicts that both utterances with longer troughs – caused by a long redundant VP compared to a short VP – and with deeper troughs – caused by utterances in a context that makes them predictable vs. a neutral context – are degraded. We test these predictions in our studies.

**Length effects.** A  $2 \times 2$  (COMPLETENESS: full form vs. VPE  $\times$  LENGTH: long vs. short) rating study with 41 subjects and 32 items like (1) (the part in italics was added for the long condition) confirmed that VPE was rated as better than the full form particularly in the long condition ( $\chi^2 = 7.66$ ,  $p < .01$ ) where the repetition of the VP would be more redundant. A self-paced reading study with 87 participants on just the full forms indicates that this preference is indeed related to troughs: We found that the second conjunct was read particularly faster (residualized cumulative reading times (Gibson & Levy 2016)) in the long condition ( $\chi^2 = 75.18$ ,  $p < .001$ ). This indicates that the processing effort is lower on average in the long condition which has also been reflected in degraded ratings for the long full forms.

- (2) a. Sam and Dean dream of becoming NFL quarterbacks some day. (predictive)  
 b. Sam and Dean dream of becoming president some day. (neutral)

**Context effects.** We conducted a  $2 \times 2$  (COMPLETENESS  $\times$  CONTEXT: predictive vs. neutral) rating study with 95 subjects and 24 items plus pretested predictive and neutral contexts (2a vs. 2b). VPEs were rated as better than the full form particularly in a predictive context ( $\chi^2 = 4.58$ ,  $p < .05$ ): The effect of predictability is significantly stronger for elliptical than for non-elliptical utterances. Surprisingly, a self-paced reading study did not reveal a significant difference in processing effort indexed by reading times of the second conjunct between the predictive and the neutral condition. In our presentation, we discuss possible explanations such as a prevailing parallelism effect overwriting the context manipulation.

**Summary.** Our data provide additional evidence for the effect of UID on encoding preferences: Speakers use ellipsis to avoid troughs in the ID profile. As hypothesized, the preference for omission increases in predictive contexts and for longer redundant elements.

Gibson, E. & Levy, R. (2016). An attempted replication of Hackl, Koster-Hale, Varvoutis (2012). arXiv preprint arXiv:1605.00178. • Levy, R. (2008). Expectation-Based Syntactic Comprehension. In: *Cognition* 106(3): 1126-1177. • Levy, R. & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In Schlökopf, B., Platt, J. C., & Hoffman, T., editors: *Advances in Neural Information Processing Systems* 19: 849-856. • Shannon, C. E. (1948). A mathematical theory of communication. In: *Bell system technical journal* 27(3): 379-423.

## A Cost Metric for Pronoun Resolution: Uncertainty Increases Processing Cost

Olga Seminck, Pascal Amsili, Université Catholique de Louvain, Université Paris 3

olga.seminck@cri-paris.org

Pronoun resolution is the process of finding the antecedent of an anaphoric pronoun. Whereas it is almost always successful — it seems even the case that the use of pronouns contributes to the fluidity of a discourse (1) — there are differences in the facility of resolution between different pronouns demonstrated by small differences in processing times. For example, in English, subject antecedents are processed quicker than object antecedents (2).

However, the list of contributing factors is long and it is difficult to combine all factors to make predictions about pronouns in corpora. Therefore, we investigated whether a broader hypothesis about processing cost induced by anaphoric pronouns can make relevant predictions when it is implemented as a cost metric.

Our cost metric is based on the following hypothesis: more uncertainty about the antecedent of a pronoun leads to higher processing cost. We use the notion of entropy (3) to estimate uncertainty: the antecedent of a pronoun is modeled as a random variable that can take the value of different discourse referents. Each of these referents can be attributed a probability that it is actually the antecedent and then the pronoun's entropy can be calculated (see the formula  $H(\text{pro})$  at the bottom of this page).

An issue with this proposal is that entropy increases when the number of discourse referents rises and will thus systematically be higher further in the text. Therefore, we propose to make a small modification on the metric and take the relative entropy (3) in which the 'distance' between the actual entropy and the maximal entropy is measured (see the formula of  $H_{\text{relative}}$  at the bottom of this page). So, when the relative entropy is low, we predict more processing cost.

We tested the uncertainty hypothesis on the English part of the Dundee Corpus (4), a corpus containing reading times of ten native speakers of English. For each anaphoric pronoun in the corpus, we estimated the probabilities of every discourse referent occurring in the text before the pronoun with a state of the art coreference resolution system (5). We used a Bayesian mixed model (6) to test whether the cost-metric contributed to the prediction of pronoun reading. We found that lower relative entropy lead to more participants fixating a pronoun: a result in line with the uncertainty hypothesis (95% credible interval).

This work illustrates that uncertainty about the antecedent influences pronoun resolution. It also illustrates how notions from Information Theory can make relevant predictions about human language processing and how NLP-systems can be used as robust tools to estimate probabilities of language.

$$H(\text{pro}) = - \sum_{a \in A} P(\text{pro} = a) \cdot \log_2(P(\text{pro} = a)) \quad H_{\text{relative}}(P \parallel Q) = \sum_{i \in P \wedge i \in Q} P(i) \log \frac{P(i)}{Q(i)}$$

- (1) Gordon, P. C., & Chan, D. (1995). Pronouns, passives, and discourse coherence. *Journal of Memory and Language*, 34(2), 216-231. (2) Crawley, R., Stevenson, R., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 4, 245-264. (3) Cover, T. M., & Thomas, J. A. (2012). *Elements of Information Theory*. John Wiley & Sons. (4) Kennedy, A., Hill, R., & Pynte, J. (2003). The Dundee Corpus. In *Proceedings of the 12th European conference on eye movement*. (5) Lee, K., He, L., Lewis, M., & Zettlemoyer, L. (2017). End-to-end Neural Coreference Resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 188-197. (6) Bürkner, Paul-Christian (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. In: *Journal of Statistical Software* 80(1), 1–28.

## The Negative Existential Cycle in Indo-European: Is a rational approach enlightening?

Shahar Shirtz, University of Oregon and Annemarie Verkerk, Saarland University  
shahar.shirtz@gmail.com, annemarie.verkerk@uni-saarland.de

This paper examines the Negative Existential Cycle (NEC, Croft 1991) in a broad sample of Indo-European languages, testing its validity both quantitatively and qualitatively. The NEC is a typological hypothesis on the diachronic relationship between different types of negative existential constructions and their relation to standard verbal negation. It recognizes six construction types and posits a unidirectional pathway across them. Recent work (Veselinova 2013, 2014, 2015, 2016) finds that the NEC often does not take the form of a cycle: the six stages of the NEC do not necessarily follow each other, languages can have different constructions belonging to different types, and there is variation in stability of the stages.

We first present a phylogenetic comparative analysis testing whether Croft's NEC explains the attested distribution of negative existential construction types better than alternative models. We discuss the requirements that Bayesian MCMC methods (*Multistate* in *BayesTraits*, Pagel and Meade 2004) impose on the sample size and the amount of crosslinguistic variation. Then, we discuss the modeling of languages with multiple negative existential construction types and the interaction between the definition of NEC construction types and models of change.

Combining quantitative and qualitative perspectives, we then ask whether our ancestral state estimations match "traditional", analytic approaches to morphosyntactic reconstruction (following, e.g., Barðdal & Gildea 2015). We ask whether we can find evidence for the origins of special negative existential constructions in older stages of languages as well as in contemporary relatives. Further, we ask whether the pathways to certain stages and construction types can be motivated by processes external to the NEC, such information-structure or interactional principles (e.g., prominence of negative quantifiers and interjections). We test whether these principles can explain instances where historical change seemingly "skips" stages in the cycle, or whether there is evidence for a rapid burst of morphosyntactic change. We then assess how well the quantitative results stand up in the face of the qualitative findings, and discuss whether a rational approach is indeed enlightening or not.

## References

- Barðdal, Johana & Gildea, Spike. (2015). Diachronic Construction Grammar: Epistemological context, basic assumptions and historical implications. In: J. Barðdal, S. Gildea, & L. Sommerer (Eds.) *Diachronic Construction Grammar*. Amsterdam: John Benjamins.
- Croft, William. (1991). The evolution of negation. *Journal of Linguistics*, 27, 1-27.
- Pagel, Mark, Meade, Andrew & Barker, Daniel. (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*, 53(5), 673-684.
- Veselinova, L.. (2013). Negative existentials: A cross-linguistic study. *Rivista di Linguistica*, 25, 107-145.
- Veselinova, L. (2014). The Negative Existential Cycle revisited. *Linguistics*, 52(6), 1327-1389.
- Veselinova, L. (2015). Special negators in the Uralic languages: Synchrony, diachrony and interaction with standard negation. In Matti Miestamo, Anne Tamm, & Beáta Wagner-Nagy (eds.), *Negation in Uralic languages*, 547-600. Amsterdam: John Benjamins.
- Veselinova, L. (2016). The Negative Existential Cycle through the lens of comparative data. In Elly van Gelderen (ed.), *Cyclical change continued*, 139-188. Amsterdam: John Benjamins.

## Reevaluating Pragmatic Reasoning in Web-based Language Games

Les Sikos, Noortje Venhuizen, Heiner Drenhaus, and Matthew Crocker (Saarland University)  
sikos@coli.uni-saarland.de

Results from work testing formalizations of Gricean maxims [1] using web-based reference games are mixed. Some studies indicate Bayesian (e.g., rational speech act (RSA)) models closely predict human (pragmatic) behavior [2]; others suggest participants rarely go beyond literal meanings in such studies [3-4]. For instance, [2] presented participants visual displays containing three objects (e.g., green-square, green-circle, blue-circle) and manipulated the number of shared features (e.g., shape, color). Using a one-shot paradigm (each participant sees a single trial), they collected separate judgments from speakers, listeners, and for salience. Results of the RSA model, which combines a speaker model (likelihood that speakers use a particular word to refer to the target) with empirically measured salience, were highly correlated with aggregate listener judgments ( $R=0.99$ ). This was interpreted as indicating that participants reasoned pragmatically. However, the reasoning required in [2] ranged from simple to more complex, thus the close fit of predicted to observed results might be driven by the simpler inferences. Consistent with this, [3] attempted a close replication of [2] focusing on more challenging items and found the basic RSA model was a poor predictor. Furthermore, [4] found that while listeners responded pragmatically in simpler conditions, they were at chance in more complex conditions.

We investigated whether listeners in such tasks reason as pragmatically as presumed. Experiment 1 ( $N=3387$ ) employed the same general methods as [2] and compared observed responses to predictions from the basic RSA model and a Literal Listener (LL) model that does not incorporate a model of the speaker. This basic LL model predicts that listeners should be equally likely to select any referent that a given word (e.g. “green”) can refer to. Because RSA relies heavily on salience, we also tested a LL+Salience model that weights its probabilities based on salience. Results showed that while RSA provided a good fit to the entire dataset (replicating [2]), both LL models performed better. Furthermore, when we analyzed only the more challenging conditions, LL+Salience performed best.

One possible explanation for this result is that one-shot web-based experiments do not engage the depth of pragmatic reasoning seen in typical human interactions. Experiment 2 ( $N=814$ ) investigated this question by testing whether increasing participant engagement leads to more pragmatic responses. Results indicate RSA performed better in Experiment 2, but still not as well as LL+Salience. Taken together, these findings indicate that a simpler model than RSA can better explain human behavior in such studies.

### References

- [1] Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). New York, NY: Academic Press.
- [2] Frank, M.C., & Goodman, N.D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336 (6084), 998–998.
- [3] Qing, C., & Franke, M. (2015). Variations on a bayesian theme: Comparing bayesian models of referential reasoning. In *Bayesian natural language semantics and pragmatics* (pp. 201-220). Springer International Publishing.
- [4] Frank, M.C., Emilsson, A.G., Peloquin, B., Goodman, N.D., & Potts, C. (2016). Rational speech act models of pragmatic reasoning in reference games. Retrieved from *PsyArXiv*.



## The effect of cue validity during input on sentence choice during output

Malathi Thothathiri

malathi@gwu.edu

Speech requires converting a thought into a sequence of words. Contrasting theories have posited two possible routes for this conversion [1]. Meaning could be mapped first to an abstract structural frame that is only subsequently populated with specific words. Alternatively, word retrieval could occur first, leading to sentence construction based on lexically specific syntactic information. Available evidence supports both sides [1,2]. Thus, whether sentence formulation relies on abstract structural versus verb-specific representations is debated. We tested a rationalist alternative to this dichotomy: The cue validity [3] hypothesis states that structural choices during speech (e.g., double-object versus prepositional-object dative) will prioritize different cues according to how reliably those cues predict structure when listening to input [4]. Competition between cues during learning can cause neural networks to reorganize such that reliable cues come to guide language production more than less reliable cues. Thus, sentence production could flexibly depend on one or the other kind of the representation depending on the input.

The studies used a language exposure + sentence production paradigm. During language exposure, participants watched videos and heard and repeated the accompanying sentences. We manipulated the statistical properties of the input, specifically how well an individual verb versus broader (e.g., semantic) properties predicted the structure. During the subsequent sentence production phase, participants watched new videos and described them as they saw fit. Analysis of the structures produced allowed us to evaluate whether participants adhered to verb-specific patterns or verb-general rules as predicted by the cue validity manipulation.

Study 1 involved artificial languages where the exposure phase indicated higher validity for verb-specific than verb-general cues, or vice versa. Consistent with our hypothesis, participants' subsequent sentence production adhered to the statistical preference of each verb in the former case, and to verb-general mappings between event type and structure in the latter case. Studies 2 and 3 extended the investigation to a natural language (English). In study 2, adults' production of dative sentences followed verb-specific or verb-general patterns depending on the relative validities of the cues. In study 3, four- and five-year-old children's production of dative sentences followed or overrode verb-specific patterns depending on whether individual verbs predicted structure reliably during exposure.

Together, these results provide evidence that the sentence production architecture can reorganize flexibly, using alternative pathways as dictated by cue validity. We are exploring this flexibility using neural measures as well. Future endeavors could add computational modeling to obtain a comprehensive picture of how speakers choose sentence structures.

[1] Konopka, A. E., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology*, 58(1), 68-101.

[2] Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, 56(3), 165-209.

[3] Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157-193). Hillsdale, NJ: Erlbaum

[4] Thothathiri, M., & Rattinger, M. G. (2016). Acquiring and producing sentences: Whether learners use verb-specific or verb-general information depends on cue validity. *Frontiers in Psychology*, 7.

## Age differences in language comprehension: older adults ‘predict more’ in a challenging listening condition

Pratik Bhandari<sup>1,3</sup>, Vera Demberg<sup>1,2</sup>, Jutta Kray<sup>3</sup>  
<sup>1</sup>Department of Language Science and Technology  
<sup>2</sup>Department of Computer Science  
<sup>3</sup>Department of Psychology  
 Saarland University, Germany  
 pratik.bhandari.np@gmail.com

The predictive processing hypothesis posits that the brain continuously makes top-down predictions of future events based on prior knowledge, experience, and contextual information. In language comprehension, incremental predictions of upcoming linguistic events (e.g., words) are generated based on the prior lexical or linguistic knowledge, world knowledge, and the contextual information available in the utterance (Kuperberg & Jaeger, 2016). However, in addition to top-down prediction, language comprehension is also modulated by bottom-up auditory processes such as the quality of the acoustic signal (Obleser & Kotz, 2010).

In the present study, we aim to investigate how top-down prediction and bottom-up auditory processes influence spoken language comprehension and how their interaction changes across the adult lifespan. It has been shown that compared to younger adults, older adults benefit from contextual information in adverse listening condition (Sheldon et al., 2008). In such a condition, they rely more on top-down semantic prediction. This can be attributed to their broader world knowledge, linguistic knowledge and their experience in spoken language conversation. Twenty-eight older adults (age 70 or above) and twenty-four younger adults (age 18 to 30) participated in an auditory sentence comprehension task. To investigate the independent effects as well as the interaction of sentence predictability (in terms of cloze probability) and intelligibility (in terms of different levels of noise vocoding), we exploited a factorial design with three levels of predictability (high, mid and low) and four levels of intelligibility. Independently, the sentences were noise vocoded to four different bands: 1-, 4-, 6-, and 8-band noise vocoding. Simple subject-verb-object sentences of German were presented and participants were asked to report the final word of the sentence, i.e. the noun.

We expected that response accuracy would be highest at the highest level of intelligibility (i.e. in 8-band noise-vocoded sentences), and also that the accuracy would be highest for highly predictable sentences. More importantly, we expected an interaction of predictability and intelligibility: older adults would rely more on top-down prediction than younger adults when the listening condition was difficult, i.e. we expected the largest age differences for the moderate intelligibility condition. Preliminary results of an ongoing study are in line with our expectations.

### References:

- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59.
- Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20(3), 633–640.
- Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*, 123(1), 489–499.

## Found in Interpreting: Detection and analysis of translationese using computational language models

*Yuri Bizzoni, Elke Teich, Tom Juzek, Katrin Menzel and Heike Przybyl*  
*University of Saarland*  
 yuri.bizzoni@uni-saarland.de

Translation is a special case of language production resulting in a so-called “third code” that shows traces of the source language (SL) while trying to conform to the conventions of the target language (TL) (House 2010). There is a wealth of descriptive, corpus-based work on SL shining-through and TL normalization in written translation, converging on observations of skewed frequency distributions of selected linguistic features in translations compared to original (non-translated) productions (“translationese”). Existing computational approaches have applied automatic classification and clustering (Rabinovich/Wintner 2015, Volansky et al. 2015, Rubino et al. 2016) to tell translations from non-translations or to guess the SL of a given TL text. With few exceptions (He et al. 2016), computational accounts typically focus on written translation and start from handcrafted, often shallow features (e.g. sentence length, type-token ratio).

We here advocate an empirical approach using computational language models (LM) to detect and interpret linguistic effects of SL shining-through and TL normalization at linguistically more informative levels (lexis, grammar). Also, we are primarily interested in simultaneous interpreting because it poses fairly severe constraints on processing and working memory (Hyönä et al. 1995) and will thus show stronger shining through effects and weaker normalization effects than written translation and possibly other effects directly stemming from cognitive limitations.

Using a richly annotated version of the EuroParl Corpus (Karakanta et al. 2018) and a comparable corpus of interpreting transcripts, we compare professional, written translations with professional, simultaneous interpretations of European Parliament speeches from English into German. We build language models for English originals, German translations and interpretations and German originals and compare models in terms of perplexity/relative entropy as follows:

- (a) to assess the relative, overall difference between translation and interpreting, we compare word-based and part-of-speech based ngram LMs for translations vs. interpretations in the target language;
- (b) to assess effects of TL normalization, we compare word-based and part-of-speech based LMs of (i) translations vs. originals and (ii) interpretations vs. originals in the target language;
- (c) to assess effects of SL shining-through, we compare part-of-speech based LMs of (i) translations vs. originals and (ii) interpretations vs. originals in the source language.

We find, for instance, that overall, interpretations exhibit less natural (more surprising) word order choices (due to SL shining-through) and tend to prefer paratactic structures, while translations tend towards hypotactic structures (a trace of TL normalization), indexed by unusual ngrams and the choice of particular conjunctions.

## Dynamic Formant Trajectories in German Read Speech: Impact of Predictability and Prosody

Erika Brandt<sup>1</sup>, Bernd Möbius<sup>2</sup>, and Bistra Andreeva<sup>2</sup>  
<sup>1</sup>Friedrich Schiller University Jena, <sup>2</sup>Saarland University  
erika.brandt@uni-jena.de

Phonetic structures increase in their duration and distinctiveness when they are difficult to predict from context compared to easily predictable structures (e.g. [1]). In American English, contextual predictability impacts on word [4] and segment duration [3]. American English vowels are also strengthened in their spectral features when they are difficult to predict from their context compared to easily predictable vowels [1,2,5]. Closely related languages, such as German [7] and Dutch [8], also seem to show the same positive relationship between vowel dispersion and predictability.

These predictability effects interact with prosodic factors, such as prosodic prominence [1,6]. Stressed segments that are highly predictable tend to be longer than unstressed, less predictable segments. These studies also show that the effects of predictability prevail even after controlling for known prosodic effects on phonetic structure.

Phonetic research from an information-theoretic perspective has so far neglected dynamic characteristics of the acoustic-phonetic signal. The present paper focuses on the effect of predictability on formant dynamics, while controlling for known effects of prosodic factors on vocalic characteristics. We use a German corpus of read speech testing the hypothesis that vowels show increased formant movement when they are difficult to predict. Interaction effects between predictability and prosody are expected. Vowel-inherent spectral change (VISC), vowel-section length (VSL), formant velocity, and discrete cosine transform (DCT) coefficients are used as metrics. Predictability is measured as surprisal based on a German language model, while controlling for word frequency. Prosodic factors are primary lexical stress, boundary, and speech rate. Results show that easily predictable German vowels have less formant change in VSL, F1 slope and velocity, and are less curved in their F2. Vowels in high-frequency words have less formant change in all measures. Interaction effects between surprisal and prosodic factors are observed, in some cases, even if there was no main effect of surprisal. This work shows that spectral expansion due to predictability effects cannot only be observed in single, pointwise measures of the spectrum but also in formant trajectories.

- [1] Aylett, M., and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, 119, 3048–3058.  
[2] Clopper, C. G., and Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America*, 124(3), 1682–1688. [3] Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates, *Laboratory Phonology*, 6(2), 243–278. [4] Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? Phonological neighbourhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4), 789–806. [5] Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: evidence from reduction in lexical production. In Bybee & P. Hopper (Eds), *Frequency and the Emergence of Linguistic Structure*, J. (pp. 229–254), Amsterdam: Benjamins. [6] Malisz, Z., Brandt, E., Möbius, B., Oh, Y., and Andreeva, B. (2018). Dimensions of segmental variability: interaction of prosody and surprisal in six languages, *Frontiers in Communication / Language Sciences*, 3, article 25, 1-18. [7] Schulz, E., Oh, Y. M., Malisz, Z., Andreeva, B., and Möbius, B. (2016). Impact of prosodic structure and information density on vowel space size. *Proceedings of Speech Prosody*, Boston, 350–354. [8] van Son, R. J. J. H., Bolotova, O., Pols, L. C., and Lennes, M. (2004). Frequency effects on vowel reduction in three typologically different languages (Dutch, Finnish, Russian). *Proceedings of Interspeech*, Jeju Island, Korea, 1277–1280.

## Pragmatically Informative Frame Identification

<sup>1,2</sup>Xudong Hong, <sup>1</sup>Frances Yung

<sup>1</sup>Saarland University, <sup>2</sup>Max Planck Institute for Informatics  
xhong@coli.uni-saarland.de

Frame identification is a semantic role labelling task based on FrameNet (FN) [1]. Given the context around a predicate (e.g. ‘the woman is *preparing* a meal.’), the task is to assign a frame (here, ‘*Cooking\_creation*’) which is a label for a prototypical situation. State-of-the-art frame identification systems trained on FN dataset obtain strong performance on in-domain text [2] but the coverage of the current FN lexicon is very limited. Performance of frame identification systems drops significantly when an ambiguous predicate is not in the FN lexicon [3] or not in the domain of FN, e.g. image captions.

To leverage this problem, we consider a frame identifier trained on a balanced corpus like FN dataset as a model of general semantic meanings to understand a predicate. When an out-of-domain context is given, humans cannot fully understand the predicate purely based on semantic information and require pragmatic inference to disambiguate the predicate. The recent proposals claim that this inference process can be modelled with recursive Bayesian inferences [4, 5]. We hypothesize that Bayesian pragmatic inference can improve out-of-domain frame identification.

We model the frame identification task as a communication game with two rational agents using Rational Speech Acts framework [6, 7]. The speaker describes a situation with a predicate given the context. The listener interprets the sentence and chooses a frame for the predicate. We consider the frame identifier [8] trained on FN 1.5 as the literal listener ( $L_0$ ):  $P_{L_0}(f|p)$  which predicts a frame  $f$  given a predicate  $p$ . The pragmatic speaker ( $S_1$ ) produce a predicate given a frame with a prior  $P_{S_1}(p)$  in the speaker’s domain:  $P_{S_1}(p|f) = P_{S_1}(p)P_{L_0}(f|p) / \sum_{i \in P_{S_1}} P_{S_1}(p_i)P_{L_0}(f|p_i)$ . The pragmatic listener ( $L_1$ )

reasons about the predicate that the pragmatic speaker would generate and infer the most possible frame given this predicate:  $P_{L_1}(f|p) \sim P_{S_1}(p|f)P_{S_1}(f)$ . We then annotate 126 verbal predicates in MS COCO captions [9] that are not in FN as a test set. Finally, we test our models by performing zero-shot frame identification. Results show that the pragmatic listener (accuracy=6.72%) outperforms the literal listener (accuracy=5.47%). We believe that our method can be generalised to other out-of-domain language understanding tasks.

[1] Fillmore, C. J., Johnson, C. R., & Petruck, M. R. (2003). Background to framenet. *International journal of lexicography*, 16(3), 235-250. [2] Hermann, K. M., Das, D., Weston, J., & Ganchev, K. (2014). Semantic frame identification with distributed word representations. In *Proceedings of the ACL 2014* (1) (pp. 1448-1458). [3] Hartmann, S., Kuznetsov, I., Martin, T., & Gurevych, I. (2017). Out-of-domain framenet semantic role labeling. In *Proceedings of the EACL 2017* (1) (pp. 471-482). [4] Jäger, G. (2012). Game theory in semantics and pragmatics. *Semantics: An international handbook of natural language meaning*, 3, 2487-2425. [5] Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, 5(1), 173-184. [6] Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998-998. [7] Lassiter, D., & Goodman, N. D. (2015). How many kinds of reasoning? Inference, probability, and natural language semantics. *Cognition*, 136, 123-134. [8] Botschen, T., Gurevych, I., Klie, J. C., Sergieh, H. M., & Roth, S. (2018). Multimodal frame identification with multilingual evaluation. In *Proceedings of the NAACL 2018* (1) (pp. 1481-1491). [9] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Proceedings of the ECCV 2014* (pp. 740-755). Springer, Cham.



## An attempt at bridging the gap between semantics and neurolinguistics – a probabilistic event semantics for the Retrieval-Integration model

Ralf Naumann and Wiebke Petersen

Institut für Sprache und Information, Universität Düsseldorf, Germany  
naumann@phil-fak.uni-duesseldorf.de

Neurophysiological research on human language processing poses at least the following three challenges for traditional formal semantic theories. First, understanding a sentence consists in the construction of a mental model of the situation that cannot be reduced to the sentence's propositional (predicate-argument) structure. Second, language comprehension is probabilistic in nature, modelling predictions. Third, according to the Retrieval-Integration model, comprehension proceeds in biphasic cycles of a retrieval and a subsequent integration operation that are indexed by the N400 and P600 effect, respectively. In our talk we outline a decompositional event semantics based on probabilistic relational models that overcomes the first two problems and show how it relates to the Retrieval-Integration model.

Situation models instantiate a script (e.g. going to a restaurant). They are represented as probabilistic relational models of a particular type. Relation models for situation models have two reference slots: 'actions' and 'participants', that link the complex event denoted by the model to its constituent events (action sequence) and to the objects participating in these constituent events, respectively. By contrast, reference slots in relational models for single events are defined on thematic roles like 'Actor' or 'Theme'. Reference slots in relational models for situation models generalize over the reference slots in relational models for (single) events. E.g., 'participants' generalizes over thematic roles like 'actor'. On the set  $SM(E)$  of situation models (events) of a particular type an accessibility relation  $\rightarrow$  is defined:  $sm \rightarrow sm'$  if  $sm'$  contains all the information contained in  $sm$  and  $sm'$  is the result of interpreting a word (a decompositional predicate) in the context of  $sm$ ; additionally, probability distributions on transitions  $Pr(SM, \rightarrow)$  and  $Pr(E, \rightarrow)$  are defined.

We distinguish two aspects of predictions (i) uncertainty: to which degree does new information reduce uncertainty about the current situation model, i.e. which situation model is actually described by the text?, and (ii) validation (verification): to which degree does new information fit into the situation model built so far? Whereas uncertainty is always related to  $(SM, \rightarrow)$  and its possible continuations, validation is related to the current event model. We will discuss the following two hypotheses. Uncertainty is quantized by entropy reduction over the probability distribution on  $(SM, \rightarrow)$ . Full or n-step entropy reduction involves n-step transitions on  $\rightarrow$  over those events possibly to be encountered in the continuation of the text/discourse. In contrast, validation is quantized by surprisal on  $(E, \rightarrow)$  involving only 1-step transitions. Retrieval (N400 effect) is modelled as updating the current situation model and hence as a transition in  $(SM, \rightarrow)$  based on  $Pr(SM, \rightarrow)$  whereas integration (P600 effect) is related to  $Pr(E, \rightarrow)$ . The differences between the N400 and P600 effect result from differences at the level of probability distributions: (i) information metrics used, (ii) different domains on which the distributions are defined and (iii) different levels of abstraction.

## Preliminaries to Acceptance-based Pragmatics

Sylvie Saget, University of Gothenburg

sylvie.saget@gu.se

While processing good enough linguistic representation, dialogue partners rely on different perspectives - their own point of view, their addressees' one, common/shared beliefs - or on existing linguistic representations built during preceding interactions. Explaining such a phenomenon is still a challenge. In this work, we propose preliminaries for a rational model of dialogue based on a distinction between belief and acceptance. Central to this model is the inclusion of acceptance. Acceptance here differs from the speech act of assent, ie. agreeing to a proposal whether or not this agreement is in line with one's mental state. Acceptance has been initially designed as a belief-like mental representation aiming at encapsulating knowledge involved in practical reasoning (Cohen, 1989; Paglieri, 2006; Saget, S. & Guyomard, M., 2006 & 2007). This model is called Acceptance-based Pragmatics. We demonstrate that adding the notion of acceptance enables to support different kinds of backgrounds, notably mixing perspective-taking and reuse.

We will firstly make explicit initial motivations and principles of Acceptance-based Pragmatics. Extending the set of belief-like mental attitudes used to characterize knowledge (privilege and shared) enables to take into account the full diversity regarding the kind of knowledge and the kind of function knowledge may have in language processing (background, construction, established mutual understanding). It also enables to specify the corresponding rational state or behavior. Such a rational model of dialog is deeply helpful as an analytical method to identify and specify subparts of a complex notion such as common ground and to express specific (several) expressions of rational behavior or state.

Secondly, we specify the basic principles of Acceptance-based Pragmatics. Finally, we present challenges of acceptance definition as well as ongoing work to refashion the belief and acceptance distinction with a fact (declarative knowledge) versus tool (procedural knowledge) distinction. Basing the distinction on one property rather than a collection of properties enables to go beyond distinctions such as voluntary/involuntary reasoning process (Hakli, 2006). or explicit/implicit memory. We explore insights both from formal epistemology and cognitive science to specify acceptance.

Cohen, L. (1989). Belief and acceptance. *Mind*, 98(391), 367-389.

Hakli, R. (2006). Group beliefs and the distinction between belief and acceptance. *Cognitive Systems Research*, 7(2-3), 286-297.

Paglieri, F. (2006). Belief dynamics: From formal models to cognitive architectures, and back again. *PhD Thesis, University of Siena*.

Saget, S. and Guyomard, M. (2006). Goal-oriented dialog as a collaborative subordinated activity involving collective acceptance. In *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue (Brandial 2006)*, 131-138.

Saget, S., & Guyomard, M. (2007). Doit-on dire la vérité pour se comprendre? Principes d'un modèle collaboratif du dialogue basé sur la notion d'acceptation. *Annales du Lamsade, MFI'07*, 8, 239-248.

**Vagueness and competition in the understanding of spatial prepositions***Fabian Schlotterbeck, University of Tübingen*

fabian.schlotterbeck@uni-tuebingen.de

Spatial language has long fascinated cognitive scientists because it provides a window into the interplay between spatial cognition and linguistic capacity. An example of one specific challenge that has attracted attention from various sub-fields of cognitive science is to give an explanation of the vagueness observed in the use and understanding of spatial prepositions, as studied experimentally using, e.g., picture rating tasks. One way to explain this vagueness is by recourse to cognitive limitations, e.g. regarding the allocation of visual attention [1, henceforth RC] or regarding representations of spatial configurations in working memory [2]. This type of approach is effectively an application of Anderson's rational analysis [3]. Although RC and related approaches have come a long way in predicting experimental data, RC suspected themselves that their account is still incomplete. They proposed that an additional factor may be at play, producing, e.g, unexpectedly steep decreases in acceptability of the preposition 'above' as the position of the so-called located object (LO) approaches horizontal alignment with the reference object (RO). Specifically, they hypothesized that lexical competition (LexComp) between alternative prepositions, e.g. 'above' vs. 'beside', may cause such decreasing acceptability. Recent computational probabilistic pragmatic models such as the rational speech act (RSA) model [4] lend themselves to formalize the LexComp hypothesis and integrate it with RC's original approach.

I present experimental data supporting LexComp and propose such an integrated RSA model to account for the data. Participants provided truth-value judgments for sentences like 'the dot is above the square' presented together with pictures showing a small dot (=LO) near a larger square (=RO). The relative angular position of LO was manipulated (0° and 180° encode complete horizontal alignment). An additional between-subjects manipulation varied whether participants saw only the preposition 'above' (exp. 1, N = 18, 36 trials, positions: 0°-180°) or did also judge sentences containing the alternatives 'left of', 'under' and 'right of' (exp. 2, N = 21, 144 trials, 0°-360°). Based on previous studies on implicatures [5], I reasoned that explicit mention of lexical alternatives would increase their salience and thus strengthen potential competition effects. As predicted by LexComp, logit mixed effects models revealed significantly lower endorsement rates in exp. 2 vs. exp. 1 if angular position was near horizontal alignment (0°-45°:  $\chi^2(1) = 11.3$ ,  $p < .001$ ; 135°-180°:  $\chi^2(1) = 9.9$ ,  $p = .002$ ) whereas no difference was found for medium angles (45°-135°:  $\chi^2(1) = 0.9$ ,  $p = .33$ ). In the computational model, a Gaussian distribution centered at the LO's true position is used to integrate noisy working memory representations. Moreover, reduced production costs of recently encountered lexical alternatives model their increased salience. This integrated RSA model thus combines a noisy semantics with informativity-based lexical competition [cf. 6]. I argue that both components are essential to account for the data.

**References:** [1] Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *J Exp Psychol Gen*, 130(2), 273-298. [2] Lipinski, J. et al. (2012). A neurobehavioral model of flexible spatial language behaviors. *J Exp Psychol Learn Mem Cogn*, 38(6), 1490–1511. [3] Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum. [4] Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998. [5] Franke, M. (2014). Typical use of quantifiers: A probabilistic model. In Bello, P., Guarini, M., McShane, M. & Scassellati, B. (eds): *Proceedings of the 36th Annual Meeting of the Cognitive Science Society, CogSci 2014*, pp. 487-492. Cognitive Science Society. Austin, TX. [6] Degen, J. et al. (under review). When redundancy is rational: A Bayesian approach to 'overinformative' referring expressions. *arXiv:1903.08237*



### On the utility of conditional answers

Jos Tellings, Utrecht University

j.i.tellings@uu.nl

Whereas most earlier work on conditional answers looks at them only as responses to conditional questions (e.g. Isaacs & Rawlins 2008), conditional answers can also naturally be responses to regular, non-conditional questions:

(1) A: Will John come to the party?

B: If he finishes his work, he will.

Such answers are given when the answerer has incomplete knowledge about A's question. These data call for an account of the utility of conditional answers: under which conditions does a speaker prefer a conditional answer over an ignorant one ("I don't know"), and if various conditional answers are possible, which one is determined most relevant?

A second reason to study the utility of conditional answers comes from the pragmatic phenomenon of "conditional perfection", i.e. interpreting conditionals as biconditionals. Recent theories propose that conditional perfection occurs when a conditional is interpreted as an exhaustive answer (e.g. Herburger 2015). In separate work, it has been proposed that "human interests" determine whether answers get an exhaustive interpretation or not (van Rooij 2004, Cariani & Rips 2017), and game theory has been applied to model this (Benz & van Rooij 2007, i.a.).

In this paper I discuss how utility-based frameworks of pragmatics can be combined with a theory of conditionals, in order to gain insight into the utility of conditional answers.

As a first step towards a theory of the utility of conditional answers, I zoom in on two case studies. First, a conditional answer ' $p \rightarrow q$ ' is relevant in the specific case in which B knows that A knows whether  $p$  (but B doesn't know whether  $p$ ;  $\neg K_b ?p \wedge K_b K_a ?p$ ). This also triggers a situation in which B may raise  $?p$  as a subissue responding to A's question (instantiating the default pragmatic assumption of 'addressee competence'), thus showing a link between conditional answers and subquestion strategies.

Second, I consider more generally the utility of a conditional in utility-based theories like Benz & van Rooij (2007). I assume that it is the conditional dependency conveyed by a conditional answer that makes it relevant. I argue that conditional answers as in (1) are conditional speech acts, rather than conditional propositions. Therefore a theory like van Rooij's (2004), which is based on the utility of *propositions*, cannot be straightforwardly applied to conditional utterances (e.g. by taking the material conditional). Instead, a conditional answer leads to a shift from A's original decision problem ("am I in a  $q$ -world?") to a new one ("am I in a  $p$ -world?"). The utility of the conditional answer is measured by the degree to which the new decision problem is easier to resolve for A than the original one.

Benz, A., & van Rooij, R. (2007). Optimal assertions, and what they implicate. A uniform game theoretic approach. *Topoi*, 26, 63–78.

Cariani, F., & Rips, L. J. (2017). Experimenting with (Conditional) Perfection. Ms.

Farkas, D., & Bruce, K. (2010). On reacting to assertions and polar questions. *J of Semantics*, 27(1), 81–118.

Herburger, E. (2015). Conditional Perfection: the truth and the whole truth. Ms.

Isaacs, J., & Rawlins, K. (2008). Conditional Questions. *Journal of Semantics*, 25, 269–319.

van Rooij, R. (2004). Utility of Mention-Some Questions. *Research on Language & Computation*, 2, 401–416.

## Information Density and the Extraposition of German Relative Clauses

Sophia Voigtmann, Saarland University

sophia.voigtmann@uni-saarland.de

A text with intertwined sentences is regarded as hard to perceive. That might result from information density (ID, Shannon 1948). ID can be defined as the “amount of information per unit comprising the utterance” (Levy & Jaeger 2007, 1). It is measured as surprisal, calculated via  $-\log_2(\text{word}|\text{context})$ . Since surprisal is connected to perceiving difficulties (Hale 2001), the impact of frequent combinations with low surprisal-values on the working memory is lower than it is for rare combinations with higher surprisal-values (Levy & Jaeger 2007, Hale 2001). To improve text comprehension producers therefore distribute information as evenly as possible across a discourse (“Uniform Information Density Hypothesis (UID)”, Levy and Jaeger 2007).

The idea behind this paper is that ID influences the extraposition of relative clauses (RC) to increase comprehension and to keep UID stable. In German the distinction of adjacent and extraposed RC depends on their position in the “Topologisches Satzmodell” (Wöllstein 2010).

Adjacent	Ich	habe	<u>ein Kind</u> ,	<i>das</i>	<i>spielte</i> ,	gesehen.		
	I	have	<u>a child</u> ,	<i>who</i>	<i>played</i> ,	seen.		
Extraposed	Ich,	habe	<u>ein Kind</u>			gesehen,	<i>das</i>	<i>spielte</i> .
	I	have	<u>a child</u>			seen,	<i>who</i>	<i>played</i> .

I have seen a child, who played.

Extraposed RC are expected to have a higher surprisal-value than embedded RC.

My research aims to find evidence for this idea in RC taken from scientific texts from the 17th to 19th century. I built a corpus of tokenized, serialized, lemmatized and normalized articles about theology and medicine from the 17th and 19th century, manually determined the RC-variants and calculated a language model (Hale 2001) with Kneser-Ney smoothing (Chen & Goodman 1999) to compute the bigram surprisal of every word of the relevant sentences.

For the periods 1650 to 1700 1258 RC and for the period 1850 to 1900 954 RC were found. A logistic regression (Generalized Linear Mixed-Effects Model, lme4 package, R) over the summed surprisal values shows a significant result ( $|z| = 3.62$ ,  $p < 0.001$ ) which indicates a correlation between surprisal values and extraposition. So for these periods it can be said that RC are more likely to be extraposed when they have a high summed surprisal value. The influence of surprisal values also seems to be stable across time. The comparison of the analyzed language periods shows no significant change (17th century:  $|z| = -3.61$ ,  $p < 0.001$ ; 19th century:  $|z| = -3.66$ ,  $p < 0.001$ ).

Bates, D., Maechler M., Bolker B., Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67(1), pp. 1-48.

Chen, S.F. & J. Goodman. 1999. An empirical study of smoothing techniques for language modelling. In Computer Speech and Language 13, no. 4, pp. 359–394.

Hale, J. 2001. A probabilistic Early parser as a psycholinguistic model. Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics.

Levy, R. & T. F. Jaeger. 2007. Speakers optimize information density through syntactic reduction. In B.

Shannon, C. E. 1948. A Mathematical Theory of Communication. Bell System Technical Journal 27. 379-423; 623-656.

Wöllstein, A. 2010. Topologisches Satzmodell. Heidelberg : Winter.

## Modeling reference games with objects of unknown categories: a Bayesian decoder for zero-shot language generation

Sina Zarriß, Bielefeld University  
sina.zarriess@uni-bielefeld.de

Recent computational work in Language & Vision has extended neural language generation models with pragmatic reasoning components implemented along the lines of "rational speech acts" (RSA) [1,2,3]. These models use a Bayesian approach to decoding a neural (literal) generation component trained, e.g., on pairs of images and verbal descriptions, to produce pragmatically appropriate image captions. Another recent strand of work in Language & Vision investigates zero-shot learning [4]. Here, the task is to correctly label objects of novel categories that the model did not observe during training. In this work, we combine these lines of research and frame zero-shot learning as a challenge for pragmatic modeling. We explore zero-shot reference games as a new setting for generation models: in this game, a speaker needs to refer a novel-category object in an image, such that an addressee (who may or may not know the category) is able to identify the target object. Previous work on language generation has commonly looked at games where a referent of a familiar category needs to be discriminated among distractor referents of identical or similar categories, thereby focussing mostly on modeling attribute selection [5]. Our work on zero-shot reference games adds an additional dimension of uncertainty to this picture, namely a setting where the category of the target itself might not be known to the model and, hence, cannot be named with reasonable accuracy.

We hypothesize that Bayesian reasoning in the style of Rational Speech Acts can extend a neural generation model trained to refer to objects of known categories, towards zero-shot learning. We implement a Bayesian decoder reasoning about categorical uncertainty and show that, solely as a result of pragmatic decoding, our model produces fewer misleading object names when being uncertain about the category. Furthermore, we show that this strategy often improves reference resolution accuracies of an automatic listener. More generally, we argue that uncertain knowledge of the world that surrounds us, including novel objects, is not only a machine learning challenge: it is a common aspect of human communication, as speakers rarely have perfect representations of their environment. Thus, when referring to objects of unfamiliar or difficult-to-name categories, even human speakers might produce utterances that avoid naming the object (e.g., the blue thingy) and avoid confusing the listener. We believe that our approach is an encouraging result for scaling models in computational pragmatics to real-world conversation and its complexities.

[1] Andreas, J. & Klein, D. 2016. Reasoning about pragmatics with neural listeners and speakers. In Proceedings of EMNLP 2016, pages 1173–1182.

[2] Cohn-Gordon, R., Goodman, N., & Potts, C. 2018. Pragmatically informative image captioning with character-level inference. In Proceedings of NAACL 2018, p. 439–443.

[3] Frank, M. C., & Goodman, N. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

[4] Socher, R., Ganjoo, M., Manning, C., & Ng, A. 2013. Zero-shot learning through cross-modal transfer. In Advances in neural information processing systems, pages 935–943.

[5] Krahmer, E., & Van Deemter, K. (2012). Computational generation of referring expressions: A survey. *Computational Linguistics*, 38(1), 173-218.

## Getting around

The conference takes place in building E 1.5 on the main campus of Saarland University. See the back cover of this booklet for a campus map. The university campus is located about 5 kilometers outside the city center. The bus ride from the center (Rathaus or Johanneskirche stops) takes about 12 minutes. The closest bus stop to the venue is “Universität Mensa”. Note that only a reduced number of buses operates on Saturdays. For up-to-date information on bus routes, we recommend checking either the website <https://saarfahrplan.de/> or the free Saarfahrplan app.

### Bus service from the city center to campus (bus stop: Universität Mensa)

Line	Destination	Service days
101	Dudweiler Dudoplatz	Thu, Fri, Sat
102	Dudweiler Dudoplatz	Thu, Fri, Sat
109	Universität Busterminal	Thu, Fri
111	Universität Busterminal	Thu, Fri

### Bus service from campus to the city center (bus stop: Johanneskirche)

Line	Destination	Service days
101	Füllengarten Siedlung	Thu, Fri, Sat
102	Altenkessel Talstraße	Thu, Fri, Sat
109	Goldene Bremm	Thu, Fri
111	Rabbiner-Rülf-Platz	Thu, Fri

### Bus service from campus to the train station (bus stop: Hauptbahnhof)

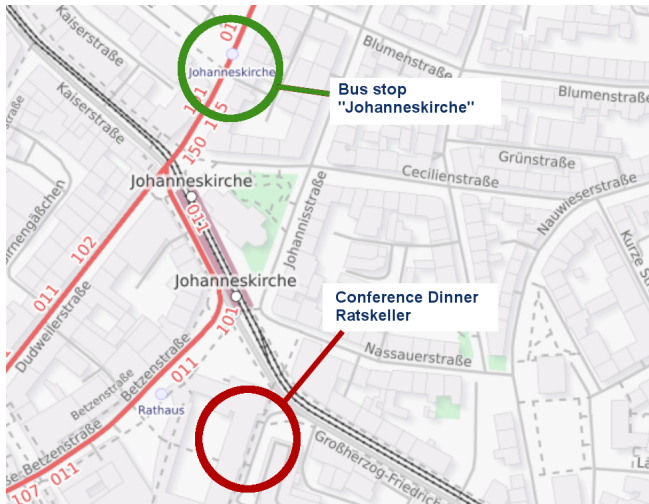
Line	Destination	Service days
102	Altenkessel Talstraße	Thu, Fri, Sat
112	Hauptbahnhof	Thu, Fri
124	Betriebshof	Thu, Fri

The University campus is serviced by taxis and buses alike. Should you need a taxi, you can contact one of the following companies:

Taxi Schneider	+49 (0) 681 71111
Taxi-Zentrale e.G.	+49 (0) 681 55000
Taxi Saarbrücken e.G.	+49 (0) 681 33033

## Conference dinner

The conference dinner will take place at the restaurant “Ratskeller”, which is located in the basement of the Saarbrücken town hall. See page 41 and the Saarfahrplan app for bus connections from campus.



## Internet and WiFi

Guests will be able to access the Internet through the wireless network. There are two ways to connect to the network:

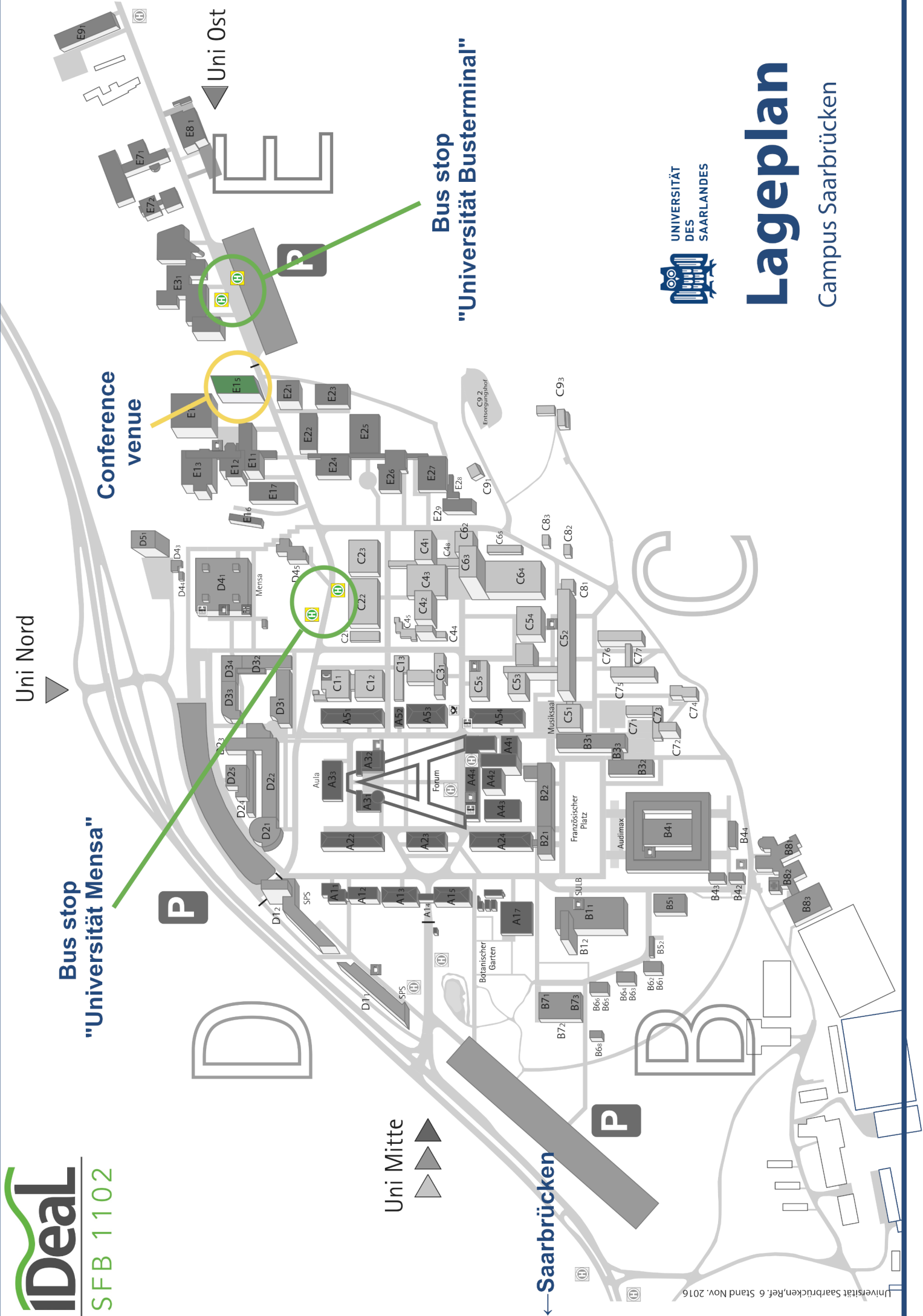
- **Academics:** Guests from academic institutions can use the Eduroam network with their institution's credentials. No extra configuration is required.
- **Industry and other guests:** We have provided individual guests accounts for the university Wifi HIZ-GUEST for the entire duration of conference. Please contact the registration desk to receive your ID and Password for Internet access. By connecting to a wireless network of the University, you agree to the Terms of Use of the Hochschul-IT-Zentrum (HIZ) of Saarland University<sup>1</sup>, along with the terms of the National Telecommunications Act.

<sup>1</sup><http://hiz-saarland.de>









# Lageplan

Campus Saarbrücken