The 4th SFB 1102 PhD Day

Location: Graduate Center (Campus, C9.3)

Date: December 8th 2017

09:45 - 10:15	**Welcome + Coffee**
10:15 - 10:45	**Examples and Specifications that prove a point: Interpreting multi-functional discourse relations** *Merel Scholman* (Project B2)
10:45 - 11:15	**Machine Comprehension Using Commonsense Knowledge** *Simon Ostermann* (Project A3)
11:15 - 11:45	**Entropy Reduction in the Visual World: Gaze-Following and Cognitive Effort** *Mirjana Sekicki* (Project A5)
11:45 - 12:45	**Lunch**
12:45 - 13:15	**Information Theory and the Usage of Fragments** *Robin Lemke* (Project B3)
13:15 - 13:45	**Integrating Convergence Capabilities into Spoken Dialogue Systems** *Eran Raveh* (associated)
13:45 - 14:00	**Poster Slam**
14:00 - 15:00	**Poster Session + Coffee** **Does referent predictability affect rate of pronominalization?** *Ekaterina Kravtchenko* (Project A3) **The Effect of Multimodal Predictability on the N400** *Christine Ankener* (Project A5) **Data Extension for Implicit Discourse Relation Classification** *Wei Shi* (Project B2) **Paraphrase Database for the Relation Types Marriage and Company Acquisition** *Seong-Eun Cho* (Project B5) **Shadowing Synthesized Speech-Segmental Analysis of Phonetic Convergence** *Iona Gessinger* (associated)
15:00 - 16:00	**SFB Meeting**

Followed by the SFB Christmas Party

Location: Graduate Centre (C9.3, Jägerheim)
Time: 16:30 – open end
More information tba!

IDeaL SFB 1102

UNIVERSITÄT DES SAARLANDES

DFG

Examples and Specifications that prove a point: Interpreting multi-functional discourse relations

Merel Scholman (B2)

Examples and specifications occur frequently in text, but not much is known about how they function in discourse. Looking at how they?re annotated in existing discourse corpora, we find that annotators often disagree on these types of relations; specifically, there is disagreement about whether these relations are elaborative (additive) or argumentative (pragmatic causal).
We conducted a crowdsourced study to investigate how readers interpret these relations. We asked English speakers (n=111) to insert connectives from a predefined list into coherence relations. The experimental items consisted of 234 relations from the Penn Discourse Treebank (PDTB, Prasad et al., 2008). In order to obtain discourse relation classification information from untrained participants via crowdsourcing, we compiled a list of connectives that unambiguously mark our target relations, drawing on a classification by Knott and Dale (1994).
The results show that they can indeed have two simultaneous functions: they can be used to illustrate/specify a situation and to serve as an argument for a claim. The results also showed large individual differences between the participants: some participants were more likely to interpret examples and specifications as elaborative, whereas others were more likely to focus on the argumentative function.
In a follow-up experiment, we asked a new group of participants to insert connectives in relations for which both interpretations were possible. The aim was to investigate whether readers have a bias for interpreting relations as elaborative or argumentative. Every participant completed four separate batches. Their insertions into items in different batches were used as a predictor for items in other batches. The results showed that readers are indeed biased towards one reading over another, and that this bias is systematic. I will discuss implications of the results of both experiments, as well as future directions.

Machine Comprehension Using Commonsense Knowledge

Simon Ostermann (A3)

Script knowledge is defined as commonsense knowledge about the events and participants involved in everyday activities, such as going to the movies or working in the garden. Project A3 is concerned with modelling script knowledge automatically, which is important both for models of information density and surprisal, and for automated text understanding. In this talk, I will concentrate on the latter part, the application of script knowledge for text understanding.
In my talk, I will present recent research on a new application area for script knowledge in natural language processing: Machine comprehension, where systems are given a text and answer questions on the text. In the first part of my talk, I will introduce MCScript, a novel dataset of narrative texts and questions about these texts that tries to make machine comprehension more accessible for models of script knowledge. The dataset complements similar machine comprehension corpora in that it contains stories about everyday activities. Also, our mode of data collection via crowdsourcing results in a substantial amount of inference questions that require commonsense, or more specifically, script knowledge, to be answered. These questions provide interesting test cases both for machine comprehension models and models of script knowledge.
In the second part of the talk, I will present a number of benchmark machine comprehension systems that have been applied to the data set, including simple classification models and more sophisticated neural models. I?ll also talk about ways of how to incorporate script knowledge. One of our preliminary findings is that while a part of the questions can very easily be answered, a major amount of questions is hard to answer.

Entropy Reduction in the Visual World: Gaze-Following and Cognitive Effort

Mirjana Sekicki (A5)

Previous research showed that a gaze cue to the target object prior to its reference induces a shift in visual attention and aids performance on subsequent tasks (e.g. Hanna & Brennan, 2007). The opportunity to employ a pupillary measure (the Index of Cognitive Activity; Marshall, 2000) while simultaneously tracking unconstrained eye-movements inspired our attempt to quantify this gaze effect online, as it appears. We hypothesized a distribution of cognitive effort during a sentence, where the reduced effort on the referent noun would be preceded by its increase on the gaze cue. In a series of experiments, we found that the cue led to a reduction of effort on the reference, even when the target did not fit the previous context. However, noting and following a gaze cue to a plausible target object proved to be effortless even when another object was preferred. Immediate cognitive effort on the cue was induced only when the target object did not fit the linguistic context.

The current study manipulates the information contributed by the gaze cue in a quantitative manner, by varying gaze specificity. From eleven plausible targets in the scene, gaze cued: a) one object; b) a group of three; or c) a group of five objects, thus reducing the scene entropy a) abruptly (11 to 1); b) significantly (11 to 3); or c) moderately (11 to 5). The results show no immediate effect on the gaze cue, but a graded modulation of cognitive effort on the reference. The same referent noun induced the least effort in *GazeToOne* condition; more in the *GazeToThree*, and the most in *GazeToFive condition*.

Congruent with our previous results, the present study finds no evidence of the distribution of cognitive effort between the (fitting) gaze cue and the referent noun. The talk will address potential interpretations of the findings, as well as further outlooks.

Information Theory and the Usage of Fragments

Robin Lemke (B3)

Nonsentential utterances, or fragments (Morgan 1973) (1-a), are frequently used instead of their fully sentential counterparts (1-b).

(1) a. [Passenger to bus driver:] To the central station.
b. [Passenger to bus driver:] I'd like to buy a ticket to the central station.

Previous research has focused mostly on the syntax of fragments (Barton 1990, Merchant 2004, Reich 2007, a.o.) and much less on the question of why we use fragments (but see Stainton 1994, Bergen & Goodman 2015). However, most, if not all, of the cited authors probably agree that the usage of fragments is possible when it is relatively easy to figure out what message the speaker meant to communicate.

The idea that those parts of the utterance which are predictable are left out suggests to address this from an information-theoretic perspective, under which fragments are preferred when they are more well-formed with respect to Uniform Information Density (Levy & Jaeger 2007), because the omission of uninformative words yields a more uniform distribution of information across the utterance.

I first present a rating study comparing the acceptability of fragments and sentences referring to predictable and unpredictable events in 2 × 2 (Predictability × Sententiality) design (2) (7-point scale, German). Predictability estimates are based on event language models trained on DeScript (Wanzare et al. 2016).

(2) Today Annika and Jenny want to cook a large serving of 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:

a.	Pour the pasta into the water, please!	Predictable
b.	Set the table, please!	Unpredictable
c.	The pasta, please!	Predictable
d.	The table, please!	Unpredictable

The study shows (i) a significant interaction between SENTENTIALITY and PREDICTABILITY: fragments are significantly better when they refer to predictable events and (ii) that fragments, but not sentences, improve significantly the more familiar a subject is with the respective script. Taken together, this suggests that fragment usage is possible specifically when the event they refer to is highly salient. The rating study is complemented by a production study, which is currently being conducted and will allow to estimate more precisely the likelihood of individual utterances in a specific situation in a script.

Integrating Convergence Capabilities into Spoken Dialogue Systems

Eran Raveh (Associated)

Nowadays, nearly every person has a device that supports some kind of spoken interaction, and a growing number of companies integrate speech-based interfaces into their products and websites. This requires human users to adapt and learn how to use these interfaces. However, while the content of speech-supporting services is fundamentally personalized (as in the case of intelligent virtual assistants (IVAs)), the way they speak back to the user is always the same - no matter how the user speaks to it, what dialect she uses, how easily she understands the device's speech, etc. Our current research deals with phonetic convergence in human-computer interaction (HCI) (e.g., Pardo, 2006; Levitan et al., 2016). More specifically, how spoken dialogue systems (SDSs) can adapt their phonetic behavior to the user's along a multi-turn interaction - both on the segmental and the supra-segmental level. This includes examining, quantifying, and modeling convergence on the Phonetic level in HCI (e.g., virtual agent, chatbots, IVAs, etc.). Ultimately, such capabilities may make the interaction easier, more natural, and more personal to the user. In this talk, a shadowing experiment carried out for examining and better understanding whether (and how) humans converge to natural and synthetic stimuli on the phonetic level will be presented (Gessinger et al., 2017). Then, the process of computationally modeling convergence will be introduced (Raveh et al., 2017). Finally, it will be demonstrated how such a model can be used in a module for SDSs to integrate convergence capabilities into them (Raveh and Steiner, 2017).

References

Gessinger, I., Raveh, E., Le Maguer, S., Möbius, B., and Steiner, I. (2017). Shadowing synthesized speech-segmental analysis of phonetic convergence. In Interspeech, pages 3797-3801, Stockholm, Sweden.

Levitan, R., Benus, S., Gálvez, R. H., Gravano, A., Savoretti, F., Trnka, M., Weise, A., and Hirschberg, J. (2016). Implementing acoustic-prosodic entrainment in a conversational avatar. In Interspeech, pages 1166-1170, San Francisco, CA, USA.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America, 119(4): 2382-2393.

Raveh, E. and Steiner, I. (2017). Phonetic adaptation module for spoken dialogue systems. In Workshop on the Semantics and Pragmatics of Dialogue (SemDial), pages 170-171, Saarbrücken, Germany.

Raveh, E., Steiner, I., and Möbius, B. (2017). A computational model for phonetically responsive spoken dialogue systems. In Interspeech, pages 884-888, Stockholm, Sweden.

Does Referent Predictability affect Rate of Pronominalization?

Ekaterina Kravtchenko (A3)

The uniform information density (UID) hypothesis suggests that speakers tend to convey information at a uniform rate, exploiting variability in language to avoid peaks and troughs in information density, across different levels of language production. One area where effects of predictability have been controversial, however, is pronominalization: e.g., "*John and Mary are friends. {John/He}...*". According to UID, pronouns, being very short linguistic units, should be used for more predictable referents, while names or complex NPs (descriptions) should be used for less predictable referents. Existing experimental evidence is mixed: there is evidence that shorter referring expressions are used in contexts where they describe referents that are, on average, more predictable (e.g. Arnold, 2001). However, there is evidence that semantic or discourse biases towards 'what will be mentioned next' do not affect pronominalization, contrary to predictions made by expectation-based accounts (Stevenson et al., 1994; Fukumura & van Gompel, 2010; Rohde & Kehler, 2014). We present two new experiments which point to limited influence of predictability on pronominalization in English.

Rohde & Kehler (2014) manipulated RE predictability in simple sentence sequences such as "*Peter admired / impressed Mary. __*", by choosing strongly subject-biased verbs (e.g. *impress*) vs. object-biased verbs (e.g. *admire*), and observing whether participants would refer back to the discourse participants with a pronoun or a name. They found that RE choice was not affected by the verb manipulation (i.e., referent predictability), although likelihood of referent mention was. Our first study attempts to replicate Rohde & Kehler (2014), but looks at a broader range of referring expressions. There is little difference in length between names and pronouns – meaning that participants, according to UID, may have little motivation to use the 'shorter' expressions in more predictable contexts. In addition to a replication, we therefore ran another similar experiment using long NPs (e.g. *The old woman in the park admired / impressed the young boy by the fountain. __*). Our results are a perfect replication of Rohde & Kehler's original results for the name condition. In the long NP condition, we find that the overall rate of pronominalization is higher than for names, but there is still no significant effect of predictability.

However, Arnold (2001) and Rosa & Arnold (2017) argue that this lack of effect is attributable to the choice of verb manipulation. Under this account, implicit causality verbs such as those used by Rohde & Kehler (2014) don't introduce sufficiently strong or reliable continuation biases (i.e., don't manipulate referent predictability strongly or reliably enough). In contrast, an effect of referent predictability on referring expression choice can be seen when one uses transfer of possession verbs: "*Peter gave / took the ball {to/from} Mary. __*", which bias towards either *goal* or *source* continuations. We therefore repeated the same experiments as above, again incorporating long NPs, but using transfer of possession constructions. To note, we used a greater number and variety of transfer of possession verbs than in previous studies. Our results do indeed show a stronger effect of the verb manipulation on referent predictability, than in the case of implicit causality verbs, and again show a higher rate of pronominalization for long NPs. However, they still show no effect of referent predictability on likelihood of pronominalization.

These experiments provide evidence that UID may not affect referring expression choice, at least in English, and more broadly that it may extend to relatively limited phenomena beyond the level of surface form predictability.

The Effect of Multimodal Predictability on the N400

Christine Ankener (A5)

A word's predictability, derived from its cloze probability, affects the N400-component on the target word (e.g. Kutas, 2011). Multimodal information, however, can also have a significant effect on Surprisal and processing effort, as reflected in the N4. By combining stable linguistic stimuli with varying visual contexts, we show that Surprisal and predictability of a target word is also sensitive to information derived from mapping multimodal information, which is reflected in a modulation of the N4.
368 visual displays were paired with 92 different plausible German sentences of the type "The man spills on saturday the water in the kitchenänd presented in RSVP-style. While the sentences were the same in each condition, the independent variable was the number of depicted objects that matched the verb constraint. Each display contained 4 clip-arts, out of which either 0, 1, 3 or all 4 were competitors (spill-able objects) for the upcoming target noun. Displays were presented with a preview time and throughout the entire sentence. If objects not matching the verb are actively excluded already at the verb, a difference in informativity and, hence, Surprisal was expected on the respective word. As a result of such an exclusion, the actual target nouns could be more or less expected, resulting in detailed differences in the N4 of the noun. A mapping of information without active exclusion of competitors was expected to result in no differences in the N4. As a result, a little to no differences in expectancy and Surprisal were expected on the noun. Results revealed an increased N4 on both relevant time windows for the mismatch condition 0, where no object matched the verb, i.e. the verb and the noun were least predictable, as well as a graded modulation in the N4 amplitude of the target noun ("water"). A modulation, that is, a significant facilitation in noun processing in condition 1 and an increased N4 for the mismatch 0, while Conditions 3 and 4 did not differ significantly from each other.
We show that the N4 is indeed sensitive to multimodal predictability andSurprisal as derived from the mapping of two modalities. We further propose that comprehenders neither actively reduce target options on the verb, nor simply map the multimodal information without having further expectations about the noun. Mapping modalities rather results in a less fine grained expectation as to whether one, many, or none of the displayed objects will be the target object.

Data Extension for Implicit Discourse Relation Classification

Wei Shi (B2)

Implicit discourse relation recognition is an extremely challenging task due to the lack of indicative connectives. It has received increased attention in recent years and various methods have been proposed.
We noticed that most existing models are trained on sections 2-21 of the PDTB and test on section 23, which includes a total of 761 implicit discourse relations. We did experiments on standard test set and cross validation among the whole corpus as well, results showed that the standard test section of the PDTB is too small and made it risky to draw conclusions about whether a feature is generally useful or not.
Neural network models become very popular on this task in nowadays, but most of them suffer from the shortage of labeled data. We addressed this problem by procuring additional training data from parallel corpora: When humans translate a text, they sometimes add (a process known as *explicitation*) connectives. We automatically back-translate it into an English connective and use it to infer a label with high confidence. We showed that a training data with several times larger than the original one can be generated in this way and gave us better performance with a simple bidirectional Long Short-Term Memory Network, even outperformed the current state-of-the-art.

Paraphrase Database for the Relation Types Marriage and Company Acquisition

Seong-Eun Cho (B5)

Relation extraction detects instances of relations between two or more arguments such as facts and events in unstructured text. These relations can be described with a wide range of encoding variants that differ in complexity and distance. In some densely-encoded variants we may observe multiple relation arguments confined within a single noun phrase (e.g., "[Microsoft] 's [2008] acquisition of [Farecast]..."), while in others the mention may span across the entire sentence ("[AOL] has gobbled up numerous online advertising companies in [2007] including : [Quigo] ."). According to the Uniform Information Density (UID) hypothesis, we would expect the spread of relation arguments of a mention to be correlated with the information density of the context in which it is expressed. More specifically, in an information-dense document we would expect the relation to be densely encoded as well (relation arguments are closer). In a document with lower average information content, the arguments are expected to be spread further apart.
We present the most recent version of our paraphrase database which will be used for testing this assumption. It contains relation mentions for the company acquisition (1.000 total) and the marriage (1.000 total) domains which were retrieved from the web utilizing relation extraction techniques and a statistical sentence filtering method called "factorial score". The database comes with the original web documents and various statistics, e.g. average information content of the document measured via the Google Web 1T n-gram database, relative position of the relation mention within the document, length of the document and average length of sentences.

Shadowing Synthesized Speech-Segmental Analysis of Phonetic Convergence

Iona Gessinger (Associated)

To shed light on the question whether humans converge phonetically to synthesized speech, a shadowing experiment was conducted using three different types of stimuli: natural speakers, diphone synthesis, and HMM synthesis. Three segment-level phonetic features of German that are well-known to vary across native speakers were examined: realization of the word-final sequence *-ig* as [ɪç] or [ɪk], e.g. in *König*, insertion or deletion of schwa in the word-final sequence *-en*, e.g. in *reden*, and realization of the word-medial vowel *-ä-* in stressed syllables as [eː] or [ɛː], e.g. in *Gerät*.
The first feature ([ɪç] vs. [ɪk]) triggered convergence in roughly one third of the cases for all stimulus types. The second feature ([ən] vs. [n̩]) showed generally a small amount of convergence. This may be due to the nature of the feature itself as schwa is unlikely to be produced in this position. The effect of the third feature ([eː] vs. [ɛː]) was clearly observable for the natural stimuli and less pronounced in the synthetic stimuli. This is presumably a result of the partly insufficient perceptibility of this target feature in the synthetic stimuli and demonstrates the necessity of gaining fine-grained control over the synthesis output, should it be intended to implement capabilities of phonetic convergence on the segmental level in spoken dialogue systems.
We conclude that humans do converge phonetically when interacting with synthesized speech. The degree of convergence however depends on the nature of the target feature, as well as its perceptibility in the stimuli, and additionally varies between participants. These findings motivate further experimentation in a more conversational setting. As spoken dialogue systems are not phonetically responsive to the user's input yet, we will work within a Wizard-of-Oz paradigm.