Alternations and Binding in Arabic

*Al Ebrahim, Hayyan* (short talk)

E-mail: hayyanebrahim@gmail.com  
Website: -  
Institute: Institut für Linguistik/Germanistik  
Project: SFB 732, B8 “Alternations and binding”

Alternations depend mainly on the meaning of a sentence which is a compositional construct in which the verb plays a crucial role. Different studies have been performed on different languages to prove the theoretical premises underlying the verb class approach (Levin 1993) and its cross-linguistic validity. They assert that verbs of a language can be grouped into semantically coherent classes according to shared linguistic properties such as the set of possible arguments structures. Many of these classes can be identified cross-linguistically. Work on German, Bengali, Korean, Chinese, French and recently Arabic has transferred a limited number of Levin classes to their respective languages. The super-classes in Levin’s collection are an especially good candidate for Arabic because they provide a high degree of generalization that allows them to be identified in many languages.

Scope-taking strategies in German Sign Language

*Bross, Fabian* (short talk)

E-mail: fabianbross@googlemail.com  
Website: http://fabianbross.de/  
Institute: Institut für Linguistik/Germanistik  
Project: -

This talk investigates the manual and non-manual encoding of modal, evaluative and speech-act operators in DGS. The central hypotheses are that the scope-taking indicated by concatenation for lower operator switches to nonmanual suprasegmental facial operator expressions for higher operators, thereby establishing the vertical axis as a scopally relevant dimension: high operators are encoded high, and low operators are encoded low. Applying a cartographic approach to clausal syntax (Cinque 1999), epistemic, permission/obligation and root modals are discussed, combined with high evaluative categories and speech-act categories. The findings have repercussions for the comparison of DGS and spoken German on the one hand, and for a general theory of scope-taking in sign languages, on the other. It is assumed that all sign languages follow this simple rule: The higher an operator is, the higher its expression on the body.

Modeling Subcategorization in Statistical Machine Translation

*Di Marco, Marion* (short talk)

E-mail: dimarco@ims.uni-stuttgart.de  
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/wellermn/index.en.html  
Institute: Institut für Maschinelle Sprachverarbeitung  
Project: SemRel “Distributional Approaches to Semantic Relatedness”
SMT output is often incomprehensible because it confuses complement types (noun phrases, NPs vs. prepositional phrases, PPs) by generating a wrong grammatical case, by choosing an incorrect preposition, or by arranging the complements in a meaningless way. Such problems are often due to differences on source and target side in terms of structure or representation of semantic information. This talk explores methods to improve the symmetry between source and target language using an abstract representation. Based on this representation, we generate context-aware synthetic phrase-table entries conditioned on the source side, to model complement types in terms of grammatical case and preposition choice. Our analysis shows that the generated phrases are used by the system and can lead to improvements during translation.

Lexical Encoding of L2 tonal contrasts: The role of L1 intonation

*Galts, Tobias* (short talk)

E-mail: tgalts@gmail.com  
Website: -  
Institute: Institut für Linguistik/Anglistik  
Project: SFB 732, A7 “Cross-linguistic interactions in second language prosody”

This study investigates the influence of the intonational system in the native language (L1) on the lexical encoding capabilities of lexical tones in a second language (L2). In a previous study, it was found that German and Russian native speakers were more accurate in distinguishing tonal minimal pairs than native speakers of French and Japanese. For this study, the previously attested difference between Russian and German native speakers in distinguishing L2 tonal contrasts will be investigated. As both L1s are comparable in terms of lexical stress, the question arises if worse results by Russian native speakers can be attributed to differences in the intonational systems. For this purpose, a cross-modal perception study design will be used. Native speakers of both L1s will be asked if a given auditory stimulus, which was learned in a previous training phase, matches the visual stimulus in three different conditions: (i) Segmental Match - Tonal Mismatch, (ii) Tonal Match - Segmental Mismatch, (iii) Complete Match. Results will have implications for the mental representation of suprasegmental knowledge as well as the acquisition of L2 phonological contrasts.

A universal modeling and query toolkit for multi-layer and multi-modal corpora

*Gärtner, Markus* (short talk)

E-mail: markus.gaertner@ims.uni-stuttgart.de  
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/gaertnms/  
Institute: Institut für Maschinelle Sprachverarbeitung  
Project: SFB 732, INF “Information Infrastructure”

The general capabilities of query systems actively being used as utility tools in NLP research can be summed up by the quote “My responses are limited – you must ask the right questions!” from the Sci-Fi movie “I, Robot”. Thus no matter the expressiveness level of the respective query language (coarsely categorized according to ISO/CD 24623-1 “Corpus Query Lingua Franca”), current systems are usually only applicable for answering a rather narrow collection of questions or operate on limited modalities and/or scales. I will present several reasons for such (design induced) constraints and show my roadmap towards a query system that builds on top of a new middleware framework for (meta)data modeling and which is able to work around most of those issues.

Towards a unified account of referential dependencies in complex sentences: Evidence from German and Jula

*Kientoré, Alassane* (short talk)
In this talk I will focus on two major phenomena of referential dependencies in complex sentences, which have long been recognized and discussed in the literature. The first one, control, has been consistently described as a referential dependence between an overt matrix argument and an embedded null subject (PRO) contained in non-finite complement clauses. The second phenomenon is the so-called “logophoricity”, which refers to an obligatory coreference relation between a special class of pronouns (logophoric pronouns) and their antecedents. Logophoric pronouns occur in embedded clauses introduced by verbs of saying, thinking and feeling. Although the general tendency in the literature consists of describing both control and logophoricity as different phenomena, I will show with evidence from German and Jula, and based on recent studies on the topic (Lee 2009, Stiebel 2010, Landau 2013, 2014, among others), that they share together properties which can justify an unified account of both.

Cross-lingual Compound Splitting for compound analysis

Müller, Stefan (short talk)
E-mail: stefan.mueller@ims.uni-stuttgart.de
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/muellesn/index.html
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, D11 “A cross-lingual approach to the analysis of compound nouns”

In SFB-project D11 I address the topic of noun compound splitting with a multilingual, corpus-based approach. I’m working on a linguistically informed compound splitting tool. In the actual state, the following languages are employed to help splitting German compounds: English, French, Spanish, Italian. I plan to include Portuguese and Greek. Furthermore, I plan to perform intrinsic (several GS formats are propagated) as well as extrinsic evaluation (within SMT-task). My current work can be found at the first level of D11’s tripartite, interdependent level approach, and is a prerequisite for the subsequent levels. In later stages of my PhD research I want to work on the two deeper levels of the model: determination of the constituents’ meaning and uncovering implicit relations, focusing on methods that involve distributional semantics. I see potential for methodological and contentual collaboration with projects also dealing with compositionality (e.g. B1, D12) or distributional information (e.g. B9, D10). Both topics are also close to the work of the SemRel-Group. Between B1/INF and D11 there’s already an ongoing collaboration established (Noun Compound Database).

Verbs in English-to-German SMT: Methods for dealing with reordering and inflection problems

Ramm, Anita (short talk)
E-mail: anita.hr@gmail.com
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/gojunaa/index.html
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, D12 “Sense Discrimination and Regular Meaning Shifts of German Particle Verbs”

In English-to-German SMT, the verbs pose a big problem. Their position differs significantly in the two languages which leads to the false placement of the verbs in the generated translations. In many cases, the English verbs are not even translated. In addition to positional problems, the rich German verbal morphology leads often to erroneously inflected verbs resulting in difficult understanding of the translations. I present a method for handling the reordering problems caused by different verb positions in English and German, as well as an approach to deal with rich German verbal morphology. Particularly the translation of tense and mood is an interesting problem which I present and discuss in detail.
Learning to Represent Meaning in Technical Documentation

Richardson, Kyle (short talk)

E-mail: kyle@ims.uni-stuttgart.de
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/kyle/
Institute: Institut für Maschinelles Sprachverarbeitung
Project: SFB 732, D2 “Combining Contextual Information Sources for Disambiguation in Parsing and Choice in Generation”

In this talk, we consider the problem of translating high-level textual descriptions to formal representations in technical documentation, e.g., programming language documentation and other how-to manuals. We approach this problem by exploiting the parallel nature of such data, i.e., textual descriptions are often paired with lower-level formal representations, such as function definitions or code templates. Translations are learned by using such parallel text-meaning pairs as the sole supervision. We report partial results on six novel datasets, including the standard library documentation for five different programming languages and a collection of Unix manuals, and for seven natural languages.

Lost in Grammaticalization – Tracking Desemantization via Entropy in Semantic Vector Spaces

Schlechtweg, Dominik (short talk)

E-mail: Snicka@gmx.de
Website: http://www.uni-stuttgart.de/ilg/institut/mitarbeiter/schlechtweg/index.html
Institute: Institut für Linguistik/Germanistik
Project: –

Change over time is a fundamental property of natural language and happens at many linguistic levels. An important process of language change is grammaticalization, i.e., “the development from lexical to grammatical forms, and from grammatical to even more grammatical forms” (Heine & Kuteva 2007: 32). In historical linguistics, grammaticalization is assumed to consist of several more specific subprocesses, one of them being desemantization (semantic bleaching). It has not been tried to model this process of language change computationally. There have been attempts to model semantic change via distributional semantics. Yet, these mostly describe the degree of change rather than qualifying different types of semantic change (Juola 2003, Gulordava & Baroni 2011). How the meaning of a word changed (given a certain degree) may, however, be very different, e.g., it may have narrowed, broadened or been lost. Few describe different types of meaning change (Sagi et. al 2009). An unsupervised computational model of different types of semantic change may contribute to historical linguistics by enabling testing of existing assumptions and making new predictions about narrowing, broadening and semantic bleaching. For this, we assume that these processes are encoded by the change of the distribution of an expression and the words it co-occurs with over time. Interestingly, there are already measures from a very different domain of research using this distribution to compute an entropy value for expressions measuring their semantic generality (Santus et al. 2014). We want to apply this method to historical linguistics by measuring changes in the semantic generality of expressions over time as consequences of narrowing, broadening or semantic bleaching.

Modular solutions for cross-disciplinary text analysis

Schulz, Sarah (short talk)

E-mail: sarschu@gmail.com
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/schulzsh/
Institute: Institut für Maschinelles Sprachverarbeitung
Project: CRETA “Center for Reflected Text Analytics”

Digital Humanities is a field of research that is developing fast. It holds its very own kind of challenges and scientific issues both interesting for Humanities and Computer Science. One of
the problems for NLP is the non-canonical nature of text especially found in projects dealing with historical text. These texts deviate from text that has been the main focus of NLP so far by lacking standardized orthography and grammar. This can often lead to dissatisfying results using tools trained on standard text for such data. In my PhD research, I focus on finding general techniques to deal with such data. I concentrate on methods for NLP tasks that re-occur throughout different projects like text normalization, PoS tagging or named-entity recognition.

Shifting Senses: From Perception to German Particle Verbs

Springorum, Sylvia (short talk)
E-mail: Sylvia.Springorum@gmx.de
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/riestesa/
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, D12 “Sense Discrimination and Regular Meaning Shifts of German Particle Verbs”

A particle verb (PV), like anstrahlen 'beam at' is a composition of the Particle (P) an and a base verb (BV) strahlen 'beam'. I assume that each P represents one or a few basic concepts, which are, in most cases, spatially grounded. In the case of an, it refers to a gravity independent condition and therefore a horizontal direction. The actual meaning contributed by the P in a PV composition is determined not only through properties which are defined through the semantics of the BV or the PV arguments, but also through perception-based characterizations of their domains and their cognitive abstractions. Examples include the number of dimensions, the influence of gravity and the concreteness, and respectively abstractness degree. This cognitive analysis of PV compositions in comparison to the pure semantic analyses proposed in former works is more expressive and can therefore also predict PV meanings in new contexts, metaphors, possible meanings of PV-neologisms and their plausibility. I will present a short summary of the theoretical ideas together with some examples from various experiments.

Prosody for Spoken Language Understanding

Stehwien, Sabrina (short talk)
E-mail: sabrina.stehwien@ims.uni-stuttgart.de
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/stehwisa/index.html
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, A8 “Investigating the Interaction between Speech and Language Processing for Spoken Language Understanding: A Case Study for Sentiment Analysis”

Spoken Language Understanding (SLU) comprises tasks that process the semantic content of speech beyond simple recognition. As it is usually implemented as a subsequent task after ASR, information that is unique to speech, such as the location of pitch accents, is lost on the way. Since prosody has a clear connection to meaning in languages like English and German, the current focus of my PhD is to integrate prosodic features in SLU and investigate their impact on the automatic understanding of speech.

Interdisciplinarity turns a speaker’s world upside down

Suditsch, Isabel (short talk)
E-mail: isabelsuditsch@gmail.com
Website: -
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, A4 “Phonetic Convergence in Spontaneous Speech”

Languages for Specific Purposes (LSP) enable speakers to communicate efficiently and explicitly within their (scientific) disciplines (vom Brocke 2013, 19 f.) since they agree concerning their scientific background, perspectives and terminologies. In an interdisciplinary context, however,
this circumstance leads to communication difficulties (Blaschke & Lukatis 1979, 70; Hoyningen-Huene 1988, 136 ff.; Balsiger et al. 1996, 128). Against the backdrop of this LSP paradox, this talk will explain the three levels of interdisciplinary ambiguity and show some evidence from empirical data. Further, some findings will be presented.

– LSP meaning is dominant in an individual’s mental lexicon
– The higher the expert status the less likely will a non-LSP meaning be present

The licensing nature of experiencer object verbs

Temme, Anne (short talk)
E-mail: anne.temme@posteo.de
Website: https://uni-stuttgart.academia.edu/AnneTemme
Institute: Institut für Linguistik/Germanistik
Project: SFB 732, B6 “Underspecification in Voice systems and the syntax-morphology interface”

Experiencer objects of verbs like annoy or worry are assumed to have a special status compared to canonical theme or patient objects. One important observation is that most accusative experiencer object (EO) verbs appear in different verbal aspects, as shown in (1). This is relevant since only non-agentive EO structures exhibit psych effects (Arad 1998).

(1) a. Nina frightened Laura deliberately / to make her go away. (agentive)
   b. The explosion / the noise / the thunderstorm frightened Laura. (causative)
   c. Nuclear war frightened Nina. (stative)

Nevertheless, the literature so far offers no theoretical clarification of the relevant factors and additionally shows a lack of robust empirical data. In my talk I will briefly present evidence for the special status of EO verbs in German and investigate the aspectual/thematic properties that distinguish them from other verb classes.

Constructing And Evaluating Syntax-Based Distributional Semantic Models

Utt, Jason (short talk)
E-mail: jasonutt@me.com
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/uttjn/
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, D10 “Incrementality in Compositional Distributional Semantics”

Distributional semantic models (DSMs) in which words are assigned vectors of their contexts in high-dimensional vector space, have proven useful as a reliable source of unsupervised representations of word meaning. They have found use across numerous applications and fields by interpreting spatial (or angular) proximity as semantic relatedness, thus allowing a geometric definition of word similarity. Such simple word-based models however lack important relational information to help distinguish or predict plausible argument role fillers. E.g. banana would be a good candidate for the object role of the verb to peel but not a good subject. Syntax-based DSMs (SDSMs) provide richer representations allowing for more fine-grained distinctions and analogies to be drawn. In the context of my thesis – where the goal is to construct and compare mono- and cross-lingually induced SDSMs – I will show and discuss different evaluation scenarios for SDSMs, in particular showing how such models can help perform tasks for which standard DSMs are unsuited.

The Utility of various Features for the Improvement of a Language-independent Compound Splitter

Ziering, Patrick (short talk)
E-mail: ziering@gmx.de
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/zierinpk/
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, D11 “A cross-lingual approach to the analysis of compound nouns”
Compounding represents one of the most productive word-formation types. There is an infinite number of compounds with very low corpus frequencies, which makes statistical methods run into data sparsity issues. Therefore, automatic compound analyzers have been developed, which break the task of interpretation down to the level of the composed lexemes. In my PhD work, I develop both multilingual and cross-lingual methods for the compound analysis. In this talk, I present a study about the utility of various features aimed at the improvement of a language-independent compound splitter. Among the inspected features, there are alternative morphological forms, cosine similarity derived from a distributional semantics model, different scoring methods and degrees of agreement between morphological rules and potential compound parts.
Progress report talks

Predicting Fine-Grained Entity Attributes from Distributional Semantics

_Gupta, Abhijeet_ (progress report)
E-mail: Abhijeet.Gupta@ims.uni-stuttgart.de
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/guptaat/index.en.html
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, D10 “Incrementality in Compositional Distributional Semantics”

The semantic representation of a named entity, for example: Germany, in a knowledge base involves associating that entity with its semantic attributes and the corresponding values for those attributes. Based on the type of values, these attributes can be broadly categorized into numerical attributes (GDP: 3.3 trillion euros) and categorical attributes (Capital: Berlin). Knowledge base creation, however, is a labour intensive process. An alternate is knowledge base completion, that starts from a partial knowledge base and extends it automatically with new attributes. The goal of my work is to develop generalized statistical models which can successfully learn and predict fine-grained attributes of entities from their distributed vectors. The hypothesis being that these vectors encode enough semantic information through which a model can be successfully made to learn and consequently predict values for these attributes. On successful fruition, these models can be applied to downstream tasks, like question-answering systems, question generation, etc. instead of other structured information sources. Our experiments show that we can successfully model attribute learning and prediction to a relatively high degree of accuracy.

Distributional Properties of Morphological Derivation

_Kisselew, Max_ (progress report)
E-mail: kisselman@ims.uni-stuttgart.de
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/kisselman/index.en.html
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, B9 “Distributional Characterization of Derivation”

The goal of my PhD project is to investigate how Distributional Semantics can contribute to our understanding of morphological derivation. On the content level of my talk I will focus on two approaches from past and ongoing work. In one approach I inspect how derived word forms in German can be predicted from their base forms using methods from compositional distributional semantics. The goal is to understand for which derivation patterns, and which lemmas, this approach works and for which it does not. Another investigation deals with conversion cases in English (e. g. tunnel (n.) → tunnel (v.)). I will show that (to some extent) it is possible to predict, which word is the base and which one the derived word by examining the properties of their distributional word vectors.

Modelling Sense Discrimination and Meaning Shifts of German Particle Verbs

_Köper, Maximilian_ (progress report)
E-mail: maximilian.koepert@gmail.com
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/koopermn/index.en.html
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, D12 “Sense Discrimination and Regular Meaning Shifts of German Particle Verbs”
The goal of my work is to model and to identify regular meaning shifts of German particle verbs (PVs) with prepositional particles, such as ab-stottern, auf-brausen, nach-weinen. The talk will present past, current and possible future work. In more detail, past work: (i) a resource of automatically generated affective norms for German, (ii) experiments on the prediction of literal vs. non-literal language usage of PVs. In addition we present ongoing work on token- and type-based approaches towards automatic word sense induction and disambiguation of German PVs.

Computational modelling of aspects of information status in interview data

Rösiger, Ina (progress report)
E-mail: roesigia@ims.uni-stuttgart.de
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/roesigia/index.en.html
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, A6 “Encoding of Information Structure in German and French”

My PhD project revolves around the computational modelling of different aspects of information status in German and English texts, mainly focusing on the task of coreference resolution, bridging resolution and information status classification. As this talk is part of my progress report, I will not focus on one specific topic, but rather give an overview of my research over the last two years as well as summarise the research questions that we want to investigate next.

The influence of rhythmic preferences on the prosodic marking of information structure: a cross-linguistic investigation

Schauffler, Nadja (progress report)
E-mail: nadjaschauffler@gmail.com
Website: http://ifla.uni-stuttgart.de/index.php?article_id=211&clang=0
Institute: Institut für Linguistik/Anglistik
Project: SFB 732, A7 “Cross-linguistic interactions in second language prosody”

Information structure and prosody interrelate in many languages. For example, in German and English, information structural categories are typically marked by pitch accent placement and pitch accent type. However, the actual choice of pitch accent type as well as the distribution of pitch accents in a phrase can vary beyond of what semantic or pragmatic factors can explain. In my dissertation I look at how rhythmic preferences contribute to this variation and how they interact with information structure in production. In particular, I investigate whether German and English speakers produce contrastive pitch accents when they violate the preference for rhythmic alternation (i.e. an alternation of strong and weak). Furthermore, I investigate how this interaction is handled in non-native speech, namely by German learners of English, and how their strategies are perceived by L1 listeners.
Dry run talks

Task-based parser output combination: workflow and infrastructure

Eckart, Kerstin (dry run)
E-mail: eckartkn@ims.uni-stuttgart.de
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/eckartkn/
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, INF “Information Infrastructure”

Generating syntactic analyses of language data is rarely a self-contained task. Thus, parser output often constitutes the input for further automatic or manual annotation steps, is utilized in downstream applications or provides a resource for data extraction. Since these subsequent tasks rely on the quality of the syntactic information they draw from the parser output, parsing errors are likely to propagate to the outcome of the subsequent task. This talk presents a task-based workflow for the combination of output from different parsers. The workflow provides a subsequent task with syntactic information which is more reliable than the information provided by each single parser. The workflow can be easily instantiated for different tasks and takes their specific requirements into account. Additionally, a classification of different combination types is introduced along with an infrastructure to support the workflow. Two case studies will show instantiations of the workflow in precision-oriented and recall-oriented settings.

P’s at the interfaces: The lexical semantics of spatial prepositions

Haselbach, Boris (dry run)
E-mail: haselbbs@ims.uni-stuttgart.de
Website: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/haselbbs/
Institute: Institut für Maschinelle Sprachverarbeitung
Project: SFB 732, B4 “Lexikalische Information und ihre Entfaltung im Kontext von Wortbildung, Satz und Diskurs”

This talk outlines a word-syntactic approach to the lexical semantics of spatial prepositions combining principles from Minimalist Syntax (Chomsky 1995), Distributed Morphology (Halle and Marantz 1993), and Discourse Representation Theory (Kamp and Reyle 1993). In particular, I take lexical-semantic features as primitives (i) that can generate sub-lexical structure according to syntactic principles and (ii) that are interpreted contextually at both interfaces, the Articulatory/Perceptual Interface (Phonological Form) and the Conceptual/Intensional Interface (Logical Form). Building on formal algebraic and geometric approaches and focusing on spatial prepositions in German, I analyze place prepositions as well as goal, source, and route path prepositions. I show that analyzing spatial prepositions sub-lexically in terms of Bare Phrase Structure with functional heads introducing building blocks of meaning does not only provide structured access to their semantic properties, but that it also allows for explaining their case assignment properties in a post-syntactic, morphological case approach (Marantz 1991, McFadden 2004).