Abstracts for the poster session

Poster 1. Automatic Identification of Paradigm Gaps in Corpora
Elmira Mustakimova
cclikespizza@gmail.com

A standard theory of word formation views inflection as the application of a certain set of rules to an input form that derives the output form. Defective words represent a major problem for such theories: there exist words to which the rules do not apply or rules do not produce any output. Any theoretical account of defectiveness in a language is based on a very practical task -- compiling a list of defective words. Usually, such lists of allegedly defective words are provided in grammars and dictionaries which, unfortunately, cannot keep pace with the language use. On the other hand, language corpora usually represent the functioning of modern language. Thus, a corpus can, in theory, show which words are defective.

Having this in mind, I postulate the following question: if we know some words that are defective in a language, how can we extract from the corpus information about other defective words? The aim of this work is to describe and implement the models for extracting a list of defective words from a corpus. In constructing and developing the models, I presuppose having a big representative corpus of a language and a program for morphological analysis. In this work, I train and test the models on the data of Russian National Corpus. I will describe the three machine learning methods that I chose for the task, present the features that are used for learning the models, and show how the accuracy depends on the size of the training sample.

Poster 2. Optimality-Theoretic Phonology as a Homeostatic Space
Jon Rawski
rawski.jon@gmail.com

Phonology learning appears seemingly different from other types of learning. A child may acquire multiple phonologic systems during a hypothesized "critical period" (CP), and then experience a large decrease in language plasticity during adolescence. The representation of simultaneous linguistic systems is not well accounted for in generative grammar, which accounts for initial and final states. However, Theoretical Neuroscience has richly developed neurally-based learning models, but cannot explain basic linguistic structure. I propose that language and language learning obey homeostatic principles and can be unified under the concept of a homeostatic space and deviations. Specifically, a unified framework for both can be achieved using Homeostatic Reinforcement Learning. I define an Optimality Theoretic Grammar as a homeostatic space, and show that it can model deviance from new phonologic constraint ranking as reduction in the space. I apply theories of homeostatic learning to a space constructed by constraint weights given by an Optimality Theoretic analysis. Preliminary findings show that defining a homeostatic space for language allows for infinite transformational fluctuation of a language state, and for multiple frameworks in a space. This new definition of language, besides accounting for a traditional
linguistic assumption, provides new prospects for cognitive theories uniting physical and abstract mechanisms rather than separating them.

Poster 3. Negative Pronouns and/or NPIs in Hittite: Corpus-Based Approach

Maria Molina

maria.molina@me.com

My presentation discusses the problem whether negative pronouns in Hittite are negative pronouns proper (like English nobody) or NPI (like English anything). I apply the corpus approach to investigate the scope of negation in sentences with negative polarity items/NPIs, the position of negative pronouns/NPIs vis-a-vis negation markers and other problems. My poster represents the ongoing project of a corpus of Hittite, one aim of which is to get quantitative data on Hittite negative pronouns — and to find more about their nature.


juliajulia95@bk.ru

The construction X Y-ed Z is used to send a message. This construction can be described as: X made a controllable action Y (a gesture or a mimic action) with a purpose to express Z (an emotion or speech acts). A physical action Y is often controllable and represents a gesture or a mimic action. Some transitive cannot be used in this construction because they require an object that expresses a part of a body that is used with Y. There are cases when Y is expressed with an action that agent cannot control (e.g. to blush). The purpose of this work is to describe compatibility and semantics of the construction X Y-ed his Z. The research shows that similar constructions are widely used with a purpose to send some message using controllable actions (a gesture or a mimic action).

Poster 5. Semantic Symmetry and Asymmetry of Russian Deictic Adverbs tut ‘here’ and tam ‘there’ (Corpus Research).

Maria Zarifyan

masha.zarifyan@gmail.com

The pair of Russian deictic words tut ‘here’ and tam ‘there’ seems to be antonymous at first glance, which is an oversimplification of the real situation. The goal of this research is to study spatial, temporal and situational meanings of tut ‘here’ and tam ‘there’ and highlight cases in which these adverbs are antonymous or not. The data shows that situational meanings develop almost complete symmetry. In other cases asymmetric contexts of use are more frequent rather than symmetric. First of all, the central meaning of tam ‘there’ is spatial, while tut ‘here’ tends to be used in situational contexts. Moreover, tam ‘there’ is less deictic. In certain spatial contexts of use it can lose the semantics of distance and develop the anaphoric function, while tut ‘here’ always refers to a speaker’s location. In addition, tut ‘here’ and tam ‘there’ develop the semantics of proximity and distance respectively in their non-spatial contexts of use. For example, both adverbs denote two successive events, but the difference is that the events linked by tut ‘here’ follow each other immediately, while tam ‘there’ separates events from each other. However, not in all contexts of use it is possible to compare both adverbs. For example, tut ‘here’ can denote an event that is close to the
moment of speech, while tam ‘there’ do not denote the distance from the moment of speech.

**Poster 6. Specialized Converbs in Nakho-Dagestani Languages: A Typological Survey**

**Maria Sheyanova**  
*shmasjanja@rambler.ru*

This work presents a typological survey of specialized converbs in East Caucasian languages. Specialized converbs are distinct from the more familiar category of general (contextual) converbs in the following way: Whereas the latter do not establish any specific semantic relation between matrix and converbal clauses and interpretation of such relations largely depends on context, the former establish specific semantic relations. These relations may be classified into temporal (for example, anterior, immediate, durative etc), logical (purposive, causal, concessive) and locative (essive, lative, elative). The inventory of semantical types varies from language to language. The main focus of the work is on comparing semantical inventories between different East Caucasian languages and reviewing of all possible semantical oppositions distinguished in the languages of the sample. It is worth mentioning that in descriptions of these languages a couple of unusual and even exotic cases were found. In addition to that, an inventory is provided of the most frequent etymological sources of specialized converb constructions. The survey is based on considering converbal systems in the following languages: Lesgian (Lezgic), Khinalug (family isolate), Akhvakh (Avar–Andic), Ingush (Nakh) and Khvarshi (Tsezic).

**Poster 7. General Converbs in Nakho-Dagestani Languages: A Typological Survey**

**Marina Kustova**  
*tanglorfefoot@gmail.com*

According to a well-known definition, a converb is a non-finite verbal category whose main function is to mark adverbial subordination [Haspelmath 1995]. Nakho-Dagestani languages tend to distinguish between general and specialized converbs. My work is an overview of the category of the general converb, which, as opposed to the special converb, does not specify the nature of the semantic relation between converbial and main clause.

I have worked with grammatical descriptions of a selection of Nakho-Dagestani languages, including Akhvakh (Avar-Andic), Khvarshi (Tsezic), Dargwa (Dargi), Ingush (Nakh) and Khinalug (separate branch), to compile a list of converb properties to compare. I considered the following features:

- the number of general converbs in the system;
- stems and markers used to form converbs;
- the use of the converbs in periphrasis;
- argument sharing strategies;
- secondary functions of converbs and converbial markers;
- co-ordination vs. subordination in converbial construction.

As a result, some generalizations about Nakho-Dagestani converbs were made:

- all languages in the sample distinguish between PFV and IPFV converbs, and some languages have additional converb forms;
• often, PFV and IPFV converbs are derived from different stems using different markers;
• IPF converbs tend to take part in formation of progressive periphrastic forms, while PFV converbs are most often used to form perfect periphrastic forms; if there are other general converbs, they never take part in forming any analytical verb forms;
• in all the languages for which enough data are available, the converbial construction showed more subordinate than co-ordinate properties;
• any argument sharing strategy is possible, though in the languages for which sufficient data are available the pattern of sharing S/A arguments dominates.

Poster 8. The Challenge of Japanese Person-Referring Terms: Pronouns or Not?
Rodanthi Christofaki
rodanthi@gmail.com

Japanese is known for its large inventory of 1st person referring terms, which are traditionally called ‘personal pronouns’ by traditional accounts such as native dictionaries of the language. Although from a syntactic point of view these items seem to resemble nouns more than pronouns (Longobardi 2008), this poster examines functional and semantic properties of pronounhood that have been put forward (Heine & Song 2011) and proposes some amendments that will serve to classify Japanese person-referring terms. This attempted classification constitutes the first step of a project that aims to assess the indexical / non-indexical distinction in language using self-reference in Japanese as its research domain.

Ekaterina Ageeva, Francis M. Tyers, Juan Antonio Pérez-Ortiz and Mikel L. Forcada
sereni.nm@gmail.com

This paper contributes observations on the viability of a strategy independently proposed in 2012 and 2013 for evaluation of machine translation (MT) for assimilation purposes. The strategy involves human evaluators, who are asked to restore keywords (to fill gaps) in reference translations. The evaluation method is applied to two language pairs, Basque–Spanish and Tatar–Russian. To reduce the amount of time required to prepare tasks and analyze results, an open-source task management system is introduced. The evaluation results show that the gap-filling task may be suitable for measuring MT quality for assimilation purposes.