THE DYNAMICS OF LEXICAL INNOVATION

Book of Abstracts

THANK YOU FOR COMING

WE HAD A LEXILICIOUS TIME!

Any new words we should monitor?

Let us know via

daphne.kerremans@anglistik.uni-muenchen.de

Gefördert von der **DFG**

Foreword

The fate of lexical innovations is not predetermined. Some nonce-formations, coined at the spur of the moment, remain ephemeral and are never used again. Others manage to spread and gain a certain degree of currency, but never make it into the common vocabulary shared by most speakers and disappear from active use after a brief period of time. Yet others diffuse rapidly and become permanent additions to the lexicon. The dynamics of these diffusion processes, and the factors influencing the extent to which new words become conventionalized (or not) have not been investigated systematically so far.

The workshop 'The Dynamics of Lexical Innovation: Data, Methods, Models' plans to provide a forum for research that investigates how lexical innovations diffuse in speech communities, domains of usage, media and text types and tries to determine which factors play a role in this process. The workshop focuses on challenges involved in data acquisition, on methodological and technical aspects, and on theoretical and computational models of lexical innovation and diffusion processes. More concretely, the organizers welcome papers on the following issues and questions, among others:

- How can neologisms be detected as close to their moment of coinage as possible?
- How can their spread be monitored systematically?
- What are suitable dependent variables for measuring the diffusion of lexical innovations and how can they be operationalized?
- Which sources are particularly useful for detecting neologisms and how can data retrieval procedures be implemented?
- What are the factors that co-determine the time course and pathways of the conventionalization of lexical innovations?
- Which types of models provide a systematic account of the process of lexical innovation and adequately predict the effects of the factors influencing this process?

Contents		Programr	ne
Programme	3	Wednesday, 2 Leopoldstr. 13, r	
Abstracts	6	16.00	Welcome
Talks	7		Approaching lexical innovation in earlier English Terttu Nevalainen
Posters		18.00	Welcome reception
Orientation	35	S	Shakespeare library, Schellingstr 3, room 207
Useful information	36		

unfollow, v.: to stop following someone on Twitter

Programme

Programme

Thursday, 29 June Seidlvilla, Nikolaiplatz 1B

9.00	The Logoscope: A semi-automatic tool for detecting and documenting the context of French new words <i>Christophe Gérard</i>
09.45	Detection and lifecycles-tracking of neologisms: Experiments in the Neoveille project Emmanuel Cartier
10.30	The NeoCrawler: When <i>frugalista</i> meets <i>bankster</i> Jelena Prokic
11.30	Coffee break
11.45	Geographical patterns of lexical innovation Jack Grieve
12.30	Exploring markers of lexical conventionalization in live monitor cor- pora <i>Piotr Pęzik</i>
13.15	Lunch break
14.30	Negativity and lexical innovation William Hamilton
15.15	Metacomments and metasignals – what can they tell us and how do they affect conventionalization? <i>Jan Svanlund</i>
16.00	Posters & coffee
17.00	Modelling the propagation of a linguistic innovation (until 18.00) <i>Richard Blythe</i>
19.00	Conference dinner (Max-Emanuel Brauerei)

Friday 30 June *Seidlvilla*

9.00	Lexical Blending in an integrated usage-based theory of word formation <i>Suzanne Kemmer</i>
09.45	Social embedding of neologisms in early English correspondence Tanja Säily und Eetu Mäkelä
10.30	Coffee break
11.15	Closing discussion (until 12 pm)

yolo,expr., aphorism: you only live once, enjoy life as if there is no next day.

Richard Blythe Edinburgh University 7

Modelling the propagation of a linguistic innovation

In this talk I will summarise what has been learnt from the study of a mathematical model for language change known as the Utterance Selection Model. This model is grounded in Croft's evolutionary framework for language change and has two main virtues. First, it can incorporate a variety of mechanisms, including various types of accommodation, Bloomfield's principle of density, social prestige effects, trend amplification and so on. Second, in many cases precise predictions for how language is expected to change under the influence of specific mechanisms can be made. More precisely, we consider the question of how an innovation (e.g., a new word for an object) competes with a convention, with particular focus on those cases where the innovation propagates to become the new convention.

As part of this work, we construct a four-way typology of language change mechanisms. The simplest mechanism (neutral evolution) appeals only to speakers passively matching the frequencies of the linguistic variants they hear from their interlocutors. More complex mechanisms appeal to variation in the rate at which speakers interact with different interlocutors, asymmetry in the weighting of different speakers in these interactions and systematic overproduction or underproduction of variants according to the identities of their users. We will argue that this latter mechanism (in which variants acquire a 'fitness') is an essential to reproduce the characteristic 'S-shaped' adoption curve of an incoming linguistic convention. Moreover, we will show that these weightings can emerge spontaneously from interactions between speakers, and lead to the formation of dialects in heterogeneous speech communities.

A number of questions remain unanswered. In particular, methods for distinguishing in detail between different mechanisms for language change with reference to available empirical data are in their infancy. I hope in this workshop to identify possibilities to bring models and theories of the type described in this talk closer to empirical data.

Abstracts Talks

Emmanuel Cartier Université Paris 13 Sorbonne

Detection and lifecycles-tracking of Neologisms : experiments in the Neoveille Project

This proposition is part of a project whose goal is to setup a web platform designed to automatically detect neologisms, describe them linguistically and track their lifecyles in seven languages through newspapers monitor corpora. The platform enables linguists to create and manage their corpora, accept or reject automatically detected neologisms, describe linguistically the validated neologisms and follow their lifecycle on a monitor corpora via visualization techniques. The focus of ourcommunication will be the parameters we have used to detect and track neology.

Several linguistic models have been proposed to explain the neologism lifecycle. Gevaudan and Koch (2010) state that lexical change is composed of three parameters: a semantic parameter expliciting a continuity or discontinuity of meaning; a stratic parameter linking the linguistic structure to its sociological context; a formal one, with four generic matrices: conversion, morphological extension, composition and clipping. Construction Grammars propose an explanatory model that stems from the notion of construction, characterized by three parameters : size (from atomic to complex), phonological specificity (from substantive to schematic) and type of concepts (from contentful to procedural) (Traugott and Trousdale 2013). The form-meaning pairs change through time is then linked to entrenchment (Langacker 1987), caused by several factors (Schmid, 2015) that affect their schematicity, productivity and compositionality. Additionally, (Traugott and Trousdale 2013) have explicited three main stages in the life of neology: emergence, dissemination and conventionalization, linking each of these to more-or-less operationalizable factors.

Computational linguistics have proposed several approaches, apart from the Exclusion Dictionary Method (EDM) (Renouf 1993; Cabré 1995; Gérard 2014; Kerremans et al. 2012; Cartier 2016), that is limited to formal neologisms : The first approach is linked to the distributional hypothesis (Harris 1970) that a significant part of word meaning is contextual. We can retrieve from a large corpus all the collocates and collostructions of a lexical unit, and classify them according to several metrics. The salient resulting context words represent the "pro-file" (Blumenthal 2009) or "sketch" (Kilgarriff 2014) of a lexical unit. Change tracking consists in comparing the word profile through time. It has been implemented in one system (Renouf 1993). The second approach is also inspired by the distributional hypothesis that lexical units sharing similar or identical contexts are most likely to be semantically close (Harris, 1954). It has been implemented in several systems (Turney 2010; Mikolov 2013) with amazing results, as synonyms, and hyperonyms are among the most similar words. In the context of semantic change Hamilton (2016) has setup the first system of this kind, based on the Google Ngrams data.

In the Neoveille Project we have tested a combination of these parameters to track semantic change: relative frequency, form, both at the morphological, syntactical, lexico-syntactical and even semantic distribution levels, diastratic and diatopic parameters. The diastratic and diatopic parameters are operationalized by adding metadata to every web source of information. The linguistic change is operationalized at the morphological level (i.e. formal neologisms) by applying the EDM. As for the lexico-syntactic level, we propose an improvement of the sketch engine system by applying a semantic distribution calculation over collocates occurrences, so as to gather collocates and/or collostructions. At this level, linguistic change is either a change in construction at the syntactical level or a change of semantic class for arguments, or both.

Our experiment is based on a large-scale morpho-syntactically annotated French newspaper corpus in three contemporary periods (1987-1988, 2005-2006 and 2010-2011). After a Frequent Patterns calculation we apply several postprocesses enabling to approach a linguistically relevant combinatorial profile for lexical units. We then complement it by a vector space model calculation (Mikolov et al. 2013), giving rise to the main synonyms for verbs. We then compare these combinatorial profiles, distributions and relative frequency between periods to track meaning change. We will present several examples of meaning change. We will also detail several constraints and preprocesses necessary to derive trackable profiles and distributions. The results and evolution of these parameters can be visualized on the Neoveille website. We will end-up with some limitations of the current system.

Christophe Gérard Université de Strasbourg

References

- Blumenthal, Peter (2009). "Éléments d'une théorie de la combinatoire des noms". Cahiers de lexicologie 94: 1.11-29.
- Cabré Maria Teresa. and Luis de Yzaguirre L. (1995). "Stratégie pour la détection semi-automatique des néologismes de presse". *TTR : Traduction, Terminologie, Rédaction* 8: 2. 89-100.
- Cartier, Emmanuel (2016) "Neoveille, système de repérage et de suivi des néologismes en sept langues". *Neologica* 10. 101-131.
- Cartier Emmanuel (2017) "Néoveille, a web platform for neologism tracking" EACL 2017, Valencia.
- Gérard Christophe, Ingrid Falk and Bernhard Delphine (2014) "Traitement automatisé de la néologie :pourquoi et comment intégrer l'analyse thématique ?". Actes du 4e Congrès Mondial de Linguistique Française (CMLF 2014), July 2014, Berlin, Allemagne 8. 2627 2646.
- Gevaudan Paul. and Peter Koch (2010). "Sémantique cognitive et changement sémantique". In: *Grandes voies et chemins de traverse de la sémantique cognitive, Mémoire de la Société de linguistique de Paris,* XVIII. 103-145.
- Hamilton William L., Jure Leskovec, and Dan Jurafsky (2016). "Diachronic word embeddings reveal statistical laws of semantic change". ACL 2016.
- Harris Zellig S. (1970). Papers in Structural and Transformational Linguistics. Dordrecht/ Holland: D.Reidel.
- Kerremans, Daphné, Susanne Stegmayr and Hans-Jörg Schmid (2012). "The NeoCrawler: identifying and retrieving neologisms from the internet and monitoring on-going change". In: Kathryn Allan and Justyna Robinson. (Eds.). Current Methods in Historical Semantics. Berlin: Mouton de Gruyter. 59-96.
- Kilgarriff, Adam, V. Baisa, V and Bušta (2014). "The Sketch Engine : ten years on", *Lexicography* ASIALEX (2014) 1:7.
- Langacker, Ronald W. (1987). Foundations of Cognitive Grammar. Vol. I: Theoretical Prerequisites. Stanford: Stanford UP.
- Mikolov Tomas, Wen-tau Yih and Geoffrey Zweig (2013). "Linguistic regularities in continuous space word representations". *Proceedings of NAACL-HLT 2013*. Atlanta, Georgia, 9–14 June 2013. 746–751.
- Renouf, Antoinette (1993). "Sticking to the text: A corpus linguist's view of language". ASLIB Proceedings 45:5.
- Renouf Antoinette and Andrew Kehoe (2013). "Filling the gaps: Using the WebCorp Linguist's Search Engine to supplement existing text resources". *International Journal of Corpus Linguistics* 18: 2. 167-198.
- Schmid, Hans-Jörg (2015) "A blueprint of the Entrenchment-and-Conventionalization Model". Yearbook of the German Cognitive Linguistics Association 3. 1-27.
- Traugott Elizabeth C. and Graeme Trousdale (2013). Constructionalization and Constructional Changes. Oxford:Oxford UP.
- Turney Peter d. and Patrick Pantel (2010) "From Frequency to Meaning: Vector Space Models of Semantics". Journal of Artificial Intelligence Research 37: 1. 141-188.

The Logoscope: a Semi-Automatic Tool for Detecting and Documenting the Context of French New Words.

The Logoscope is a tool especially developed to detect new words of the French language, to document them and allow a public access through a web interface (http://logoscope.unistra.fr). This semi-automatic tool collects new words daily by browsing the online versions of well-known French newspapers.

In contrast to other existing tools essentially dedicated to dictionary development, the Logoscope attempts to give a more complete account of the context in which the new words occur. In addition to the commonly given morpho-syntactic information it also provides information about the textual and discursive contexts of the word creation; in particular, it automatically determines the (journalistic) topics of the text containing the new word.

After giving a general overview of the developed tool, we will defend the importance of a contextual approach of neology, especially as far as the relation between themes and new words is concerned.

covfefe, ?: ????

Jack Grieve University of Birmingham

grievejw@gmail.com

Geographical patterns of lexical innovation

In this presentation, I discuss how we can track and visualise the geographical spread of new words by monitoring social media. Specifically, I look at regional patterns of lexical innovations in American English over the course of 2014 based on a 9 billion word corpus of geo-coded American Tweets. First, I consider how to extract emerging words from a large corpus of time-stamped data and use this technique to identify 54 such forms in the Twitter corpus, all of which were very uncommon at the start of 2014 but rose dramatically in frequency over the course of the year. After exemplifying various ways in which the spread of words can be mapped, I identify five main hubs of lexical innovation on American Twitter by subjecting the maps for the complete set of emerging words to a multivariate spatial analysis. Finally, I offer explanations for these results and consider how they inform theories of lexical variation and change.

Negativity and lexical innovation

It is often argued that natural language is biased towards negative differentiation, meaning that there is more lexical diversity in negative affectual language, compared to positive language. However, we lack an understanding of the diachronic linguistic mechanisms associated with negative differentiation. In this talk, I will review key con-cepts related to negative differentiation and discuss how I am using diachronic word embeddings to test whether negative lexical items are more semantically unstable than positive ones. Preliminary results suggest that lexical innovation may be more prevalent for negative affectual language, compared to positive language. I will finish my talk by discussing some practical consequences of this positive/negative asymmetry for sentiment analysis tools.

bromosexual, n.: a man in a close friendship with another man without being homosexual

cyberbullying, n.: the use of Internet and mobile phones to send embarrassing or hurting messages

William Hamilton Stanford University

Suzanne Kemmer Rice University

kemmer@rice.edu

Lexical blending in an integrated usage-based theory of word formation

Lexical blends are undeniably a significant word formation type among neologisms in present-day English. The sheer number of coinages means that lexical blends can no longer be dismissed as marginal to or even, as sometimes claimed, outside of the overall system of English word formation. As such, morphological theories that aspire to empirical adequacy require some account of them. Further, any linguistic theory that is committed to accounting for how speakers understand and produce instances of language, i.e. any usage-based theory, must incorporate them, since English speakers produce and interpret lexical blends every day, and blends are found as neologisms in other languages as well. What is needed is an account that is empirically well-grounded, consistent with what is known of cognitive and social processes involved in creation, interpretation, and propagation of linguistic units, and that accommodates the various types of blends naturally and integrally, despite their heterogeneity and distinctive properties.

Recent work has advanced our knowledge of the empirical range of blends and the identification and treatment of methodological and theoretical issues relating to them (Gries 2004, 2012, Lepic 2016, and the work in Renner et al. 2012). Building on Kemmer (2003) and taking account of recent work, I describe in general terms the essential formal and functional characteristics of lexical blends, and posit cognitive mechanisms for how they integrate formal and semantic information of various types. I sketch a theoretical account in Cognitive Grammar that relates blend formation types to one another and to the morphsyntactic processes closest to them. It is argued that the sub-morphemic word parts found in blends are not qualitatively different from morphemes, and that viewing morphemes along a cline of symbolic units defined by various properties provides a natural account of the properties observed in blends and gives a basis for integrating them into linguistic theory, although there are still methodological and theoretical issues to deal with. Finally, I will suggest how studying blends can give insight into the cognitive and social processes involved in the novel creation and spread of lexemes.

References

- Gries, Stefan Th (2004). Shouldn't it be *breakfunch*?A quantitative analysis of blend structure in English. *Linguistics* 42: 3. 639-667.
- Gries, Stefan Th. (2012). Quantitative corpus data on blend formation: Psycho- and cognitive-linguistic perspectives. In: Renner et al. (eds.) 2012, pp. 147-167.
- Kemmer, Suzanne (2003). "Schemas and lexical blending". In: Cuyckens, Hubert et al. (Eds). *Motivation in Grammar*.. Amsterdam: Benjamins.
- Lepic, Ryan. 2016. "Lexical blends and lexical patterns in English and in American Sign Language". In: Jenny Audring, Francesca Masini, and Wendy Sandler (Eds.). Online Proceedings of the Tenth Mediterranean Morphology Meeting MMM10 (Haifa), 7-10 September 2015, pp. 98-111. lilec-mmm.it/?page_id=738
- Renner, Vincent, François Maniez, and Pierre Arnaud (Eds.) (2012). Cross-Disciplinary Perspectives on Lexical Blending. : Berlin: de Gruyter.

Terttu Nevalainen University of Helsinki

Approaching lexical innovation in earlier English

Addressing three of the challenges presented for this workshop, my paper focuses on neology detection, monitoring of the diffusion of neologisms and their potential sources in earlier English. The period I am largely concentrating on is Early Modern English (1500–1700). Some comparisons will also be made with processes of neologizing in 19th-century Finnish.

By way of background I will discuss some Early Modern English trends in lexical innovation based on historical dictionaries that have come to be accepted as received wisdom (Nevalainen 1999). Now that we have access to large historical corpora and digital databases, speaker innovations in particular are up for reevaluation (e.g. Elliott & Valenza 2011). Although accessing the moments of coinage of new words in earlier English may be beyond the means of historical lexicographers and lexicologists, the new tools make it possible to better monitor their process of diffusion over time. This is particularly the case with loan words in Early Modern English, which is characterized by intense lexical borrowing (Durkin 2014). For a more accurate detection of native neologisms we could follow, for example, the rise of the Finnish literary language, which was a conscious creation of the country's academic elite in the 19th century.

References

Durkin, Philip (2014). Borrowed Words: A History of Loanwords in English. Oxford: Oxford UP. Elliott, Ward E. Y. and Robert J. Valenza (2011). "Shakespeare's vocabulary: Did it dwarf all others?" In: Ravassat,

Mireille and Jonathan Culpeper (Eds.). *Stylistics and Shakespeare's Language*. London: Continuum. 34-54. Nevalainen, Terttu. 1999. "Early Modern English lexis and semantics." In: Lass, Roger (Ed.). *The Cambridge History of the English Language*. Vol. III. Early Modern English 1476–1776. Cambridge: Cambridge UP. 332-458. 16

piotr.pezik@gmail.com

Exploring markers of lexical conventionalization in live monitor corpora

Lexical innovation can be a complex process triggering multiple stages of conventionalization. New word formations which score similarly on a scale of institutionalization, i.e. reduction of ambiguity (Bauer 1983) may differ greatly in their diffusion across genres, topical domains and channels of communication. Recent advances in corpus linguistics and natural language processing have made it possible to use large reference text collections to verify the diffusion of neologisms and neosemanticisms and monitor their transition from compositional nonce-formations to fully institutionalized lexemes.

This paper showcases the role of live (i.e. continuously updated), webbased monitor corpora in tracking markers of conventionalization and dispersion of lexical innovations. More specifically, the paper introduces the Monco corpus search engines available for English (6.5 billion words) and Polish (4.1 billion words) at monitorcorpus.com. We show how these tools can be used to study lexical innovations in terms of their a) dispersion across sources, b) diachronic trends, c) morphological productivity, d) development of phraseological and selectional patterns as well as e) co-occurrence with metalinguistic markers indicating novelty.

The different degrees of conventionalization emerging from the exploration of these collections are exemplified by selected innovations found in the aforementioned reference corpora. The first of these items is the adjective *hangry* which emerged as a blend of *hungry* and *angry*. An analysis of the Monco data suggests that this adjective is relatively sporadic in spontaneous use and that it tends to be introduced by metalinguistic markers which explicitly describe it as a new and potentially ambiguous word. Furthermore, the dispersion of *hangry* across sources is rather low and it seems to inherit most of its phraseological and semantic arguments from its constituent adjectival stems. The second item considered in this study is the compound noun *post-truth*, whose first use goes back at least to the early 1990s. What makes this item interesting is that its frequency spikes in the Monco data seem to have been a direct result of its announcement as "the Oxford Dictionaries Word of the Year 2016", thus illustrating the significance of explicit promotion in the diffusion of such innovations. Since its recent re-appearance, *post-truth* has begun to recur in a small set of fixed phrases, such as the commonplace expression we live in a post-truth world/ era/ society, which testifies to its partial conventionalization.

The diffusion of *post-truth* can be contrasted with that of the noun *emoji* which owes its high frequency to more naturally dispersed proliferation across different internet text genres. The established status of *emoji* is further confirmed by the fact that it has developed a number of type-bound and restricted collocations, e.g. *winking/ crying emoji, text an emoji, to caption (a/ one's message) with an emoji,* etc. This study also looks at additional markers of conventionalization such as spelling adaptation and morphological productivity, which are particularly important in Polish as a less influential, but highly inflectional language. In particular, we consider two sets of lexical innovations which originally appeared as loanwords from English and which have resulted in numerous derivational and inflectional variants based on the stems *hejt.** and *fejs.**. Apart from presenting the current functionality of Monco and its applications in tracking lexical innovation, the paper also discusses the main limitations of these tools and outlines plans for future improvements.

References

Jelena Prokić, Daphné Kerremans, Quirin Würschinger & Hans-Jörg Schmid LMU Munich

jelenaprokic@gmail.com

The NeoCrawler: When frugalista meets bankster

In this talk I will present the NeoCrawler¹ – a tailor-made webcrawler, which identifies and retrieves neologisms from the Internet and stores data necessary for the systematic monitoring of their early diffusion in the language and the speech community. The NeoCrawler consists of two modules: Discoverer and Observer. While the Discoverer attempts to detect new words on the whole Web as closely to their date of coinage as possible, the Observer performs weekly searches for selected neologisms and stores retrieved HTML pages. The NeoCrawler relies on Google Custom Search API to search for neologisms and retrieve corresponding web pages and their weekly frequencies. The retrieved HTML pages are automatically post-processed by means of meta-data cleaning, tokenization, co-text extraction and POS tagging. These pages can then be used for linguistic analysis of the neologisms.

In the *Incipient diffusion of lexical innovations* project², whose main objective is to shed more light on the incipient diffusion of neologisms and the factors influencing the trajectory of the diffusion, we rely on the NeoCrawler to monitor close to 600 newly coined words with respect to the degree of the spread within the speech community as represented on the Internet. A number of factors responsible for the behaviour of the newly coined words will be investigated, including morphological and semantic transparency, social status of the coiner, style and mode of use, as well as context-related features in which the word is encountered. In the second part of my talk, I will address technical and methodological challenges involved in the automatization of data collection, annotation and analysis. I will conclude the talk with a case study that highlights the benefits and potential of the NeoCrawler in studies on lexical innovation.

Social embedding of neologisms in early English correspondence

This study is part of a project that focuses on sociolinguistic variation in the spread of new words, tracing lexical productivity and creativity over time and space in the Corpora of Early English Correspondence. The CEEC represents a wide social spectrum, richly documented in the metadata associated with the corpus. Coupled with the new interface to be developed in the project, this will enable us to ask new kinds of questions about neologisms in personal letters:

- 1. Who are the innovators? Which social groups do they represent?
- 2. How do the new words spread socially, geographically and diachronically?
- 3. Which semantic domains do the neologisms represent?
- 4. Why are the neologisms created and established? Can they be linked to specific historical events or changes in culture and society? What kinds of social meanings are associated with them?

The goal of the project is to provide a more balanced picture of the history of English neologisms as a wider range of social groups, including women and the lower ranks, is taken into account. The diffusion of lexical innovation in the long diachrony is for the first time studied in a maximally representative social setting that resembles spoken conversation, the hotbed of language change, which may lead to surprising discoveries. The tools and methods developed in the project will enable others to conduct similar types of research.

The identification of neologisms in a moderately sized historical corpus is a challenging task that sets a number of requirements to the toolkit developed in the project, from detecting spelling variants to retrieving related lexicographical data from the Oxford English Dictionary (OED) and its Historical Thesaurus (HT). Previous collaboration between some of the project participants has already produced tools for analysing lexical diversity (Säily & Suomela 2009; Suomela 2016) as well as for filtering and categorising corpus search results (Mäkelä et al. 2016), which may be used as starting points for further development.

¹ http://www.neocrawler.anglistik.uni-muenchen.de

² http://www.anglistik.uni-muenchen.de/abteilungen/sprachwissenschaft/ research/research_projects1/dfg-projekt/index.html

Tanja Säily^a, Eetu Mäkelä^b & Jukka Suomela^b ^aUniversity of Helsinki, ^bAalto University

tanja.saily@helsinki.fi, eetu.makela@aalto.fi

Jan Svanlund Stockholm University

We will present two case studies of innovative nominalisations in *-erand -ity in* the late 18th century, comparing the answers they yield to our research questions and considering their implications for (semi-)automating the process of identifying and tracing neologisms in the corpus. We are currently negotiating with Oxford University Press to gain access to local copies of some of the OED and HT data, which will facilitate our tool development.

We look forward to feedback on our plans from the workshop participants.

References

- Mäkelä, Eetu, Tanja Säily and Terttu Nevalainen (2016). "Developing an interface for historical sociolinguistics." Paper presented at the Digital Humanities Congress, Sheffield, September 8–10.
- Säily, Tanja and Jukka Suomela (2009). "Comparing type counts: The case of women, men and ity in early English letters." In: Renouf, Antoinette and Andrew Kehoe (Eds.). *Corpus Linguistics: Refinements and Reassessments*. Amsterdam: Rodopi. 87–109.
- Suomela, Jukka. 2016. "types2: Type and hapax accumulation curves." Computer program. http://users.ics.aalto.fi/suomela/types2/.

Metacomments and metasignals – what can they tell us and how do they affect conventionalization?

Many lexical innovations receive explicit metacomments during their conventionalization:

(1) *Plötsligt har begreppet "bokstavsbarn" fått en nedsättande klang.* 'Suddenly the concept "letter child" has received depreciatory ring.'

(2) Göran Tägtström värjer sig inför uttrycket "stafettläkare" som används på många håll i landet.

'Göran Tägtström opposes the expression "relay doctor" which is used in many places around the country.'

(3) Så kallade curlingföräldrar är föräldrar som sopar banan framför sina barn.

'So-called curling parents are parents who sweep the track for their children.' There also exists several ways of metalinguistically emphasizing e.g. a new word as a word in running texts. These can be called *metasignals*. The citations above include some examples: quotation-marks "X" (nr 1 and 2), the hedge så kallad X 'so-called X' (nr 3), and metacategorizing labels, like *ordet/uttrycket/begreppet X* 'the word/expression/concept X' (nr 1, 2). All these signal a metaperspective on the word X. They are methodologically convenient, since many of them are easy to find in large corpora. The reasons for metasignalling vary considerably. Metasignals are not always accompanied by explicit verbal comments. Sometimes they function as implicit metastatements by themselves. They can indicate, for instance, that the word is odd, deviant, contentious, controversial, or that the writer wants to dissociate herself or himself from the actual wording, which is said to be used by others. Sometimes they merely mark that the word is being introduced.

In other instances, metasignals are also combined with explicit comments, as in the examples above. Thus, it is sometimes useful to distinguish metasignals and actual metacomments. The term *metafocusing* can be used to cover them both. A large study of Swedish neologies, currently being concluded, indicates that metasignals and metacomments are natural features of conventionalization. More than 95 percent of the neologies were metafocused some time (in spite that all explicit comments could not be conclusively identified). However, new words receive metafocus to extremely varying degrees.

Some of the 360 neologies (representing more than 500 000 citations from the Swedish Language Bank) displayed some metafocus in almost half of the citations. A few others were never accompanied by such features at all. On average, only 3 percent of the citations were metafocused.

Metafocusing is likely to be more intense in early phases and later to decline. Thus, conventionalization seems to affect metafocusing, but does metafocusing at the same time affect conventionalization? Does it speed up the process and make words more well-known more quickly? So far, the project has not been able to show any long-term effects of metafocusing in the corpus data, such as statistical correlations between early high degree of metafocus and later frequency growth or dispersion to more text types. However, further elaborations of the project are intended to analyze possible effects on cognitive accessibility among language users. Metafocusing is potentially useful in the study of neologies, but we still don't know enough about its role in conventionalization.

Abstracts Posters

upcycle, v.: to convert waste into a useful product

burquini/burkini, n.: a fully-body swimsuit

Daniel Kjellander Umeå University

daniel.kjellander@umu.se

Cognitive constraints in English lexical blending: towards the development of a data collection methodology and an explanatory model

Research on English lexical blends such as *foolosophy* (*fool + philosophy*) constitutes a growing and multifaceted field. Despite the well-attested complexity of this word formation process, recent studies report a number of significant patterns (Beliaeva 2014; Gries 2004, 2006, 2012; Kelly 1998; Lehrer 1996; 2007 to name a few). Indeed, the importance of lexical blending appears to be a powerful process in the change and progress of the English language (Bauer 2006; Cannon 1986; Lehrer 2007). Yet, many aspects of the formation and development of blends remain unresolved, especially as regards semantic and conceptual concerns. Collecting lexical blends in a systematic way presents another challenge. Their ephemeral nature (Bauer 2006; Lehrer 2007), the sociopragmatically influenced distribution (Fandrych 2008), and their character of novelty (McMahon 1994) indicate that traditional sources of retrieval are limited. Instead, the emergence and growth of digital technologies have prompted a methodological development widening the explanatory scope (Gries 2012).

- The current study of lexical blends is concerned with mainly two questions;
- i) how can data collection methods be constructed so as to maximize the significance and relevance of the data, and
- ii) what governing principles can be discerned in their formation and development?

To address these issues, a method was devised with which the initial focus was shifted from lexical blends as such to their source words. Three such matrix words (i.e. lexical items providing a structural base) were chosen to form the basis for a number of corpus searches of truncated forms in the GloWbE corpus¹. This resulted in a database consisting of 94 types of lexical items (2126 tokens), of which 43 (210 tokens) were considered lexical blends. The data analysis revealed not only how certain structural properties of the matrix words either fueled or hampered the production of lexical blends, but it also highlighted the technical possibilities to detect patterns of formation. The findings are subsumed under the broad notion of cognitive constraints, which in this context is intended to encapsulate a set of fundamental cognitive principles governing the development of lexical blends. One such constraint is related to the impact of morphological lexicalization, or secretion (i.e. the gradual development of a lexical fragment to a morpheme) while another is called *schema transfer effects*. The results suggest that the former process appears to impede the production of lexical blends, while the latter fuels blend formation through the imposition of a morphological schema on a source word; for instance, *republican* is treated as a complex lexeme prefixed with *re*- (cf. Cannon 1986; Lehrer 1996; Soudek 1971) which apparently has inspired coiners to form novel lexical blends such as *demopublican* and *rubepublican*.

In all, the study demonstrates the possibilities to capture natural language data illustrating various stages of the development of neologisms. Furthermore, it also constitutes the starting point of a current research project exploring what further cognitive constraints can be identified in lexical blending and how these interact.

References:

- Bauer, Laurie (2006). "Compounds and Minor Word-formation Types". In: Aarts, Bas and April S. McMahon (Eds.). The Handbook of English Linguistics. Oxford: Blackwell. 483-506
- Beliaeva, Natalia (2014). "A study of English blends: From structure to meaning and back again". *Word Structure* 7: 1. 29-54.
- Cannon, Garland (1986). "Blends in English word formation". *Linguistics* 24: 4. 723-753.
- Fandrych, Ingrid (2008). "Pagad, Chillax and Jozi: A multi-level approach to acronyms, blends, and clippings". Nawa: Journal of Language & Communication 2: 2. 71-88.
- Gries, Stefan Th. (2004). "Shouldnt it be *breakfunch*? A quantitative analysis of blend structure in English". *Linguistics*, 42: 3. 639-667
- Gries, Stefan Th. (2006). "Cognitive determinants of subtractive word formation: A corpus-based perspective. *Cognitive Linguistics* 17: 4. 535-558.
- Gries, Stefan Th. (2012). ⁴Quantitative corpus data on blend formation: Psycho- and cognitive-linguistic perspectives". In: Renner, Vincent, François Maniez and Pierre Arnaud (Eds.). *Cross-Disciplinary Perspectives on Lexical Blending*. Berlin: De Gruyter. 145-167
- Kelly, Michael (1998). "To "brunch" or to "brench": Some aspects of blend structure". Linguistics 36: 3. 579-590.
- Lehrer, Adrienne (1996). "Identifying and interpreting blends: An experimental approach". *Cognitive Linguistics* 7: 4 359-390.
- Lehrer, Adrienne (2007). "Blendalicious". In: Munath, Judith (Ed.). *Lexical Creativity, Texts and Contexts*. Amsterdam: John Benjamins. 115-133.

Kenjiro Matsuda Kobe Shoin Women's University

kenjiro@shoin.ac.jp

Using the parliamentary minutes for the identification of group language neologisms

One of the challenging problems of empirical studies concerning lexical innovation is how to locate the time and date of the first use of the neologism in a systematic and reliable way. This paper is a report of using the minutes of the Japanese parliament as a means of locating the first use of a group language in the parliament. Group language is a cover term for slang, jargon, etc. that emerge as the network among group members grows dense (Shibata 1958), and as such, it could also grow among the Members of Parliament. What makes the minutes ideal for an empirical study of neologism is that they enable us to observe the birth and growth of every new group language, with exact dates and the linguistic contexts of all the uses.

We report two cases of neologisms here: terebi-iri 'TV-in' and okyooyomi 'sutra chanting'. In the Japanese parliament, the Main and the Budgetary Committeesessions are broadcast on the national TV network, and as this is a great opportunity for the MPs to appeal to their local constituents, most MPs are rather conscious of the broadcast, explicitly stating in their speeches that the current session is on TV. This used to be expressed in a sentential form, but after an MP used its nominalized version in February 2004, the new form, terebi-iri, quickly spread among the MPs and has come to be used widely in their speeches. The word has come to be used in blogs and newspapers as well, but their use of the neologism outdates the first use in parliament, corroborating the fact that the minutes did catch its first use. Okyoo-yomi refers to the monotonous speech by ministers that explains the bill they have put before the parliament. The term was first used in March 1965, and soon became a staple in Japanese 'parliamentarese'. In this case, too, the exact date of emergence and its later development are clearly observable.

Note that in both cases, the new words were introduced by the members of the opposing party, and spread in a party-by-party manner. Thus, *terebi-iri* was first used by the Democratic Party members, then the Liberal Democratic Party members began to use it, followed by the members of other smaller parties. Similarly, *okyoo-yomi* was first used by the Socialist Party, followed by other parties.

This process is ascribable to the nature of group language, as it was born out of the dense network among the party members that symbolizes their group membership. It is also found that the transmission of the neologism occurs through the question and answer interaction in the committee sessions, where the physical distance between the government and the opposing parties is much closer than that in the Main sessions.

Our study strongly suggests that the analysis of the group language neologisms in parliament is a promising method for the empirical study of neologisms.

slacktivism, n.: a form of activism that seeks projects and causes that require the least amount of effort, often online.

Dong Nguyen^a, Taha Yasseri^b and Barbara McGillivray^c Alan Turing Institute at ^aEdinburgh University, ^bOxford University, ^cCambridge University dnguyen@turing.ac.uk

Urban Dictionary as a source for studying lexical innovations

With the rise of the web 2.0 and user-generated content, new sources to study lexical innovations have become available. In this talk, we explore the use of Urban Dictionary, an openly edited online dictionary with mostly slang words and phrases. In contrast to traditional dictionaries, which are curated by a small set of editors, Urban Dictionary is collaboratively created by a large, (semi-)anonymous, online crowd. Users submit defi-nitions and these definitions are then voted up or down by the community. Many words contain multiple definitions added over time and thus these different definitions compete against each other. As a result, the most agreed on definitions emerge. For example, the word *emo* contains over 1,200 definitions in our dataset, with the top definition having 71,536 thumbs up and 28,327 thumbs down. Compared to other corpora, Urban Dictionary offers several advantages: (i) it captures infrequent, highly informal language; (ii) submitted definitions are added quickly, due to a large group of volunteer editors; (iii) various meanings and usages of a word are represented, rather than only the conventional ones.

We will present an exploratory study based on the collected data to illustrate: What types of word formation processes can be found in Urban Dictionary? Why are some definitions more popular than others? How do the meanings of new words evolve over time? As an illustration, searching for frequent words (with over 20 definitions) that appeared for the first time in 2014 resulted in: *hashtag abuser, truffle butter, trap queen, meninist, meeking, idghp, flappy bird, on fleek*, and *catching the crazy turkey*. While *Flappy Bird* is the name of a game (which received immense popularity early 2014), (*eyebrows*) on fleek is a phrase that gained popularity in 2014 and its meaning and spread has been a topic of discussion¹ and *meninist* is a term used in the men's right movement². While Urban Dictionary is a promising source to complement existing sources for studying lexical innovation, it also introduces various challenges due to noise and biases in the data. We will end with a reflection on these challenges.

Anabella-Gloria Niculescu-Gorpin and Monica Vasileanu

The 'lorgu lordan – Alexandru Rosetti' Institute of Linguistics, Bucharest , monica.vasileanu@gmail.com

Diffusion of Lexical Borrowings in Present-day Romanian. A Case Study¹

The dynamics of the Romanian vocabulary has been a matter of great concern over the past 25 years, with an emphasis on the massive import of lexical Anglicisms (e.g. Avram 1997, Stoichiţoiu-Ichim 2001, 2006, Niculescu-Gorpin 2013). However, most researchers seem to have failed to provide real criteria and/or methods suitable to assess and predict the diffusion of lexical Anglicisms in Romanian. For instance, 'luxury' Anglicisms, i.e. Anglicisms that have one or several long-established Romanian equivalents, such as *job, look, trend*, have been perceived as unnecessary, sometimes even useless and thus their life expectancy has been considered very low (Stoichiţoiu-Ichim 2001, 2006); Nevertheless, they are still present in everyday language, have developed extended families and some, if not all, are even preferred by Romanian native speakers (Niculescu-Gorpin & Vasileanu 2016, in press).

In their current research, Niculescu-Gorpin (2013) and Niculescu-Gorpin & Vasileanu (2016, in press) have shown that psycholinguistic testing of Romanian native speakers may hold the key to understanding the current language change phenomena and that such an approach may shed some light not only on the dynamics of the Romanian vocabulary, but also on the whole language system that is currently undergoing an unprecedented metamorphosis under the English influence.

During this presentation, we are going to show how psycholinguistic tools such as preference/acceptability questionnaires and the study of reaction times by means of E-prime-based experiments can be used as methods and criteria to evaluate the impact of particular lexical Anglicisms on present-day Romanian and can be reliable instruments for predicting their future influence.

We will discuss a particularly interesting experiment applied to over 100 Romanian native speakers in various age groups. In a nutshell, respondents were given 30 sentences and were asked to decide whether they made sense or not. Out of those, 10 were fillers, i.e. grammatically and semantically ill-constructed sentences, and 20 targets, i.e. meaningful ones. The experiment had two versions – in one, the targets were luxury Anglicisms, and in the other one, the long-established Romanian word.

¹ http://knowyourmeme.com/memes/eyebrows-on-fleek.

https://www.google.com/trends/explore?q=eyebrows%20on%20fleek

http://www.languagejones.com/blog-1/2015/2/23/socal-is-getting-fleeked-out

² https://www.buzzfeed.com/rossalynwarren/men-are-

calling-themselves-meninists-to-take-a-stand-against

¹ This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-2480.

Cecile.Poix@univ-lyon2.fr

Preliminary results show that some lexical Anglicisms are more easily processed than their Romanian equivalents (e.g. *shopping, look, trend* are processed faster than their Romanian equivalents *cumpărături, înfățișare, tendință*). Though word length may have triggered such preference, e.g. *look* is shorter than *înfățișare*, in other situations this has obviously not been the case (e.g. *background* was processed faster than its Romanian equivalent *fundal*).

The response times suggest that these borrowings are extremely well-integrated into the Romanian native speakers' lexicon, sometimes replacing established lexical items. Further analysis and research may provide highly valuable information about how other processing effort influencing factors (recency or frequency of use, linguistic complexity, etc.) may explain our findings, and can be correlated with the principle of relevance – that all cognition is geared towards maximization of relevance.

References

Ciobanu, Georgeta (1997). Romanian Words of English Origin. Timişoara: Editura Amphora.

Görlach, Manfred (Ed.) (2001). A Dictionary of European Anglicisms. Oxford: Oxford UP.

Mitchell, Mark L. and Janina M. Jolley (2010). Research Design Explained. Seventh edition. Belmont: Wadsworth Cengage Learning.

Niculescu-Gorpin, Anabella-Gloria (2013). Influențe Cultural-lingvistice Asupra Evoluției Recente a Limbii Române în Context European și Mondial. București: Editura Muzeului Național al Literaturii Române

Stoichiţoiu-Ichim, Adriana (2001). Vocabularul Limbii Române Actuale. Dinamică, Influenţe, Creativitate. Bucureşti: ALL.

Niculescu-Gorpin, Anabella-Gloria and Monica Vasileanu (2016). "Romanian lexical Anglicisms: norms and usage, speakers' attitudes and official recommendations". *Romania Orientale XXIX*. (in press).

Stoichiţoiu-Ichim, Adriana (2006). Aspecte ale Influenţei Limbii Engleze în Româna Actuală. Bucureşti: Editura Universității din București. Nonce-words in children's literature: some key features of morphological coinages.

The overwhelming presence of nonce-words in English literature for children implies that lexical innovation is essential to the fantasy-driven world of children's books. An opaque neologism often stands out like a magic formula. The fate of such nonce-words depends on their playfulness – and a possible film adaptation of the book. As such, Lewis Caroll's *unbirthday* has entered the lexicon. The coinage of nonce-words is here a recreational process to create idiosyncrasy and produce an effect on the reader/listener. My analysis of over 550 nonce-words in a corpus of British English children's literature showed that neologisms could be classified (following Jean Tournier's matrix of lexicogenesis) into three types of coinages:

(1) morphosemantic neology

(2) semantic neology

Cécile Poix Université Lyon 2

(3) morphological neology.

For each standard word-formation process, there is a recreational de-formation equivalence, and de-formation processes are often combined with one another. However, the characteristics of playful neologisms are not limited to lexicology. Stylistic features – such as reduplication – and the alteration of customary collocations are also word de-formation processes.

Semantic neology might be impossible to detect computationally. However certain features of morphological nonce-words can provide a model. This article presents the analysis of word length and string repetition of neologisms inventoried in the corpus.

Neology can be sound-motivated as books are often read out loud to the child. Coinages often play with the reduplication of sounds. I analyse reduplication as a word formation process under four different forms:

(1) pure reduplication: *thingalingaling*

- (2) rhyme-motivated: pish dish
- (3) ablaut-motivated: *clip clap clop*
- (4) alliterative compounds: <u>crackety-crack</u>

Prototypically, English words are monosyllabic or disyllabic. Scientific studies reveal that word length in spontaneous speech averages 1.2 syllables per word for children and ranges around 1.4 for adults. Following the least effort principle, polysyllabic words in English tend to be clipped and initials and acronyms are an increasingly productive word-formation process. On the contrary, in children's literature, signifiers are extended and 58% of nonce-words in the corpus are over three syllables long.

- a. scrumdiddlyumptious > scrumptious
 - b. rhinostossterisses > rhinoceroses > rhinos

The above examples illustrate that the idiosyncrasy is often achieved through the length of the lexemes and the complexity of their pronunciation. To emphasize complexity, letters can be repeated. Through repetition, reduplication and length, playful coinages can be identified.

References

(1)

- Carroll, Lewis (1998). Alice's Adventures in Wonderland and Through the Looking-Glass. Penguin Classics. Crystal, David (1990). "Linguistic Strangeness." In Margaret Bridges (Ed.). On Strangeness. Swiss Papers in English Language and Literature.
- Dahl, Roald (2007). The BFG and The Witches. Penguin Books Ltd.
- Flipsen, Peter, Jr. (2006). Syllables per Word in Typical and Delayed Speech Acquisition". Clinical Linguistics & Phonetics 20: 4 293-301
- Lecercle, Jean-Jacques (1990). The Violence of Language. London: Routledge.
- Milligan, Spike (2000). A Children's Treasury of Milligan. Virgin Books Ltd.
- Milne, Alan Alexander (1991). Winnie-The-Pooh. Reed Books.
- Tournier, Jean (2007). Introduction Descriptive à la Lexicogénétique de l'Anglais Contemporain. Slatkine Erudition.

Susanne Vejdemo Stockholm University

Predictive regression models combine with descriptive color inguistics to explain lexical replacement patterns

In this talk I wish to highlight two issues:

1) How insights from a top-down multiple regression model of factors that affect lexical replacement, can be combined with results from a bottom-up detailed qualitative and experimental case study of lexical innovation, diffusion and change in the PINK and PURPLE color categories of Swedish.

2) How different bottom-up data sources and methods used to investigate change in the Swedish color lexicon tell different and sometimes contradictory 'stories' of the process - yet these stories can be complimentary pieces of the puzzle.

In a 2016 paper, we show that the likelihood that a non-grammatical lexeme would be replaced with another lexeme could be predicted from several variables - a high number of synonyms had a positive correlation, while frequency, imageability, and the number of senses had negative correlations. Our multiple regression model predicts 34% of non-grammatical concepts' lexical replacement variation: This is considerably better than other published models for non-grammatical concepts. The model uses 167 non-grammatical items from the 200 item Swadesh list and the rate of lexical replacement from Pagel et al (2007), based on data from 87 language varieties.

But what does such a top-down statistical model mean when confronted with actual lexical data of (ongoing) lexical change? For example: when the model indicates that a high imageability rating for a lexical item (i.e the lexical item is easier pictured in the mind) insulates it somewhat from being replaced – what does this mean in practice for a particular lexical item in a particular semantic domain?

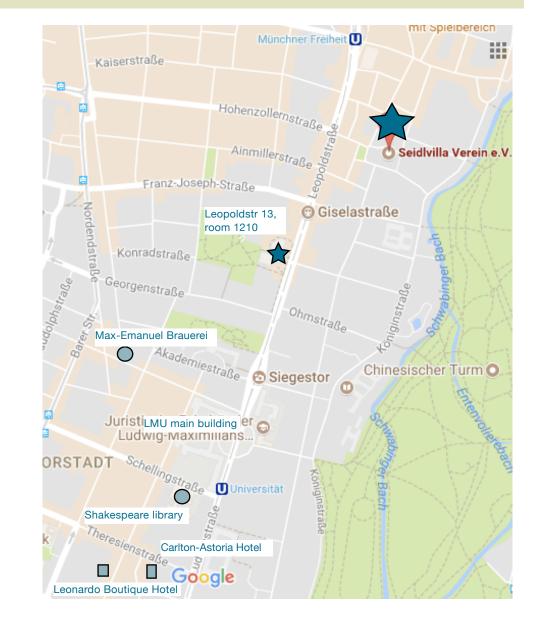
I argue that while some semantic domain-independent rules for lexical innovation and replacement can be posited, the fate of non-grammatical lexical items is also governed by domain-dependent processes. In the case of the color domain, this includes specific semantic change tendencies. For example: high imageability of a color term can be tied to a high consensus level for a group of speakers in color elicitation experiments for labeling stimuli with that color term. Different kinds of language data (e.g. lexicons, fiction books) have different levels of inertia when it comes to recording lexical and semantic change in language communities. I will discuss color elicitation data from experiments with two generations of speakers, as well as different textual data sources from the last two centuries: a corpus study of fiction, changing flower descriptions in ten botanical encyclopedias, and several successive lexicons and dictionaries.

The factors affecting lexical replacement in the domain-independent statistical model can be interpreted in the light of the data from this specific semantic domain. In the domain-specific data we see that low frequency is a natural condition of disappearing color terms (but it is also a feature of emerging terms); high lexical competition among color words leads to many near synonyms; a high level of imageability for a color term is similar to a high level of consensus in a group of speakers for the denotation of that term; and that having several different senses can insulate a word from lexical replacement.

References

- Pagel, Mark, Quentin.D Atkinson and Andrew Meade (2007). "Frequency of word-use predicts rates of lexical evolution throughout Indo-European history." *Nature* 449: 7163. 717–20.
- Vejdemo, Susanne, and Thomas Hörberg (2016). "Semantic factors predict the rate of lexical replacement of content words". PLOS ONE 11: 1.

Orientation Area map



Useful information

Workshop venues Leopoldstr 13 (June 28) & Welcome reception

June 28: Leopoldstraße 13, room 1210 = Haus 1, 2nd floor, room 210

How to get there: U-Bahn U3/U6 Giselastraße (exit F), bus 154 (Georgenstr), bus 54 & 150 (Giselastr) 10 min from LMU main building

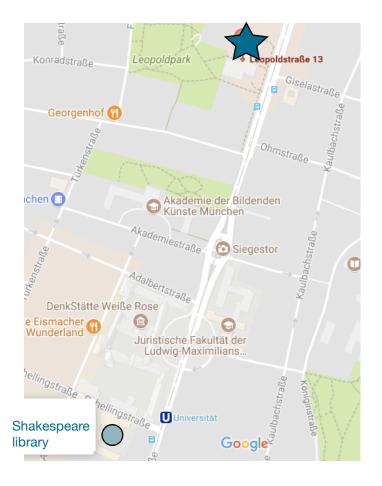
June 29-30: Seidlvilla, Nikolaiplatz 1a

How to get there: U-Bahn U3/U6 Giselastr (exit C), bus 54 (Hohenzollernstr), bus 53 & 59 (Münchner Freiheit)

20 mins walk from LMU main building

Munchner Freihett Feilitzschstraße Feilitzsch The welcome reception takes place at the Shakespeare library at the Institute of English Studies, room 207 (2nd floor), Schellingstraße 3 back building

How to get there: U-Bahn U3/U6 Universität (exit D) bus 154 (Schellingstr) 2 mins walk from LMU main building, 10 mins from the Leopolstr



Useful information Lunch places around the Seidlvilla

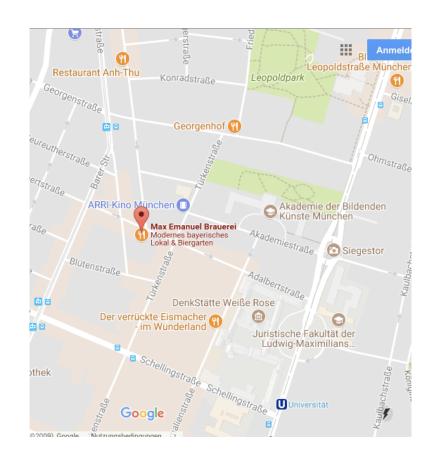
- ① Ruff's Burger, Occamstraße 4
- ② Alles Wurscht, Nikolaiplatz 3
- ③ Bapas/Mauerer Cafe, Leopoldstraße 52A
- ④ Dean and David, Leopoldstr 52
- ⑤ Asahi running sushi and buffet, Leopoldstr 27
- 6 Bachmair Hofbräu, Leopoldstraße 50
- ⑦ Don Luca, Leopoldstr 44
- 8 L'Osteria, Leopoldstra



Useful information Conference dinner

The conference dinner takes place on Thursday, June 29 at 7 pm at the **Max-Emanuel Brauerei**, Adalbertstr. 33 (max-emanuel-brauerei.de/).

How to get there: U3/U6 Universität (exit A) 20 minutes on foot from the Seidl villa



Notes	Notes

Notes	Notes