

Research Topics

Kellert, Olga

<https://www.uni-goettingen.de/de/566300.html>

[Google Scholar Profile](#)

Topic Summaries:

Topic 1: Exploring Gender and Equity through Linguistic Expressions: A Comprehensive Review of Gender-Neutral Language (GNL)

It is generally assumed that gender-neutral language (GNL) is “a way of speaking that does not perpetuate gender-based stereotypes, and includes all gender identities” (Nissen, 2013). Lack of GNL is assumed to foster inequality against women (Erdocia, 2021). However, GNL is still socially, politically, and linguistically not recognized in many countries (Erdocia, 2021, **Kellert, 2022**). This research project endeavors to survey GNL usage across diverse nations and societal groups. The aim is to discern societal and linguistic adaptations regarding GNL, while uncovering the distinct factions resisting its utilization and their reasons for doing so. The findings will contribute invaluable insight into societal transformation and linguistic changes that have important political implications such as promoting gender-inclusive policies by ensuring equal linguistic representation for all genders.

Olga Kellert (2022) ‘Gender Neutral Language in (Greater) Buenos Aires, (Greater) La Plata, and Córdoba: An analysis of social context information using textual and temporal features’, *Frontiers in Sociology*, doi: 10.3389/fsoCarlos2022.805716.

Topic 2: Diagnosing misinformation in minority communities with NLP:

Indigenous people belong to a particularly vulnerable group that is disproportionately affected by epidemics and other crises, as acknowledged by the United Nations. In addition to socioeconomic marginalization which can affect factors such as health-care accessibility (especially in remote or rural areas), the sparsity of information presented in indigenous languages can lead to a high level of misinformation across all areas of life with potentially severe consequences when it comes to issues such as understanding instructions for use of medications. This project explores development of automated natural-language processing tools to improve assessment of misinformation and health education of ethnic minority groups in Latin America in the context of the recent COVID-19 pandemic.

Olga Kellert & Mahmud Zaman (2023) Use of NLP in the Context of Belief states of Ethnic Minorities in Latin America. In *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*, pages 1–5, Toronto, Canada. Association for Computational Linguistics

Olga Kellert, Claudia Crespo, Marleen Haboud, Fernando Ortega, Stavros Skopeteas (2023) Factores sociolingüísticos y demográficos que influyeron en la transmisión de información sobre COVID-19 en poblaciones indígenas y no indígenas rurales de Perú y Ecuador. Talk presented at the Hispanistentag at the University of Graz (Austria) in February 2023 and Talk presented at the University of Quito, September 2023.

Olga Kellert (2021) La desinformación en las redes sociales en América Latina, *Multilingüismo y migración en la era de las humanidades digitales* (23.09.2021, Bochum, invited talk)

Topic 3: NLP for heritage speakers and low-resource languages

AI, machine learning and large-language models like Chat-GPT are promising to revolutionize our way of life with dramatic social and economic implications. Training these tools, however, requires enormous language datasets that necessarily limits their application to the dominant languages of the world. Minority populations and heritage speakers are therefore at risk of being excluded from the benefits of this revolution and facing additional pressure to abandon their native languages. Development of natural-language processing tools specialized for handling low-resource languages therefore can represent a critical avenue not only for gaining insight into the social circumstances of the associated communities but fighting their marginalization as well (see agenda of [AmericasNLP](#)).

Olga Kellert & Mahmud Zaman (2023) Use of NLP in the Context of Belief states of Ethnic Minorities in Latin America. In *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*, pages 1–5, Toronto, Canada. Association for Computational Linguistics

Topic 4: Linguistic mapping and smart cities

In Smart Cities, a wide variety of geographic data is collected from citizens and infrastructure with the aim of improving operations and delivery of services by adapting to behavior patterns. Geographic patterns of language use, however, are nearly unexplored for this purpose. Social media data, which combines text with geolocation coordinates and user meta-data, represents a rich source of information which opens up many opportunities for correlating linguistic behavior with socioeconomic and geographic factors on small spatial scales. Such studies promise to add a new dimension to the Smart-City concept. In this project, tools were developed to extract geographic patterns of multiple linguistic features with connections to social behavior.

Olga Kellert (2023a) Using geolocated tweets for probing Language Geography and Migration. In Sandra Issel-Dombert, Ignacio Andrés Soria, Laura Morgenthaler García (Eds.). *Language, migration and multilingualism in the age of digital humanities*. De Gruyter. 129-137. doi: 10.1515/9783110746082-007

Olga Kellert (2023b) Linguistic variation in Twitter: a case study of Italian loanwords in Spanish of South America, in Natascha Pomino, Eva-Maria Remberger, Julia Zwink (eds.), *From formal linguistic theory to the art of historical editions: The multifaceted dimensions of Romance linguistics*, V&R unipress Publishers. 347-359.

Olga Kellert (2023c) Probing Sociodemographic Influence on Code-Switching and Language Choice in Quebec with Geolocation of Tweets. *Front. Psychol. Language Sciences*. Volume 14 - 2023 | doi: 10.3389/fpsyg.2023.1137038

Olga Kellert & Nicholas Matlis (2022) ‘Geolocation of multiple sociolinguistic markers in Buenos Aires’, *PLoS ONE* 17(9): e0274114.

Olga Kellert & Nicholas Matlis (2022) ‘Social context and user profiles of linguistic variation on a micro scale.’ *Proceedings of the Ninth Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 14–19, October 16, 2022. ©2022 Association for Computational Linguistics, <https://aclanthology.org/2022.vardial-1.2.pdf>

Topic 5: Mapping sociolinguistics in the modern age

Bilingualism is at the heart of the linguistic phenomenon of language change fueled by the coexistence of multiple languages of disparate levels of political and socioeconomical dominance. Yet much is still not known about which social factors affect bilingualism, proficiency in heritage languages and their impacts on various aspects of life, particularly on the small spatial scales within cities. Classic data sources such as governmental census surveys or personal interviews are time consuming, expensive and not well adapted to addressing sociolinguistic questions on a large data scale. Social media data offers a solution by providing large amounts of data in the form of natural communication bundled together with time stamps, precise geolocation coordinates and other user information enabling tracking of linguistic phenomena in time and space and providing a way of correlating them to social factors. In this project, we show that multiple linguistic markers, including dialect, speech formality and bilingualism can be spatially mapped on scales down to city blocks, providing an unprecedented view of social dynamics.

Olga Kellert (2023a) Using geolocated tweets for probing Language Geography and Migration. In Sandra Issel-Dombert, Ignacio Andrés Soria, Laura Morgenthaler García (Eds.). *Language, migration and multilingualism in the age of digital humanities*. De Gruyter. 129-137. doi: 10.1515/9783110746082-007

Olga Kellert (2023b) Linguistic variation in Twitter: a case study of Italian loanwords in Spanish of South America, in Natascha Pomino, Eva-Maria Remberger, Julia Zwink (eds.), *From formal linguistic theory to the art of historical editions: The multifaceted dimensions of Romance linguistics*, V&R unipress Publishers. 347-359.

Olga Kellert (2023c) Probing Sociodemographic Influence on Code-Switching and Language Choice in Quebec with Geolocation of Tweets. *Front. Psychol. Language Sciences*. Volume 14 - 2023 | doi: 10.3389/fpsyg.2023.1137038

Olga Kellert & Nicholas Matlis (2022) 'Geolocation of multiple sociolinguistic markers in Buenos Aires', *PLoS ONE* 17(9): e0274114.

Olga Kellert & Nicholas Matlis (2022) 'Social context and user profiles of linguistic variation on a micro scale.' *Proceedings of the Ninth Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 14–19, October 16, 2022. ©2022 Association for Computational Linguistics, <https://aclanthology.org/2022.varodial-1.2.pdf>

Topic 6: Sentiment analysis for small data

Sentiment analysis (SA) is a widely-used tool for gauging popular opinion with applications ranging from commerce to politics. Implementation of SA is frequently done via machine learning which typically requires large, annotated datasets to achieve sufficient accuracy. However, many domains of interest exist for which large amounts of annotated data do not exist. Methods such as syntactic parsers, which incorporate dictionaries and linguistic rules, can achieve superior accuracy as well as speed in domains with small data sets, such as low-resource languages, opening up new domains for analysis. Such tools, which are attractive not only to businesses but also researchers in sociolinguistics, require development of parsers and dictionaries that are tailored to the specific language variety or domain.

Carlos Gómez-Rodríguez, Muhammad Imran, David Vilares, Elena Solera & **Olga Kellert** Dancing in the syntax forest: fast, accurate and explainable sentiment analysis with SALSA, SEPLN – CEDI 2024. Seminar of the Spanish Society for Natural Language Processing at the 7th Spanish Conference on Informatics, CEUR Workshop Proceedings, A Coruña, Spain, 2024. ISSN 1613-0073.

Olga Kellert, Mahmud Zaman, Nicholas Hill Matlis, Carlos Gomez-Rodriguez Experimenting with UD Adaptation of an Unsupervised Rule-based Approach for Sentiment Analysis of Mexican Tourist Texts. CEUR Workshop Proceedings, Vol. 3496, Rest-Mex paper 15, Alvarez- Carmona et al. 2023, eds. <http://www.grupolys.org/biblioteca/KelZamMatGom2023a.pdf>