Arabic sociolinguistics and dialectology: Taking stock

University of Bayreuth, 10 - 11 July 2025

(University Campus, GW1, Room K5)

The interest in Arabic dialects has a long tradition going back, depending how one defines it, to Sibawaih. Since WWII it has grown into a significant, independent sub-discipline of Arabic in the west (Europe in particular). Arabic sociolinguistics (again ignoring Sibawaih), having grown up in the shadow of (mainly English-based) sociolinguistics is more recent, but it still has reached a point where it has achieved its own independence within the pantheon of Arabic linguistic sub-disciplines.

Sociolinguistics is an adventitious sub-discipline. Its input is defined by whatever language norms are discernible in the larger linguistic landscape: In the case of Arabic, Standard Arabic is one norm, a local dialect or dialects constitutes another. By the same token, Sociolinguistics offers insights into language which a structurally-based input (Standard Arabic, a given dialect) cannot offer. In particular – and Labov saw this as one of the key insights which could be gained from its study – it allows inferences and deductions to be made about language change.

But how much really can sociolinguistic variation tell us about language change as a general phenomenon? Arabic dialects display an intriguing degree of individual variation across the entire palette of linguistic phenomena. While some of this variation has been treated and elucidated in sociolinguistic terms, for instance the omnipresent “qaaf” variable, only a small portion of observed dialect differences have been treated using sociolinguistic methodology. A meta-question which poses itself therefore is, to what degree is it possible to link sociolinguistic results–change in progress for instance–to the broader question of how changes occurred in dialects which led to dialects as we know them today.

This is obviously a huge question, more interesting we think in its statement than in the ability in the contemporary research environment to answer it. Nonetheless, we would like to take a small step forward in this endeavor in recognizing the issue in our conference.

The empirical study of language starts with the case study. However, herein lies what we might call the case study – grand panorama divide. Case studies are necessary to elucidate the detailed reality

of language in speech communities. More often than not, however, their connection to a grander panorama goes no further than to define the study as a contribution to exemplification of particular phenomena of a particular dialect. The problem inheres in the sociolinguistic – dialectology dichotomy and is by no means a problem only for Arabic.

The conference "Arabic sociolinguistics and dialectology: Taking stock" has two orientating landmarks. On the one hand it features case-study level sociolinguistic studies, which also address the "grand panorama" perspective. In particular they address the question how the detailed study of individual variables elucidates an understanding of how distinctive dialect features emerged in the first place. On the other hand, other talks approach the issue from the opposite perspective. Beginning with a grand panorama, they address the question how does a dialect area reveal what finer level sociolinguistic processes might have been necessary to produce the dialect in its present-day form?

Program

Thursday 10 / 07	**14:00-14:15**	**Opening**
	14:15-15:00	**The Arabic dialects of northern Oman: some key features of the lexicon** *Clive Holes*
	15:00-15:45	**The emergence of the *šāwi* and (rural) *gilit* dialect areas in the light of historical sociolinguistics** *Stephan Prochazka*
	Coffee break	
	16:15-17:00	**Exploring Arabic dialect relationships across Africa through dialectometry: Insights and limitations of this approach** *Carolina Zucchi*
	17:00-17:45	**Morphosonority and indexical sociohistorical linguistics** *Jonathan Owens & Ajid Lawan*
	19:30 - Social dinner	
Friday 11 / 07	**9:00 – 10:00**	**Linguistic homogenisation and its effect on the loss and reconfiguration of traditional dialects** *Enam Al-Wer*
	Coffee break	
	10:30 – 11:15	**Short-term accommodation in a multilingual and multidialectal context, and the way forward** *Valentina Serreli*
	11:15 – 12:00	**The Role of Migration and Displacement in the Life Course of an Arabic Variety: A Linguistic and Ideological Analysis of Variation in the Dialect of Arab al-Maslakh in Beirut (Karantina)** *Ana Iriarte Díez*
	Lunch at Mensa	
	13:30 – 14:15	**Genitive markers and the Historical Development of Maghrebi Arabic dialects** *Felipe Benjamin Francisco*
	14:15 – 15:00	**Bifurcation** *Jonathan Owens*
	Coffee break	
	15:30-16:30	**Workshop session on CorpusCompass** *Muhadj Adnan & Nicolò Brandizzi*
	16:30	**Closing**

Abstracts

Linguistic homogenisation and its effect on the loss and reconfiguration of traditional dialects

Enam al-Wer, University of Essex

Sociolinguistic investigations of variation and change in Arabic vernaculars have normally focused on the trajectory of change, that is where an observed variable feature is heading. Indeed, our collaborative research in this area has found that linguistic change is moving towards supralocal features, which are normally characteristic of dominant city dialects or burgeoning conurbations. The accumulation of data from multiple communities, particularly in the Levant, Saudi Arabia, Oman and Iraq, has made it possible to identify common tendencies including koineisation, which involves for the most part the levelling out of localised linguistic features in favour of supralocal or shared features.

A different way of looking at the outcome of koineisation is to consider its flipside; that is, how it looks from the perspective of the local dialects; specifically, what effect it has on the coherence and identity of these dialects, and on dialect configuration in the respective region. From this perspective, what is happening through koineisation is loss of local distinctive features, which can potentially lead to the dilution or total loss of traditional dialects, a process that Peter Trudgill (1996) dubbed 'dedilectalisation'. The outcome of this process is reduction in variability and increase in homogeneity across the dialects of a particular regions. From a cultural perspective, dedialectalisation may also lead to or expedite the loss of local culture, which is embodied in the local dialect. In this talk, I illustrate the effect of dedilaectalisation within Jordan, based on dialectological and sociolinguistic data that span nearly one hundred years.

Genitive markers and the Historical Development of Maghrebi Arabic dialects

Felipe Benjamin Francisco, University of Bayreuth

The traditional classification of Maghrebi Arabic dialects into pre-Hilali and Hilali varieties—based on waves of Arabization—has long failed to reflect the region's complex linguistic reality. Cases such as the "mixed" dialects of cities like Meknes and Marrakesh or the typologically ambiguous varieties of southern Morocco challenge this binary model, both historically and linguistically. This paper

examines the diachronic development of one specific variable: the genitive marker. Focusing on the distribution of the variants mtāʕ and dyāl/d- (< *dil-), commonly linked to Hilali and pre-Hilali dialects respectively, the analysis draws on both contemporary dialectal data and historical documents dating from the 15th century onwards (ANTT/Yale archives). By tracing the emergence, evolution and distribution of these genitive forms, the study seeks to shed light on the broader historical dynamics that have shaped Maghrebi Arabic.

The Role of Migration and Displacement in the Life Course of an Arabic Variety: A Linguistic and Ideological Analysis of Variation in the Dialect of Arab al-Maslakh in Beirut (Karantina)

Ana Iriarte Díez, University of Vienna

This study presents historical, ideological, and sociolinguistic data on Arab al-Maslakh (ArMS), using it as a case study to explore how migration and displacement can fundamentally shape a community's social and linguistic emergence, development, and future trajectory. ArMS refers to an umbrella community originally composed of various self-designated Bedouin groups historically engaged in cattle herding, trading, and butchery, who settled in Beirut's Karantina neighborhood in the 20th century, drawn by economic opportunities in the municipal slaughterhouse and related industries (Stocker et al. 2026, forthcoming). Karantina's key institutions, along with its strategic location near the port, made it a hub for labor migrants and successive waves of refugees, including Armenians (1920s), Kurds (1930s), Palestinians (1948–67), and Syrians (since 2012), as well as economic migrants from southern Lebanon, the Bekaa Valley, Egypt, and Ethiopia (Lteif 2022).

A pivotal moment in the community's history occurred during the Lebanese Civil War (1975–1990). In January 1976, Christian militias perpetrated a brutal massacre in Karantina. Within hours, the ArMS community was attacked, besieged, and later forced into displacement. For at least 17 years, community members resettled in various locations across Lebanon (e.g., Saida, Naameh, Baalbek, Khalde) and abroad (e.g., Germany, Cyprus). This prolonged period of forced migration disrupted old social networks while simultaneously exposing members to a wide range of new social and linguistic environments — both of which left a lasting impact on linguistic practices. Those who returned to Karantina after the war (1992-1995) encountered a transformed urban environment, shaped by a post-war economy and a fractured social fabric, which soon received new waves of migration. More recently, the financial crisis that began in 2019, followed by COVID-19 and the 2020 Beirut Port explosion — which severely affected the neighborhood due to its proximity to the epicenter — further

exacerbated instability and triggered additional emigration, especially among the younger generation, with many relocating to Germany, Norway, and Sweden.

We hypothesize that migration and displacement have shaped linguistic variation in distinct ways across generations, with each wave of migration and every experience of displacement contributing to unique changes in the community's linguistic repertoire. The second part of this study examines these dynamics in detail through a sociolinguistic analysis of variation in the ArMS dialect, drawing on data collected during two fieldwork campaigns in 2022 and 2023 (Iriarte Díez & Laaber, 2026, forthcoming). Using both intra-speaker and inter-speaker approaches, we identify the main sociolinguistic variables in the speech of a sample of 30 community members. These variables are selected based on a thorough analysis of ideological data that allows us to classify them as markers or indicators, illustrating the perceived social meanings attached to their use. We then examine the distribution of different variants across speakers, paying particular attention to generational differences in contact-induced linguistic change. By adopting a stylistic approach informed by third-wave sociolinguistics (Coupland 2007; Eckert 2000, 2012), we explore how individual speakers negotiate linguistic variation in relation to their personal migration trajectories, social positioning, and interactional styles.

By integrating historical, ideological, and sociolinguistic perspectives, this study contributes to understanding how migration and displacement may shape language variation in Arabic dialects, while documenting the disappearing linguistic practices and narratives of a marginalized Bedouin community at the heart of Beirut.

References

Coupland, Nikolas. (2007). *Style: Language Variation and Identity*. Cambridge: Cambridge University Press.

Eckert, Penelope. (2000). *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Malden, MA: Blackwell.

Eckert, Penelope. (2012). "Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation." *Annual Review of Anthropology*, 41(1), 87–100.

Iriarte Díez, Ana & Laaber, Claudia. (2026, forthcoming). *Transforming Traditions: Dialect Contact and Sociolinguistic Variation in the Arabic of Arab al-Maslakh in the Migration Hub of Karantina (Beirut)*. Amsterdam: John Benjamins.

Lteif, Diala. (2022). "One Hundred Years of Refuge: Migrants and the Making of Karantina, Beirut (1918-2018)." Unpublished PhD Thesis. University of Toronto.

Stocker, Laura, Claudia Laaber, and Ana Iriarte Díez. (2026, forthcoming). "Arab (Bedouin) Urban Communities in Lebanon: Challenging Traditional Understandings of 'Tribe' in the Light of

Historical and Sociolinguistic Data." In *Interdisciplinary Perspectives on the 'Tribe' in the Middle East and North Africa*, edited by L. Stocker, *Welten des Islams*. Berlin, Boston: De Gruyter.

The Arabic dialects of northern Oman: some key features of the lexicon

Clive Holes, Magdalen College, Oxford

The Arabic dialects of Northern Oman are now much better described than hitherto. New field data has enabled us to place them more accurately in the Arabian dialectal context, and also suggests historical links they may have with Arabic dialects outside Arabia and far removed from them in both space and time. Research into Omani and indeed Gulf Arabic more broadly has focussed hitherto almost exclusively on the distribution of certain phonological and morphosyntactic variables, and there has been little analytical work done on the lexicon. This is understandable, as by its nature lexical analysis requires large data bases if any geographically or socially representative patterns are to be accurately discerned.

This paper is a first step down the road of describing the distribution of certain key lexical elements of the northern Omani Arabic dialects as represented in a large data base of natural spoken data gathered (1) in recordings made in the course of fieldwork done by the author between 1985 and 1987 when he was resident in Oman, and (2) through the analysis of approximately 100 hours of videoed interviews with elderly Omanis first broadcast on Omani public TV around ten years ago in a number of TV series devoted to recording memories of the older generation predating 1970, and recently uploaded en masse to Youtube.

The data base of contextualised Omani sentence examples that has resulted is still unfinished (in fact has no logical endpoint), but currently runs to over 700 A4 pages, and is based on the speech of 170 individuals (approx 85% men) from 75 different locations in northern Oman. Every piece of speech data in this corpus is coded for the location in which it was recorded, which facilitates data searches in addition to those routinely available in Word. The intention is to publish this glossary (late 2025/2026) in the Cambridge Semitic Languages and Cultures Series published by Open Book, downloadable from the website free of charge.

This paper presents data from half-a-dozen or so common verbs which have a modal or aspectual function and describes their distribution in terms of the Omani dialect typology already established by previous studies of phonology and morphosyntax. Some comment will also be offered on the occurrence/ non-occurrence of these verbs in neighbouring Gulf dialects.

Bifurcation

Jonathan Owens, University of Bayreuth

In a given geographical area it is an easy matter to define opposing varieties of Arabic well profiled for differing linguistic attributes. Such dialectal differences span all manner of social life, from compact urban areas such as Baghdad to larger spaces such as eastern vs. western Libyan Arabic. I propose rebranding these contrastive dialectal areas in terms of "bifurcation". A bifurcated feature is a variable distributed within different segments of a speech community. It has inherently sociolinguistic ramifications, but is extracted from its sociolinguistic underpinnings and projected on to long term historical stability or change of variables. By definition it is a socio-historical construct.

Bifurcation is described in terms of five general attributes

- Speech community
- Linguistic attributes
    - Paradigmatically constrained
    - Open ended
- transfer
- focus-ambience
- a socio-historical concept

These categories will (time allowing) be illustrated in describing:

- Bahrain
- Baghdadi Arabic of the 1960's
- West Sudanic Arabic
- The -k perfect of Yemen
- (Sabaic perfect)

The study proceeds from contemporary speech communities to increasingly distant diachronic destinations, from speech communities whose communal attributes and historical origins can be relatively richly documented, to those circularly definable largely by the linguistic attributes which they display.

Bifurcation is definitionally an historical linguistic concept, which, however, complements rather than displaces traditional instruments for understanding Arabic language history. As will emerge in the course of the presentation, it underscores the idea that Arabic is well served by tailoring linguistic

concepts to elucidate the interpretive interests of the language and in so doing contributes directly to defining broader issues in historical linguistics.

Morphosonority and indexical sociohistorical linguistics

Jonathan Owens, University of Bayreuth & Ajid Saleh Lawan, University of Maiduguri

This paper describes the development of a new conjugational class in a dialect of Arabic, sensitive to the sonority structure of a sub-class of verb stems. The morphophonological change is triggered by a subject suffix and is realized by the stressed stem vowel /á/ (e.g. bi-[šár]-u 'they buy'). The change emerges from a complex chain implicating:

1. The widespread role of sonority in LCA (Lake Chad Arabic) syllable structure
2. The shift of pharyngeals to laryngeals (*ʕ/ħ → ʔ/h) in LCA
3. The variable loss of final laryngeals which, however, leaves behind a CC trace whose residual effects are manifest in stress
4. The sonority structure of a class of weak final verbs
5. The emergence of a distinctive "morphosonority template" {[CáCson]-Vsbj}out of factors (1-4), but based on an inherited /aClaryngeal/ propensity going back to Old Arabic

The dialect in question shares (1-4) with a wider dialectology of the LCA area, but on the basis of a sociolinguistic trend study it is shown that (5) is a unique development in Arabic. The development of the unusual morphosonority is argued to represent diachronic dialect vitality and index a relatively closed speech community in which the change could be nurtured.

The emergence of the šāwi and (rural) gilit dialect areas in the light of historical sociolinguistics

Stephan Procházka, University of Vienna

In Arabic dialectology, the term '*šāwi*-Arabic' refers to a bundle of closely related dialects spoken in various regions of the Fertile Crescent, particularly in Syria and Jordan, but also in southeastern Turkey and Lebanon. However, typologically similar dialects are also spoken in many rural parts of Iraq, which makes it reasonable to group the *šāwi* dialects and the rural Iraqi *gilit*-type dialects

together. This dialect continuum has been labelled as “Syro-Mesopotamian fringe dialects” or, more recently, “Dialects of the Syrian Steppe and adjacent regions” (henceforth SySt AR).

Although Arabicization of the region in question started in Late Antiquity, i.e. centuries before the emergence of Islam, the linguistic ancestors of today’s dialects probably began to migrate to the Syrian Steppe and the adjacent Jazeera in the late 10th century when the region witnessed significant ethnic changes (above all due to the migrations of Turkic and Kurdish tribes). In Iraq, large-scale re-Bedouinisation of the countryside happened later, in the aftermath of the Mongol conquests of the 13th century. However, many tribes were always on the move and migration to and from as well as inside the region itself frequently happened up to the 20th century.

In the light of this evidence, we can assume a relatively high degree of dialect contacts, both between different Bedouin-type dialects and between Bedouin-type dialects and Sedentary-type dialects—in the east with *qəltu*-dialects, in the west with Levantine dialects, in the centre with the traditional dialects spoken in the oases of the Syrian Steppe (e.g. Palmyra, Soukhne). Additionally, this region witnessed long-standing contact with other languages, particularly with Turkish and Kurdish.

The paper will shed light on the emergence of SySt AR Arabic by analysing an array of linguistic features which are regarded to be typical of at least a larger number of sub-dialects. We start from the hypothesis that there are:

(1) Inherited features — suggested features for analysis here are: the raising of low vowel /a/, *ġ > q, *gahawa* syndrome, nominal linker, retention of feminine *-t* in attributive phrases, 1st person pronouns, imperfective inflection suffixes, future and lexical features

(2) Internal innovations — suggested features for analysis here are: the genitive particle/exponent, perfective inflectional suffix 3pl.m, interrogative adverbs

(3) Innovations through contact —suggested features here are: affrication of *g and *k, resyllabification of *CaCaCv, indefinite marker /fard/, *b*-marker for present, repetition with *m-*, emphatic imperative prefix *d-,* and lexical calques.

Sociolinguistic factors may have played a role mainly in contacts between nomadic and sedentary populations as they differed not only in their lifestyles, but also in prestige, military power, economy and access to education. Noteworthy are also the hierarchical and social differences among the Bedouins themselves, especially between the camel breeder tribal and the sheep and goat breeder tribes.

The whys and hows of short-term accommodation in a multilingual and multidialectal context, and the way forward

Valentina Serreli, Sapienza University of Rome

The presentation will outline the findings of a sociolinguistic study that has been designed to observe the behaviour of local and supralocal Arabic variants in a context that is both multilingual (Amazigh-Arabic) and multidialectal (Shahibi-Cairene Arabic). Drawing from group conversations among Shahibi speakers and Siwi speaker aged between 18 and 70 years, phonological, morphological, and lexical variables have been statistically analysed. The texts collected display heterogeneity at all levels of grammar and lexicon and discourse. All in all, the study demonstrated the vitality of the local (Shahibi) Arabic and the large diffusion of supra-local (Cairene) Arabic features and points to traditional patterns of social variability. The patterns of microvariation observed reflect broad categories such as Siwi minority status, oasis remoteness, ethnolinguistic group cohesion, as well as local and national language ideologies.

Exploring Arabic dialect relationships across Africa through dialectometry: Insights and limitations of this approach

Carolina Zucchi, University of Bayreuth

Arabic dialects spread rapidly across Africa as a consequence of the early Islamic conquests, reaching Egypt by 639, present-day Tunisia by 670, Tangier by 708, the Iberian Peninsula by 711, Sicily by 827, and as far as Mauritania by the 15th century. Arabic expanded into the Sudanic region from Upper Egypt at various stages, with some sources indicating that Arabic-speaking communities had already moved west from there to reach the Lake Chad area by the late 14th century (Owens 2003: 721). All major Arabic dialect areas in Africa originated from migrations of speech communities that initially settled in Egypt following the Arab conquest, making Africa a crucial domain for studying dialect contact and development.

Reconstructing the historical development of these dialects is complex, requiring consideration not only of initial settlements, but also of later migrations (e.g. the 11th-century Banū Hilāl movements, whose impact on North African Arabic remains debated, see Benkato 2019), sociolinguistic factors (e.g. Jewish-Muslim dialect coexistence within the same North African city, prestige-driven linguistic shifts), and contact with both neighboring languages and among different dialects. Isoglosses do not

bundle cohesively, complicating classification. For instance, the only linguistic trait considered to be characteristic of Maghrebi dialects is the presence of 1SG n- and 1PL n-...-u imperfect affixes (Vicente 2008: 39, 41), but this isogloss also appears in some Egyptian and Sudanic dialects. Behnstedt and Woidich (2005: 132) note that if this feature alone defined Maghrebi, some Egyptian dialects would qualify, which is counter-intuitive given their other characteristics. Similarly, the retention of a plural gender distinction is rare but scattered across all major dialect areas.

One step towards understanding these dialects' relationships is visualising their linguistic features collectively across geographical space. Dialectometry, a quantitative method based on the measurement of dialect distances developed in European linguistics (Séguy 1971, 1973; Goebl 1984), by using both cartographic and statistical tools, has the potential to reveal patterns that may be imperceptible using traditional methods. It involves three core steps: (1) compiling a dataset where dialects are assigned values for linguistic features (e.g., relative to the feature “reflex of *q”, the Ḥassānīya dialect of Mauritania would be assigned value "g" and Cairene "ʔ"); (2) applying a metric to compute linguistic distances between dialects, resulting in a matrix of similarity scores, and (3) visualising these relationships using statistical and digital mapping tools.

Traditionally, dialectometry has been applied to dense datasets in relatively small areas, usually within a single country or region (e.g., the Netherlands, Switzerland, Japan). In Arabic dialectology, some of its methodological tools have been used by Behnstedt and Woidich (2005: 106-135) for Egypt and the Levant and by De Jong (2011: 329-336) for Sinai. This study expands its scope across a 5,500 km area (from Egypt and the Sudan to the east to Mauritania to the west) to explore the potential of this method to detect whether dialects separated by vast distances exhibit significant linguistic similarities, which could suggest shared ancestry or contact.

In this talk, I will explore the extent to which dialectometric methods can provide insights into the historical evolution of Arabic in Africa, while also discussing their limitations. I apply such methods to a dataset of 100 linguistic features—including phonetics, phonology, morphosyntactic categories and particles—compiled based on published sources and covering 52 Arabic dialects across Egypt, North Africa, and the Sudanic/Lake Chad region. Using multidimensional scaling (MDS) and clustering techniques, I examine how dialect clusters emerge and compare them with traditional classifications. Among other findings, I discuss how different Egyptian oases dialects align more closely with Maghrebi or Sudanic varieties, how Upper Egyptian dialects relate to those of the Sudanic region, and how the Douz, Eastern Libyan and Western Egyptian Bedouin dialects form a distinct cluster. Finally, I evaluate how these results contribute to our understanding of dialect history and assess the broader potential of dialectometry in Arabic dialectology.

References

Behnstedt, Peter and Manfred Woidich. 2005. *Arabische Dialektgeographie: Eine Einführung*. Leiden: Brill.

Benkato, Adam. 2019. "From Medieval Tribes to Modern Dialects: on the Afterlives of Colonial Knowledge in Arabic Dialectology." *Philological Encounters* 4(1–2). 2–25.

De Jong, Rudolf. 2011. *A Grammar of the Bedouin Dialects of Central and Southern Sinai*. Leiden: Brill.

Goebl, Hans. 1984. *Dialektometrische Studien: anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*. Tübingen: Niemeyer.

Mackintosh-Smith, Tim. 2019. *Arabs: A 3000-year History of Peoples, Tribes and Empires*. New Haven; London: Yale University Press.

Owens, Jonathan. 2003. "Arabic Dialect History and Historical Linguistic Mythology". Journal of the American Oriental Society, Vol.123(4): 715–740.

Séguy, Jean. 1971. "La relation entre la distance spatiale et la distance lexicale." *Revue de Linguistique Romane*, 35: 335–57.

Séguy, Jean. 1973. "La dialectométrie dans l'Atlas linguistique de la Gascogne." *Revue de Linguistique Romane*, 37: 1–24.

Vicente, Ángeles. 2008. "Génesis y Clasificación de los Dialectos Neoárabes", in Federico Corriente, Ángeles Vicente and Farida Abu Haidar (eds.). *Manual de Dialectología Neoárabe*. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo.

CorpusCompass

Workshop Session : CorpusCompass

Muhadj Adnan & Nicolò Brandizzi, Independent researchers

This workshop introduces CorpusCompass, an open-source tool for optimizing data extraction, dataset generation, and annotation quality in Corpus Linguistics. It is particularly valuable for studies of linguistic variation and under-resourced languages, enabling transformation of annotated corpora into structured datasets for statistical analysis.

The session will cover the tool's foundational context, a technical overview, and a guided practical application. Attendees are encouraged to independently select their own data, define their annotation styles and features, and experiment with CorpusCompass prior to the workshop. This preparation will allow the workshop to serve as a focused session for addressing specific questions in a dedicated Q&A session. Participants will gain practical knowledge on how to integrate this software into their linguistic research, thereby enhancing their corpus linguistics workflows and improving annotation accuracy. Furthermore, they will learn to approach quantitative methodology with greater confidence and accelerate their linguistic analyses.

As this marks the inaugural public release of CorpusCompass, we highly value constructive feedback.

* * * * *

Organised by Valentina Serreli and Jonathan Owens
Funded by the University of Bayreuth