Mapping Linguistic Epistemicity

General information

NWO Vici grant, number VI.C.231-014
1/9/2024 – 31/8/2029
PI Jenneke van der Wal
Hosted at Leiden University

Summary

When people communicate, they not only exchange information, but they also indicate how the speaker's and addressee's knowledge relate to that information. What is the source of the information, how certain is the speaker of it, is it unexpected for the speaker or addressee? Interestingly, the linguistic strategies to express these aspects are typically multifunctional, indicating fuzzy boundaries between the concepts of source, certainty etc. For example, saying 'They DID eat it' can indicate a contrast with the addressee thinking they didn't, but also that this is surprising, and that the speaker is certain. All the concepts related to the speaker's and addressee's knowledge are thus interconnected in our mind: together they form one conceptual space. How is that space organised? And is it the same for speakers of all languages?

While previous analyses unfruitfully focus on rigid categories (such as 'contrast' and 'evidentiality'), the current project takes advantage of the fuzzy boundaries to create a network map of this *intersubjective epistemicity*. We will investigate a number of African languages, which have not had a chance to contribute to this area, in collaboration with native speaker linguists, and making use of virtual reality stimuli. For these languages, we study how they express the detailed aspects of the speaker's and addressee's knowledge in relation to the information. Given that concepts that are expressed together must be more closely related than those that are not, the resulting data can be used for a proximity network analysis. This then forms a map that shows how the human mind organises this conceptual space, and by comparing crosslinguistically, we can discover to what extent our conceptual organisation is the same or influenced by the language(s) we speak.

Research proposal 'How languages show what speakers know'

(This is a large part of the submitted research proposal, with updated references)

1. The conceptual space of intersubjective knowledge

"Speakers of any language continuously keep track of what others know and how their own knowledge can be related to the knowledge of others" (Bergqvist & Kittilä 2020: 12). In communication, therefore, we are not only exchanging factual information, but we also communicate how each of us relates to that information, and how the information should be added to our knowledge base. We indicate, for example, the source of the information, its certainty, a contrast with previous information, or whether it is surprising. These aspects of intersubjective knowledge frequently overlap: in saying 'I heard that Sydney went paragliding', I indicate that I have indirect evidence, but also potentially that the information is not certain. Similarly, the sentence 'They DID finish it!' indicates that there is a contrast with previously thinking that they didn't, but can also indicate that this is surprising. It is difficult, therefore, to tease these aspects of meaning apart.

The fact that concepts such as 'source of information' or 'contrast' have fuzzy boundaries with the concepts of 'certainty' and 'surprise' indicates that they form part of the same conceptual space (i.e., the mental organisation and representation of concepts, cf. Aikhenvald's 2021 'Web of knowledge'). What does that space look like? And is it the same regardless of which language you speak or does language influence how these concepts are organised? In the proposed project, we chart the map of the conceptual space of how the speaker's and addressee's knowledge relate to the content of the utterance by investigating a number of African languages. Such a map provides a way to understand how different concepts are related to each other and how they are mentally arranged. In this way, we discover not only the intimate details of a part of how the human mind organises intersubjective knowledge, but also to what extent this organisation is influenced by the language you speak.

2. Language as a gateway to the conceptual space

How can we access this subconscious conceptual space? Croft (2001: 364) suggests that "the conceptual space is the geography of the human mind, which can be read in the facts of the world's languages in a way that the most advanced brain scanning techniques cannot even offer us". We should therefore investigate how languages express the aspects of **intersubjective epistemicity**, defined as the way that the speaker's and addressee's knowledge relate to the utterance content (henceforth **ISE**, comparable to Behrens' 2012 'metadiscourse' or Bergqvist's 2017 'epistemic perspective'). As an example of the expression of ISE, in Ikoma, a Bantu language of Tanzania, there is a choice of two forms in the perfective conjugation, depending on whether you have direct evidence or indirect evidence. If your information is based on directly seeing or hearing it yourself, as in (1a), the verb is marked by *-ka-*, whereas if you have indirect evidence, you use the form marked by *-iri*, as in (1b).

Ikoma (Roth 2018: 89, glosses adapted)

(1) Context: A child is taking a nap in the bedroom of a house.

a. A-**ká**-βook-a.
1SM-DIR.EV-wake.up-FV
'S/he is waking up.' / S/he has just woken up.' (speaker can see/hear the child)

b. N-a-βook-**iri**.
FOC-1SM-wake.up-PFV
'S/he has just woken up.' / 'S/he is awake.' (speaker cannot see or hear the child)

The same is found in Dutch for the modal verb *moeten* 'must' in (2), indicating indirect evidence in this context:

Dutch (de Haan 2000: 74)

(2) De film **moet** uitstekend zijn.
DET.DEF film must excellent be
'The film is said to be excellent.'

Various categories have been proposed over the years to capture ISE concepts in languages that have specific markers for them (this is a non-exhaustive list, and definitions differ between authors):

1. Evidentiality: what is the source of the information?
2. Engagement: who has access to the information?
3. Egophoricity: who has authority over the information?
4. Epistemic modality: how probable/certain is the information?
5. Mirativity: is the information surprising?
6. Givenness: is the information active in the interlocutors' mind? ⎫ Information
7. Contrast: is there a contrast with alternative information? ⎭ structure

For each of these categories, there is a hot debate on whether it forms its own category (e.g., de Haan 1999; Speas 2010; Aikhenvald 2015, 2021), and if so, what then defines the boundaries of each category. This debate is not productive in furthering our understanding of how interlocutors organise their knowledge, for two reasons. The first is that by focusing on the core categories, many aspects of meaning and use are missed ('if all you have is a hammer**,** everything looks like a nail', see Matić & Wedgwood 2013, Bergqvist & Grzech 2023). When setting out to document information structure, for example, an out-of-the-blue sentence will be interpreted as 'thetic' (the information structured presented as one piece), missing out on the fact that it could also encode surprise (mirativity). Consequently, there is a high risk of misdiagnosing and underdiagnosing linguistic strategies and functions, which adversely affect theoretical modelling of human language. The second reason that the focus on categories is unfruitful is that a universal definition of a category does not do justice to the rich variation between languages. Very little is to be gained from a debate about the (non-)existence of a category if the parties base themselves on different languages – an illustrative example is the exchange between DeLancey 1997 (entitled 'Mirativity: The grammatical marking of unexpected information'), Hill 2012 ('Mirativity doesn't exist'), and Hengeveld & Olbertz 2012 ('Didn't you know? Mirativity does exist!').

The fact that the relationship between these notions remains debated and/or elusive (Grzech et al. 2020: 286) is in large part due to the fuzziness or multifunctionality that linguistic strategies tend to have in the area of ISE. To illustrate, the particle *kamaŋ* (3) in the Nilo-Saharan language Fur (spoken in Sudan) indicates that the speaker has witnessed an event (evidentiality), as well as unexpectedness for the addressee (mirativity).

Fur (Waag 2010: 260, glosses adapted)

(3) D-íí-ŋ bára **kamaŋ** ʔɛ́la.
SG-2SG-GEN brother EV 3SG.come.PFV
'Your brother has really come.' [I have seen him]

Another example of overlapping functions comes from !Xun, spoken in Namibia. König (2013: 79) states that the marker *cālā* "belongs to the domain of focus", but at the same time it fits in the evidential system as a marker of indirect knowledge (a non-firsthand evidential), and it can indicate uncertainty as well as "contrast with the expectation of the addressee", as illustrated in (4).[1]

[1] König (2013: 79) notes that "the question marker can be seen as an indicator of the focal status that the statement gets with the non-firsthand evidential: the question marker is simply triggered by the non-firsthand evidential".

!Xun (König 2013: 80, glosses adapted)

(4) Hà á **cālā** kē ú |ˈàn hà.
PRO Q NONFIRSTH.EV PST go with PRO
'He is said to have left together with her (but I doubt it).'

For these two examples (and countless others), a debate can be held about whether they are actually evidential markers or are better treated as mirative/focus/epistemic markers, and how broad or narrow those categories should be defined.

Rather than continuing this unfruitful debate, it is becoming clear that we have to look **beyond the categories** as discrete entities. For all subparts of the ISE area, recent studies suggest that they are related and should be studied in conjunction: Masia (2022) argues that information structure overlaps with evidentiality, Yliniemi (2023) shows the same for contrastive focus and mirativity, Boye (2012) studies epistemic modality and evidentiality together, and the most encompassing work is probably Aikhenvald (2021), showing the relations between epistemic modality, evidentiality, mirativity, and egophoricity (see also Cornillie 2009, Behrens 2012, García Macías 2016, Grzech et al. 2020, Ozerov 2018, Sandman & Grzech 2022). In all these studies on the fuzzy boundaries of different ISE categories, the idea that all belong to a single conceptual space emerges as a common theme, yet the full space has not been studied as such – partly because data on African languages are lacking and have not had a chance to contribute to the debate.

It is now time to capitalise on this idea, using the fuzziness between the categories to our advantage. By studying where the expression of "categories" overlaps, we can determine which aspects of ISE are more closely related and which are more distinct. This is based on the connectivity hypothesis: "any relevant language-specific and construction-specific category should map onto a *connected region* in conceptual space" (Croft 2001: 96, see also Haspelmath 2003). Therefore, by systematically checking which aspects of meaning are jointly expressed by a particular linguistic strategy, we deduce that those aspects must be closer together in the conceptual space than aspects of meaning that are not part of the interpretational range of the expression. For example, if the concept 'hearsay evidence' is expressed by the same linguistic strategy as 'speaker uncertainty', they must be closer to each other than the concept 'contrasting with addressee's knowledge' if that is expressed by a different strategy. This information can then be plotted, forming a proximity map (Croft & Poole 2008), illustrated for lexical data in .

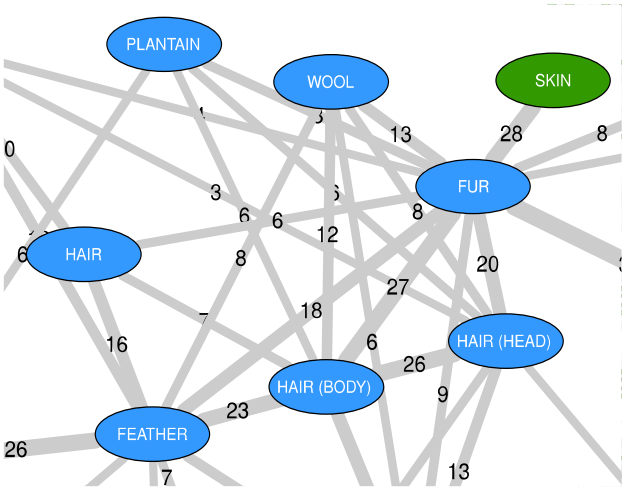

Figure 1 – Detail of colexification clusters (CLICS3, Rzymski et al. 2020: 6)

The nodes (or 'vertices') on a proximity map represent universal concepts, such as HAIR and FEATHER in the lexical map in . In the case of the proposed project, nodes could be DIRECT VISUAL EVIDENCE and UNPREPARED MIND OF THE ADDRESSEE, for example. The *connections* (or 'edges') between the nodes represent whether those concepts are connected, as evidenced by their being expressed together or distinguished in different languages. By investigating which groupings of concepts exist cross-linguistically, we can see whether concepts are frequently expressed together or not, and thereby deduce how close or distant those concepts must be, represented by the relative distance and strength, indicated in numbers in Figure 1.

3. Research questions

The main questions for this research project are therefore:

M1 How is the conceptual space of intersubjective epistemicity organised?
M2 Is this organisation independent of which language one speaks?

To find the answers to these questions, we investigate the following subquestions:

S1 Which nodes (core concepts) in the space of intersubjective epistemicity form the basis of investigation?
S2 How are these nodes grouped by linguistic strategies?
S3 What does the grouping tell us about the relations between the nodes, intra- and crosslinguistically?
S4 What do the relations between the nodes reveal about our linguistic ability, our conceptual organisation, and the possible interaction between them?

The following sections explain how each subquestion (S1-S4) will be addressed.

3.1. Which nodes in the ISE space form the basis of investigation? (S1)

As the study is oriented from function to form (asking for particular concepts how they are expressed in languages), we first need to establish which precise concepts in the ISE space will be taken into account. These concepts must be language-independent and cross-linguistically applicable, or in other words, we need an *etic grid* (Levinson & Meira 2003: 487; alternatively called *comparative concepts* by Haspelmath 2010, or *analytical primitives* by Cysouw 2007).

In order to determine this etic grid, we need to again go beyond the given categories (evidentiality, mirativity, etc.) but in a different way, namely to look at more detailed aspects of interpretation. For example, we want to talk specifically about 'unprepared mind of the speaker' and not about 'mirative' in general. In a recent talk (10/07/21), Alexandra Aikhenvald says: "I think it's all about disentangling lots of little kinds of domains that grammars of languages (especially minority languages, not very well described languages all over the world) express, and there is a lot of overlap". We have already seen the overlap, and Aikhenvald (2012) unravels five notions within mirativity: (i) sudden discovery/revelation/realization, (ii) surprise, (iii) unprepared mind, (iv) counter-expectation and (v) new information. García Macías (2016) also illustrates this in his typological survey: the notion of 'theticity' (information structured as one piece) actually breaks down into two subtypes (presentational and existential, cf. Sasse 1996, 2006), and languages further distinguish between *unexpected* events and *misexpected* events. For information structure (with the key terms topic, contrast, focus), Ozerov (2018, 2021) equally makes a plea to investigate the more detailed aspects and discourse motivations underlying the posited categories. The same notional breakdown is seen for evidentiality, with epistemic authority/ownership/primacy, access, and source playing a role, and at least 6 types of semantic parameters being described within the latter category of source (Aikhenvald 2004). We thus see that each of the areas shows a need for further detail, and such detailed aspects form the nodes of the conceptual space of ISE.

By scrupulously perusing the relevant literature, we can harvest as many notions as possible to form an initial etic grid – a task that is already partly done: I have extracted some 42 detailed notions from the literature on 38 languages, e.g. 'reactivation', 'speculative evidence', 'above addressee's expectation', 'privileged speaker access to knowledge'. It is inevitable that the etic grid will change with further investigation in the project, but, as Haspelmath (2003: 231) indicates, "the typical experience is that after a dozen languages have been examined, fewer and fewer functions need to be added to the map with each new language".

In this way, the project extends the domain of investigation beyond the work that has produced semantic maps for the adjacent areas of modality and tense/aspect (see e.g. Anderson 1982 on tense/aspect; Van der Auwera & Plungian 1998, Van der Auwera et al. 2009 on modality). We thus

take the conceptual mapping and notional grid into the realm of *epistemic and attentional management*, that is, providing information about the *knowledge states* of the speaker and addressee, without widening the scope to include their social interpersonal relation (as relevant for honorifics, speaker stance and evaluation), to keep the project manageable.

3.2. How are the nodes grouped by linguistic strategies? (S2)

In order to answer the second subquestion S2, relevant data are needed from a number of languages. How will we obtain the relevant data, and which languages will be studied? The methodology and two data-gathering subprojects are explained below, one on ISE in 6 languages and one on expressing and highlighting truth in Cinyungwe.

3.2.1. How do we obtain the relevant data? (Methodology)

Intersubjective epistemicity crucially involves the speaker and addressee in their conversational negotiation of information, therefore being highly context-dependent (e.g. Bergqvist 2017, Bergqvist & Grzech 2023). The investigation of ISE should therefore always be grounded in contextualised data, important to avoid misdiagnosis. Grzech et al. (2020: 298) postulate that "the type of elicitation techniques that can deliver the best results in epistemic fieldwork are tools that elicit 'staged communicative events'".

Practically, even though each "category" (evidentiality, mirativity, contrast etc.) has proven complex to investigate in a targeted manner, recent developments and innovations in methodology make such research much more feasible now:

- The BaSIS project has developed an extensive methodology (available here) for the study of information structure (building on Van der Wal 2016 and Skopeteas et al. 2006), including more natural visual stimuli, which has proven to work for comparative purposes (Kerr et al. 2023, Kerr & Van der Wal 2023).
- The volume by Grzech et al. (2020) gathers and develops diagnostics for documenting and testing a range of notions in the domain of epistemicity (evidentiality, egophoricity, epistemic authority and engagement).
- Peterson (2017) provides a set of tests that can be used to probe (inherent and implied) mirative meanings.
- Zeisler (2016) provides a questionnaire on evidentiality, inferentiality, and speaker's attitude.
- Our successful pilot using **virtual reality** (VR) shows that virtual environments can be employed productively for fieldwork (González et al. 2024). In traditional linguistic fieldwork, translations are elicited, or a reaction is requested to a verbally or pictorially presented situation. But these stimuli can only provide part of the context, as we can never fully control what the speaker may imagine in addition to the given stimulus. We can get much closer to a full and natural context if we present speakers with a 360° audiovisual environment through VR. Such an environment stimulates natural speech while still controlling carefully for the linguistic factors to be studied. This is particularly relevant for systematically eliciting aspects of ISE, e.g. presenting contrasting or surprising information, or presenting evidence directly or indirectly: with VR the speakers feel as if they are direct witnesses, our pilot shows. Trial videos in a natural East-African context have already been recorded in Uganda, for example a set of two videos, one with someone accidentally dropping a tray with a cup of tea (direct evidence), and one with people finding a shattered cup on the floor (indirect evidence). Asking participants what happened might result in different descriptions for the two VR experiences.

Figure 2 – Tsonga speaker participating in our VR study

These resources will be extended and appended in the first phase of the project, building on my experience in intensive collaboration with native speaker linguists. The resulting methodology approaches the field in a novel way, firstly because it targets not just 'engagement' or 'contrast'

as an individual category, but the *whole* functional area of ISE (evidentiality – engagement – egophoricity - epistemic modality – mirativity – givenness –contrast); and second, because it targets the more detailed and contextualised aspects of interpretation (the nodes as explained under S1 in §3.1). This provides an important way out of the current fixation on these larger categories, shedding light on the smaller ingredients that actually determine ISE strategies, and inductively establishing the relations between these.

The resulting data will consist of form-function pairs. For these pairs, a distinction has been made between dedicated markers for a certain meaning vs. co-opting a strategy that has another core meaning. For example, Aikhenvald (2004) differentiates between grammaticalised, obligatory 'evidentials' and non-obligatory, pragmatically conditioned 'evidential strategies'; and Peterson (2017) discusses 'parasitic' and 'non-parasitic' mirativity. To give a concrete example, in Lubukusu (spoken in Kenya), the object marker on the verb (*ka-* in example (5)) can optionally cooccur with the object noun phrase (*kamalwa* 'beer' in (5)), resulting in the range of interpretations indicated for example (5) – but it cannot be used as a statement out of the blue.

Lubukusu (Lippard et al. 2024)

(5) Wekesa a-**ka**-nyw-a **kamalwa**.
1.Wekesa 1SM.PST-6OM-drink-FV 6.beer
'Wekesa drank beer.'
Available interpretations:
- The sentence is an emphatic confirmation of the act. (verum)
- It is a well-known fact that Wekesa does not drink, so seeing him drink is surprising. (mirative surprise)
- It is a well-known fact that Wekesa is a drunkard; any doubt is met with this sentence. (mirative reprimand)
- The sentence expresses the sheer amount of alcohol consumed. (intensity)

One possible option for analysis (that we will not follow) would be to posit one core function, for example 'givenness of the object', suggesting that the other interpretations are parasitic on this interpretation, depending on the context. In the current endeavour, however, we analyse object marking here as a multifunctional strategy for verum, speaker surprise, reprimand, and intensity. The motivation is that, first, as part of the discovery procedure we cannot *a priori* distinguish core vs. peripheral meaning; second, the notion of obligatoriness is problematic (e.g., only syntactic obligatoriness or also discourse-wise?); third, the "parasitic" evidential/mirative/focus etc. strategies may over time develop into "true" evidential/mirative/focus markers (Willett 1988, among others); and fourth, it is not arbitrary which grammatical strategies are co-opted for ISE: only concepts that are 'in the neighbourhood' qualify for it (according to the connectivity hypothesis). This makes it worthwhile studying all strategies regardless of whether they are considered to be "true" or "parasitic" (to the extent that these can be distinguished; see also Mushin 2013): both types are informative for establishing which concepts are closer, regardless of further possible claims on core meaning or pragmatic influence.

In the form-function pairs, the linguistic strategies to express ISE are somewhere on a continuum between lexical and grammatical. Lexical expressions of ISE include for example periphrasis, as in 'I am certain that…' (epistemic modality), 'Unlike what you think, …' (contrast), or 'I heard that…' (evidentiality). The lexical expression of ISE is very likely universal – every language will have some lexical way to describe the interlocutors' states of knowledge – and is therefore not as interesting for the current project. On the other hand, languages differ in the grammatical expression of ISE, defined as providing secondary, non-at-issue information (Boye & Harder 2009, 2012). Under this definition, not only complementisers or clitics are considered, but certain adverbs may also have a grammatical function. The proposed project thus targets the grammatical expression of ISE, tapping into the subconscious intuitions of the human language

faculty. A further reason to focus on the grammatical expression of ISE is that this will inform us on which features of ISE actually enter from the pragmatics into the syntax (see S4 in §3.4).

The project methodology will be made available not only in written form (and open access), but also through an online masterclass, as described below.

3.2.2. Which languages?

As Evans et al. (2018a: 110) indicate, "Human language offers rich ways to track, compare, and engage the attentional and epistemic states of interlocutors. While this task is central to everyday communication, our knowledge of the cross-linguistic grammatical means that target such intersubjective coordination has remained basic". This means that there is an exciting and necessary task that this project will take up. The project will consider a number of languages from Africa, and consists of two lines for data-gathering, as presented next.

3.2.2.1. Intersubjective epistemicity in Africa

The Himalayas and South America are relatively well studied as hotbeds for evidentiality, egophoricity, mirativity, and engagement, but there is very little work on African languages in these domains. The fact that researchers are actively studying the Himalayan and South-American but not African languages could lead to Africa being pushed to the fringes as the research in this area develops, with a skewed typological basis for comparative research as a consequence. To ensure that we are not missing out on the insights from African languages, the proposed project will investigate the expression of ISE in 7 African languages from different language families on the continent.

Although the continent has been overlooked for many of the ISE aspects, recent reports on various African languages show that there is a world of knowledge to be gained. For evidentiality, the following quote is typical for the general image of the continent: "It appears that expressing evidentiality as a verbal affix or clitic is the most common strategy. With the exception of Africa, it occurs on every continent" (De Haan 2013). Storch (2018: 610) however adds that "It seems as if the contrary is true and that precisely this strategy is also present in a number of African languages". This also became apparent in the workshop on mirativity and evidentiality in October 2021: "ongoing semantic research in Bantu languages continues to uncover systems that are primarily evidential in their semantics, as well as other grammatical categories that can be exploited secondarily to express evidential distinctions" (Crane et al. 2024).[2] As Dimmendaal (2014) and Storch & Coly (2014) demonstrate with their chapters on *the grammar of knowledge* in the languages Tima (Sudan) and Maaka (Nigeria): it is there – we just haven't been looking for it. The same goes for mirativity, which is only now being properly recognised in various descriptive analyses for such diverse strategies as object marking, sentence-final particles, and complementisers, see e.g. Asiimwe (2024), Jordanoska (2020), König (2013), Lippard et al. (2024), Sikuku, Diercks, and Marlo (2018), Sikuku & Diercks (2021), Storch (2009). Aspects of information structure (contrast, givenness) have received more attention in African languages, showing that information structure has a large influence on the grammar (see Amfo 2018, Ermisch 2009, Fiedler & Schwarz 2010, Güldemann et al. 2015, Van der Wal 2015 for general overviews). This makes African languages not just an obvious area for the proposed research, but also a necessary one for reasons of typological and theoretical balance (see further under S4 in §3.4).

For the expression of ISE in African languages, I expect the following areas in particular to reveal new insights (based on the literature referred to above):

- Demonstratives (cf. relation with deixis; e.g. De Haan 2005, Rooryck 2021)
- Personal pronouns (e.g. logophoric, Dimmendaal 2001, Speas 2004)
- Complementisers and quotatives (e.g. Güldemann 2002, 2008)

[2] See for such discoveries in the Bantu languages Bassong (2014), Botne (1997), Crane (2024), Crane et al. (2024), Gluckman & Bowler (2016), Kanijo (2020), Roth (2018); and for overviews for Africa, see Botne (2020) and Storch (2018).

- Discourse particles
- Verb conjugations

The proposed project will therefore widen and deepen our empirical knowledge about these languages (increasing the basic knowledge of crosslinguistic grammatical means for ISE, see Evans et al.'s quote above). This much-needed diversification of data also has the potential to change both the comparative-typological picture as well as the conceptual-theoretical development (see S4 in §3.4), preventing a skewed basis for both (cf. Ameka & Terkourafi 2019 on the importance of non-Western views and starting points). Targeting a local level, the project values deeper insights into fewer languages over superficial data on more, and even with a limited number of (genetically unrelated) languages we can answer the main research questions of how the conceptual space is organised and whether this is language-independent (M1+2). After this first step, we will be able to then address bigger questions of universality, and the comparison with other languages will be sought in a conference on ISE worldwide towards the end of the project.

3.2.2.2. Data-gathering line 1: Expression of ISE in 6 languages

Grzech et al. (2020: 297) remark that "Epistemic marking [in a broad sense - JW] is a prime example of a phenomenon where morphosyntactic resources employed in speech cannot be considered in isolation from 'the cultural knowledge, attitudes and practices of speakers' (Enfield 2002: 3). It follows that collaborative fieldwork models [...] are particularly well-suited for carrying out fieldwork on epistemic marking" (see also Mithun 2020). This means that the selection of languages in the proposed project must be based on both the language properties and the availability of native speaker linguists.

To select the optimal combinations of expert linguists and relevant languages, an open search process will be conducted for 6 collaborators on the project, rather than determining languages and linguists beforehand. While my network includes sufficient native speaker linguists, the open search and the masterclass (see below) will increase the chance of including languages that were previously under the radar but that show interesting properties in the expression of ISE, so that those languages will be able to contribute to the project (and thereby to linguistic typology and theory).

The 6 collaborators will be involved in the project on a task-based contract, with the tasks to create inventories of the linguistic strategies and the detailed aspects of ISE in their language (description and analysis), as well as to (co-)author articles on the analyses. A first training meeting in person (see S3 in §3.3) and regular online meetings will create a growth environment in which team members can learn from each other and develop further skills. In the case that too few suitable collaborators apply (which is mitigated already by the masterclass – see below), further collaborators will be invited through my extensive network, the African Linguistics School alumni, Afranaph collaborators, etc.

To bring collaborators on board, the project will develop an **online masterclass** about ISE. This will free three birds with one key: 1) preparation for teaching will ensure that the notions and elicitation materials are crystal clear, thus going hand in hand with the methodology development; 2) the availability of an online masterclass means that an introduction to the topic and the methodology will be available worldwide; 3) the participants in the masterclass are all potential collaborators, with the benefit that they will be motivated and already have the basic knowledge required for the task. The masterclass is aimed at advanced students of linguistics and professional linguists in Africa, and will be realised in collaboration with the online learning professionals at the Leiden Learning and Innovation Centre (LLInC) and the Humanities Faculty's centre ECOLe (Expertise Centre Online Learning). It will involve videos ('knowledge clips'), readings, and quizzes to help participants grasp the basics of ISE, as well as online meetings for interaction and critical reflection. The masterclass will thus contribute to a change in how linguists think about and practically approach ISE, it provides an opportunity for a new generation

of international researchers, and it will form part of the training of the research team (collaborators and PhD).

In addition to the collaborative data collection in this line, the project also dives deeper into one particular aspect where all areas of ISE seem to come together: marking of truth. This forms the second line of data gathering.

3.2.2.3. Data gathering line 2: The truth in Cinyungwe

The notion *verum* (Latin for 'truth') borders on practically all aspects of ISE. Stating that something is true (and/or emphasising its truth) can come with a range of extra information about the interlocutors' knowledge, as shown in Figure 3. Given how verum seems to connect aspects of ISE, it forms a particularly interesting case study to investigate in more detail for one language.

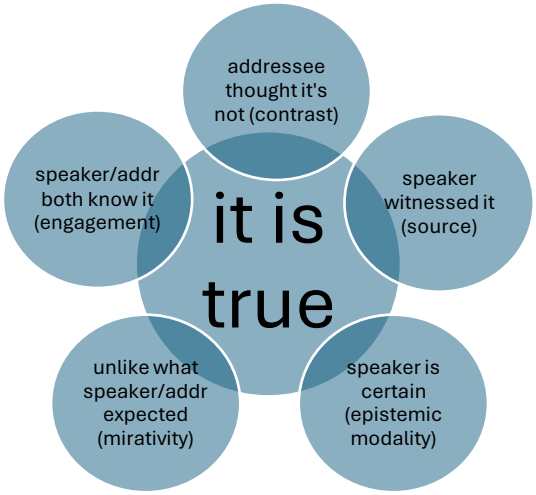

Figure 3 – Verum interfaces with aspects of ISE

The language selected for this strand is Cinyungwe, a Bantu language spoken in Mozambique. Cinyungwe is ideal for an in-depth investigation, as it shows a range of morphosyntactic strategies used in verum contexts (see Kerr & Van der Wal 2023). Three of these are object marker doubling (see Langa da Câmara et al. forthcoming), predicate doubling (6), and the particle =*di* (7). Strategies may also be combined, as indicated by the brackets in (6).

(6) Ku-**nemb**-a w-a-**nemb**-a(=di).
INF-write-FV 2SG.SM-PST-write-FV=VERUM
'You DID write.' (I didn't think you did) / 'You really wrote!' (more than expected)

(7) A: I am not convinced that he can sing. He told me he can. Have you ever heard him sing?
B: A-ni-yimb-a=**di**
1SM-PRS-sing-FV=VERUM
'He DOES sing!'
Interpretation: "I am confirming that he can really sing. I heard him singing; I am the witness."

We see here that one strategy can have multiple interpretations: predicate doubling in (6) shows both verum and mirative interpretations, and =*di* in (7) not only expresses verum but also a direct witness evidentiality.

The questions and hypotheses for Cinyungwe are thus as follows:

A. How do the strategies used in expressing verum differ in their interpretation and use?
Hypothesis: Each strategy covers a slightly different range of nodes/aspects; there is no full redundancy.

B. What is the effect of combining the strategies?
Hypothesis: By marking two or more aspects that are closely related at the same time, a paralinguistic effect of emphasis is brought about.

C. What does the expression of (various aspects of) verum tell us about the status of verum?
There is an ongoing debate on whether verum is an independent notion or part of information structure (see Gutzmann et al. 2020, Goodhue 2022, among others). In Kerr & Van der Wal (2023), we add hitherto undiscussed strategies based on new data from the Bantu languages, adding a new theoretical option. The proposed project takes the debate to a next level by approaching it within the larger space of ISE.
Hypothesis: Verum as an independent use-conditional notion (the contextual felicity

equivalent of truth-conditionality; Gutzmann 2015, Gutzmann et al. 2020) interacts with different aspects of the grammar to form language-individual categories and interpretations (see the Interactional Spine Hypothesis under S4 in §3.4).

The first data collection in this subproject will consist mostly of spontaneous interactional language use. Analysis of these data will result in more specific hypotheses and predictions, and these predictions can then be tested in the second data collection period using targeted VR stimuli. The VR stimuli to be developed for targeted elicitation will consist of 360° videos (as described in §3.2.1). As mentioned, a fully controlled virtual environment allows for testing individual linguistic factors while keeping the other factors constant, still encouraging semi-spontaneous speech. Like the languages studied by the collaborators (data-gathering line 1), the Cinyungwe data will also be used to create a conceptual map (see S3 in §3.3), but in addition this subproject provides a deeper insight into the expression of verum within the conceptual space of ISE.

To summarise the approach to S2 ('How are the nodes grouped by linguistic strategies?'), the data collected and analysed in the two lines will reveal which aspects of ISE are grouped together in 7 African languages, and with these groupings we can establish the relations between the nodes in the conceptual space of ISE (subquestion S3).

3.3. What does the grouping tell us about the relations between the nodes? (S3)

Having established which nodes form the basis (S1) and how languages group them (S2), these groupings are then used for a network analysis, which will provide a visualisation of the groupings as a proximity map, and at the same time highlight how the nodes in the conceptual space of ISE are related. Inspired by the now widely used colexification networks (as first proposed by Cysouw 2010, later expanded by List et al. 2013, and most prominently presented in the CLICS database), we will construct intra-linguistic and cross-linguistic co-expression networks and analyse them both quantitatively and qualitatively.

On the **intra-linguistic level**, as soon as one language employs multiple strategies to express different parts of the grid/space (which all languages are expected to), a network can be created of just that language, showing the proximity relations between the individual nodes. In this way, the ISE relations in individual languages form the input for the last subquestion (S4) and the second main question (M2): are the individual networks compatible with each other? A computational comparison for similarity and compatibility will enable us to answer that question.

On the **cross-linguistic level**, taking all the groupings from the investigated languages together, we arrive at an impression of what the (hypothetically universal) conceptual space of ISE could look like, and plot this as a map. A pilot analysis of a small sample based on existing literature is visualised in Figure 4. The network analysis thus answers the first main research question M1 of how the conceptual space of ISE is organised.

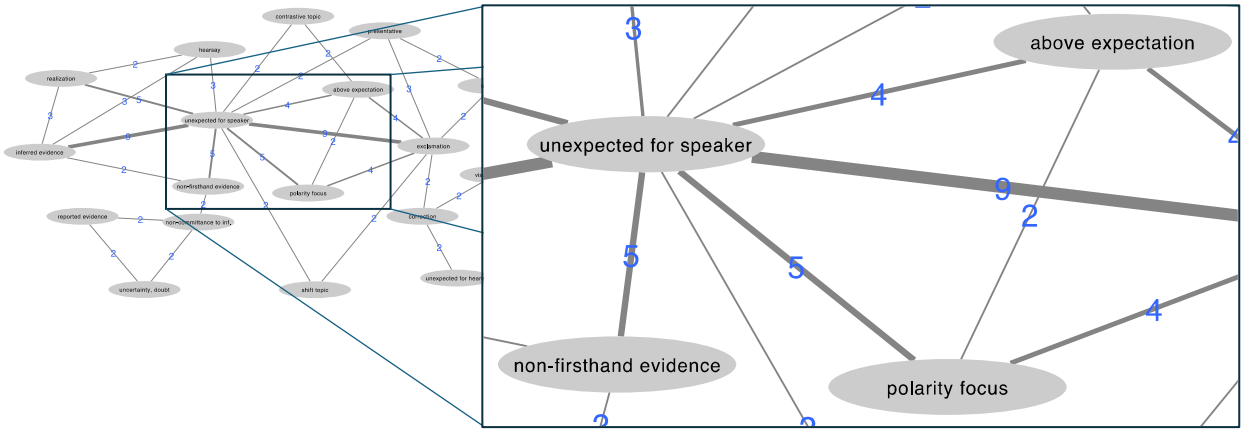

Figure 4 – Visualisation of relations between ISE concepts based on a small sample from the literature (courtesy of J. Mattis List). Note that 1) the nodes here are based on the description in the literature and not on the detailed semantics foreseen for the proposed project, 2) some nodes therefore represent broader concepts than others (e.g. direct evidence is a superset of visual evidence), and 3) some nodes' names are abbreviated, e.g. not indicating whether it is speaker- or addressee-related.

For this network analysis, expert input is required, which is organised in three ways. First, a Postdoctoral researcher will be employed after data collection to provide the computational analysis of the gathered data, as well as the visualisation. Second, in order to ensure that the data are gathered, organised, and stored in such a way that the Postdoc can work with them, the computational linguists on the advisory board have agreed to provide an in-person intensive training workshop before data collection and a check-in shortly after the start of the data collection. Third, a data steward will be appointed part-time from the first training onwards and overlapping with the Postdoc so that the PhD and collaborators always have recourse to expert data management support. The data steward will also participate in the training workshop, along with the PhD, PI, and collaborators, to make sure the whole team is up to speed and can help each other too.

3.4. What do the relations between the nodes reveal about our linguistic ability, our conceptual organisation, and the possible interaction between them? (S4)

The insights into how the nodes in the ISE space are related (results of S3) inform and impact linguistic typology as well as theoretical linguistics.

First, the **crosslinguistic map** (taking into account all data from all languages) shows whether and where clusters/hotbeds exist, and this provides insights into which conceptual relations are particularly strong. This can in turn inspire historical, functional, and structural explanations. The project will focus on the latter two, but historical linguistics can also benefit from the findings, as these relations are likely to form subsequent stages in diachronic development (e.g., Aikhenvald 2011, Narrog & Van der Auwera 2011, Van der Auwera 2013, Willett 1988).

For the functional approach, the relations and maps in S3 can be used to both **test and extend the typological generalisations** and the proposed explanations of multifunctionality/polysemy. Examples of such generalisations and explanations to be tested include the following:

- "A visual or a direct evidential presupposes personal access to information on behalf of the speaker, and thus has an in-built overtone of egophoricity" (Aikhenvald 2021: 36)
- When correcting the addressee (contrast), the speaker must assume that they are the only person to have the right information, i.e. have epistemic primacy (egophoricity) (Grzech 2020, Evans et al. 2018b)
- If some information is given, it is accessible to both interlocutors (engagement) (Evans et al. 2018ab)

- Lack of firsthand information (evidential) implies the speaker's non-participation and lack of control, suggesting an unprepared mind, which may result in surprise (mirative) (Aikhenvald 2004: 39).

Interestingly, the generalisations may contradict each other, as in the case of a possible link between direct evidentials and mirativity. On the one hand, Peterson (2016: 1329) states that "only certain kinds of evidentials can be used to express surprise; specifically, in systems with two or more grammatical evidentials, the evidential that is most specialized for sensory evidence is usually the one that expresses mirativity." On the other hand, Aikhenvald (2021: 35) posits that "Visual or direct evidentials hardly ever have any mirative extensions." These (and other) generalisations and predictions form concrete questions in the investigation (see the literature review as described under S1 in §3.1). Systematic data from the African languages in this project can shed light on these generalisations and the conceptual logic proposed on the basis of other languages.

Regarding structural explanations, the crosslinguistic variation and tendencies will inform **theories about the pragmatics-syntax interface**, contributing to a recently revived debate about the 'syntactisation' of interactionality (Wiltschko 2021, Miyagawa 2022, among others). Descriptive research contributes to the modelling of our human language capacity, in which binary hierarchical structures play a fundamental role. It is broadly accepted that such structures have a place for core syntactic notions such as negation or complementisers, but a pressing question is whether aspects of ISE should also be incorporated in this model, and if so, whether this should be universally so or could differ from language to language (and by which parameters).

Wiltschko's (2021) Interactional Spine Hypothesis forms a hybrid proposal between universal and language-specific, whereby interactional language is universally modelled as two parts of the Grounding layer: GroundAddressee and GroundSpeaker (see Figure 5). This is a promising account to understand ISE, as it can account for the common features and the cross-linguistic variation in different aspects of ISE. Which aspects are grammaticalised depends in this model on which language-particular Unit of Language is used to value GroundAdr/Spkr. For example, it is possible that the Cinyungwe particle =*di* values GroundAdr (removing any doubt from the Addressee's mind), and that in Rukiga the particle -*o* values it for contrast with what was in the Addressee's ground. Furthermore, this model actually predicts the multifunctionality of ISE markers, depending on whether they value GroundSpkr or GroundAdr, and whether the valuation is positive or negative – for example, the same linguistic strategy may indicate surprise for alternatively the speaker or the addressee.

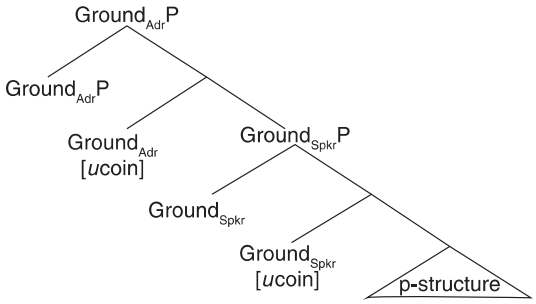

Figure 5 – The interactional spine linking the Addressee's and Speaker's ground to the propositional structure (Wiltschko 2021: 82)

The answers to subquestions S1-3 in the project will form the input to test this model by investigating how the interactional higher parts of the spine (as in Figure 5) interact with the lower layers in the spine (Wiltschko 2014), specifically Linking (thought to be related to information structure) and Point-of-View (associated with aspect but also evidentiality, cf. Speas and Tenny 2003). In this way, the model can help us understand which aspects belong to the universal human language capacity (the Grounding layer) and which are language-specific (the Units of Language) to form language-individual categories. Vice versa, the new data and insights will improve the model.

Second, the comparison of the **intra-linguistic maps** allows us to assess the similarity of the conceptual organisation between two (or more) languages. If the maps are compatible, we will have a strong indication that the human mind organises the concepts of ISE in the same way, independent of language. If the maps turn out to be incompatible (for example, two concepts are

closely related in one language, but far apart in another), then we will have discovered a potentially even more interesting conclusion, namely that language influences the way we conceptualise and organise intersubjective epistemicity (entering into the area of linguistic relativity, which has been rejected, e.g. Pinker 1994, 2007; McWorther 2016, but has regained interest, e.g. Lucy 1992, Boroditsky 2001 – see Casasanto 2016 for a recent overview and references). This answers the second main research question M2.

4. Impact

With each subquestion building on the previous, we discover how our mind organises the knowledge of speaker and addressee with respect to the information in an utterance, and whether this organisation is influenced by the language we speak. The proposed project thus concerns fundamental research into human cognition. Its impact is expected in four areas:

1. We gain an insight into **human mental organisation and conceptualisation**. The results in this project provide insights in how the human mind works, how we understand the world by organising mental concepts, and how we use linguistic strategies to express them. Whether the language we speak influences this organisation is a question many people are interested in, both academic and lay (see e.g. Deutscher's 2010 popular book 'Through the language glass: why the world looks different in other languages'). This question of language relativity should be addressed empirically and the proposed project offers a novel way to do so.
2. We uncover the area of interactive epistemicity in **African languages**. It is important that linguistic and cognitive models are based on more than a handful of languages, as we now have a too narrow understanding of language and cognition: how can we say anything about universal mental organisation (or tendencies for categorisation) if a large part of the empirical area has been left out? Not only does the study of African languages in this project contribute to the data and description of these languages, but the new comprehension of their grammatical systems will influence current thinking and theorising about it. This goes two ways: through presentations and publications that analyse the new data in a theoretical model we understand the data better, and vice versa we adapt and improve the model inspired by the new data. A broader descriptive base thus brings the insights into ISE to a higher level.
3. New methods will help linguist(ic)s go beyond the fixed categories. The more encompassing methodology will help to **prevent misdiagnosis of linguistic phenomena**. By investigating more detailed notions in the whole domain and always involving context (and also implemented in virtual reality), the resulting data will be more reliable as a basis for typological and theoretical generalisations. The core concepts and tools will be shared in the online masterclass, and the materials will remain available for everyone around the globe, so the new methods have a potentially large reach. As such, they can contribute to a change in how linguists think about and practically approach interactive epistemicity.
4. **Training and collaborative opportunities for native speaker linguists** have a potentially huge impact. In meetings with colleagues from Kenya, Uganda, Tanzania, and Malawi, it was indicated that participating in international linguistics courses would be a very welcome catalyst to further linguistic research and education. Training the students and collaborators in the masterclass will boost the professional field. This is because understudied languages are fundamental for our understanding of the human language capacity, and these languages will then be researched with improved knowledge, skills, and tools. This will bring more in-depth and diverse data and knowledge to linguistic debates. The collaborators additionally profit from participation in the project, through the opportunities for joint work and the positive influence on their careers. The effect is anticipated to continue beyond these students and collaborators, because they can in turn teach their own students and colleagues, thus planting a seed that can continue growing.

Abbreviations

CLICS	(database of) Cross-Linguistic Colexifications
ISE	intersubjective epistemicity
LLInC	Leiden Learning and Innovation Centre
VR	virtual reality
DEF	definite
DET	determiner
DIR	direct
EV	evidential
FV	final vowel
INF	infinitive
OM	object marker
PRO	pronoun
PRS	present tense
PST	past tense
Q	question
SG	singular
SM	subject marker

References

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2011. The grammaticalization of evidentiality. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Grammaticalization*, 605-613.

Aikhenvald, Alexandra Y. 2012. The essence of mirativity. *Linguistic Typology* 16. 435-485.

Aikhenvald, Alexandra Y. 2015. Evidentiality and its interrelations with other categories. *Linguistic Typology* 19. 239-77.

Aikhenvald, Alexandra Y. 2021. *The web of knowledge*. Leiden: Brill.

Ameka, Felix K. & Marina Terkourafi. 2019. What if…? Imagining non-Western perspectives on pragmatic theory and practice. *Journal of Pragmatics* 145. 72-82.

Amfo, Nana A.A. 2018. The expression of information structure in African languages. In Augustine Agwuele & Adams Bodomo (eds.), *The Routledge Handbook of African Linguistics*, 243-261. London: Routledge.

Anderson, Lloyd B. 1982. The 'perfect' as a universal and as a language-particular category. In Paul J. Hopper (ed.), *Tense-aspect: Between semantics & pragmatics. Typological studies in language*, 227–264. Amsterdam: John Benjamins.

Asiimwe, Allen. 2024. On the expression of mirativity in Rukiga. Studies in African Linguistics 42 (supplement 14). 28-46.

Asiimwe, Allen & G.J. van der Wal. 2021. The multifunctionality of *-o* in Rukiga: pronoun, contrastive topic, and mirative marker. *Nordic Journal of African Studies* 30 (1).

Bassong, Paul Roger. 2014. Information structure and the Basa'a left peripheral syntax. PhD dissertation, Université de Yaounde I.

Behrens, Leila. 2012. Evidentiality, modality, focus and other puzzles. In Andrea C. Schalley (ed.), *Practical theories and empirical practice*, 185–243. Amsterdam: John Benjamins.

Bergqvist, Henrik. 2017. The role of 'perspective' in epistemic marking. *Lingua* 186/187. 5–20.

Bergqvist, Henrik & Karolina Grzech. 2023. The role of pragmatics in the definition of evidentiality. *STUF Language Typology and Universals* 76(1). 1-30.

Bergqvist, Henrik & Seppo Kittilä (eds.). 2020. *Evidentiality, egophoricity and engagement*. Berlin: Language Science Press.

Boroditsky, Lera. 2001. Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology* 43(1). 1-22.

Botne, Robert. 1997. Evidentiality and epistemic modality in Lega. *Studies in Language* 21. 509–532.
Botne, Robert. 2020. Evidentiality in African Languages. In Chungmin Lee & Jinho Park (eds.), *Evidentials and Modals, 460-501*. Leiden: Brill.
Boye, Kasper. 2012. *Epistemic meaning, a crosslinguistic and functional-cognitive study*. Berlin: De Gruyter Mouton.
Boye, Kasper, and Peter Harder. 2009. Linguistic categories and grammaticalization. *Functions of Language* 16(1). 9-43.
Boye, Kasper & Peter Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88(1). 1-44.
Casasanto, Daniel. 2016. Linguistic relativity. In Nick Riemer (ed.), *The Routledge Handbook of Semantics*, 158-174. New York: Routledge.
Cornillie, Bert. 2009. Evidentiality and epistemic modality: On the close relationship between two different categories. *Functions of Language* 16(1). 44-62.
Crane, Thera. 2024. Looking for evidentiality (and mirativity) in Bantu. Studies in African Linguistics 52 (supplement 14). 9-27.
Crane, Thera, Tim Roth, Hilde Gunnink & Ponsiano Kanijo. 2024. Aspect and evidentiality in four Bantu languages. In Astrid De Wit, Frank Brisard, Carol Madden-Lombardi, Michael Meeuwis & Adeline Patard (eds.), *Beyond aspectual semantics: Explorations in the pragmatic and cognitive realms of aspect,* 177-206. Oxford: Oxford University Press.
Croft, William. 2001. *Radical construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.
Croft, William & Keith T. Poole. 2008. Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. *Theoretical Linguistics* 34. 1-37.
Cysouw, Michael. 2007. Building semantic maps: The case of person marking. In Bernhard Wälchli & Matti Miestamo (eds.), *New challenges in typology*, 225-248. Berlin: Mouton de Gruyter.
Cysouw, Michael. 2010. Semantic maps as metrics on meanings. *Linguistic Discovery* 8(1). 70-95.
De Haan, Ferdinand. 1999. Evidentiality and epistemic modality: setting boundaries. *Southwest Journal of Linguistics* 18. 83-101.
De Haan, Ferdinand. 2000. Evidentiality in Dutch. In Jeff Good & Alan C.L. Yu (eds.), *Proceedings of the twenty-fifth annual meeting of the Berkeley Linguistics Society*, 74-85. Berkeley: Berkeley Linguistics Society.
De Haan, Ferdinand. 2005. Encoding speaker perspective: Evidentials. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.), *Studies in Language Companion Series* 72, 379-397. Amsterdam: John Benjamins.
De Haan, Ferdinand. 2013. Coding of evidentiality, in Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, chapter 78. Leipzig: Max Planck Institute for Evolutionary Anthropology.
DeLancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1. 33-52.
Deutscher, Guy. 2010. *Through the language glass: why the world looks different in other languages*. New York: Henry Holt & co.
Dimmendaal, Gerrit J. 2001. Logophoric marking and represented speech in African languages as evidential hedging strategies. *Australian Journal of Linguistics* 21(1). 131-157.
Dimmendaal, Gerrit J. 2014. The grammar of knowledge in Tima. In Alexandra Y. Aikhenvald & Robert M.W. Dixon (eds.), *The grammar of knowledge*, 245-259. Oxford: Oxford University Press.
Enfield, Nicholas J. 2002. Ethnosyntax: Introduction. In Nicholas J. Enfield (ed.), *Ethnosyntax: Explorations in culture and grammar*, 1-30. Oxford: Oxford University Press.

Ermisch, Sonja. 2009. *Focus and topic in African languages*. Cologne: Rüdiger Köppe.
Evans, Nicholas, Henrik Bergqvist & Lila San Roque. 2018a. The grammar of engagement I: framework and initial exemplification. *Language and Cognition* 10(1). 110-140.
Evans, Nicholas, Henrik Bergqvist & Lila San Roque. 2018b. The grammar of engagement II: typology and diachrony. *Language and Cognition* 10(01). 141-170.
Fiedler, Ines, & Anne Schwarz (eds.). 2010. *The expression of information structure: a documentation of its diversity across Africa*. Amsterdam: John Benjamins.
García-Macías, José Hugo. 2016. *From the unexpected to the unbelievable: thetics, miratives and exclamatives in conceptual space.* Albuquerque: The University of New Mexico.
Gluckman, John & Margit Bowler. 2016. Expletive agreement, evidentiality, and modality in Logooli. In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds.), *Semantics and Linguistic Theory (SALT)*, 1063-1082. LSA and CLC Publications.
González, Paz, G.J. van der Wal, Claudia Berruti, Victor van Doorn, Irina Morozova, Jai von Raesfeld Meyer, Thomas Vorisek. Submitted. Using virtual reality for linguistic fieldwork.
Goodhue, Dan. 2022. All focus is contrastive: on polarity (verum) focus, answer focus, contrastive focus and givenness. *Journal of Semantics* 39. 117-158.
Grzech, Karolina. 2020. Epistemic primacy, Common Ground management and the epistemic perspective domain. In Henrik Bergqvist & Seppo Kittilä (eds.), *Evidentiality, egophoricity and engagement* (Studies in Diversity Linguistics 99), 23-60. Berlin: Language Science Press.
Grzech, Karolina, Eva Schultze-Berndt & Henrik Bergqvist (eds.). 2020. Knowing in interaction: Empirical approaches to epistemicity and intersubjectivity in language. Special issue of *Folia Linguistica* 54(2).
Grzech, Karolina, Eva Schultze-Berndt & Henrik Bergqvist. 2020. Knowing in interaction: An introduction. *Folia Linguistica* 54(2). 281-315.
Güldemann, Tom. 2002. When 'say' is not say: the functional versatility of the Bantu quotative marker *ti* with special reference to Shona. In Tom Güldemann & Manfred von Roncador (eds.), *Reported discourse: a meeting ground for different linguistic domains.* 253-287. Amsterdam: John Benjamins.
Güldemann, Tom. 2008. *Quotative indexes in African languages: A synchronic and diachronic survey*. Berlin: Mouton De Gruyter.
Güldemann, Tom, Sabine Zerbian & Malte Zimmermann. 2015. Variation in information structure with special reference to Africa. *Annual Review of Linguistics* 1(1). 155-178.
Gutzmann, Daniel. 2015. *Use-conditional meaning: Studies in multidimensional semantics*. Oxford: Oxford University Press.
Gutzmann, Daniel, Katarina Hartmann & Lisa Matthewson. 2020. Verum focus is verum, not focus: Cross-linguistic evidence. *Glossa: a journal of general linguistics* 5(1):51. 1-48.
Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure, Volume 2*, 211-42. Mahwah, NJ: Erlbaum.
Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86. 663-687.
Hengeveld, Kees & Hella Olbertz. 2012. Didn't you know? Mirativity does exist! *Linguistic Typology* 16(2). 487-503.
Hill, Nathan W. 2012. Mirativity does not exist: Hdug in Lhasa Tibetan and other suspects. *Linguistic Typology* 16(3). 389-433.
Jordanoska, Izabela. 2020. The pragmatics of second position and sentence final particles in Wolof. PhD dissertation, University of Vienna.
Kanijo, Ponsiano S. 2020. Evidential strategies in Nyamwezi. *Studia Orientalia Electronica* 8(3). 81-98.

Kerr, Elisabeth J., Allen Asiimwe, Patrick Kanampiu, Zhen Li, Amani Lusekelo, Ernest Nshemezimana & G.J. van der Wal. 2023. Bantu word order between discourse and syntactic relations. *Linguistique et Langues Africaines* 9(1).
Kerr, Elisabeth J., and J. van der Wal. 2023. Indirect truth marking through backgrounding: evidence from Bantu. 42(3). *Zeitschrift für Sprachwissenschaft* 42(3). 443-492.
König, Christa. 2013. Source of information and unexpected information in !Xun — evidential, mirative and counterexpectation marker. In Alexandra Y. Aikhenvald & Anne Storch (eds.), *Linguistic expression of perception and cognition*, 69-94. Leiden: Brill.
Langa da Câmara, Crisófia, Michael Diercks, Madelyn Colantes, Brendan Ly, Jackson Kuzmik & Hannah Lippard. Forthcoming. Object marking in Cinyungwe. *Proceedings of ACAL 51/52*.
Levinson, Stephen C. & Sérgio Meira. 2003. 'Natural concepts' in the spatial topological domain – Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 79(3). 485-516.
Lippard, Hannah, Justine Sikuku, Crisófia Langa da Câmara, Rose Letsholo, Madelyn Colantes, Kang (Franco) Liu & Michael Diercks. 2024. Emphatic interpretations of object marking in Bantu languages. Studies in African Linguistics 52 (supplement 14). 78-109.
List, Johann Mattis, Anselm Terhalle & Matthias Urban. 2013. Using network approaches to enhance the analysis of cross-linguistic polysemies. In *Proceedings of the 10th International Conference on Computational Semantics—Short Papers*, 347-352. Stroudsburg, PA: Association for Computational Linguistics.
Lucy, John A. (ed.). 1992. *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.
Masia, Viviana. 2022. Remarks on information structure marking asymmetries: the epistemological view on the micropragmatic profile of utterances. In Davide Garassino & Daniel Jacob (eds.), *When data challenges theory. Unexpected and paradoxical evidence in information structure*, 57-91. Amsterdam: John Benjamins.
Matić, Dejan & Daniel Wedgwood. 2013. The meanings of focus: The significance of an interpretation-based category in crosslinguistic analysis. *Journal of Linguistics* 49(1). 127-163.
McWhorter, John H. 2014. *The language hoax: Why the world looks the same in any language*. Oxford: Oxford University Press.
Mithun, Marianne. 2020. Context and consciousness: documenting evidentials. *Folia Linguistica* 54(2). 317-342.
Miyagawa, Shigeru. 2022. *Syntax in the treetops*. Cambridge, MA: MIT Press.
Mushin, Ilana. 2013. Making knowledge visible in discourse: Implications for the study of linguistic evidentiality. *Discourse Studies* 15(5). 627-645.
Narrog, Heiko & Johan van der Auwera. 2011. Grammaticalization and semantic maps. In Heiko Narrog & Bernd Heine (eds.), *Handbook of grammaticalization*, 318-327. Oxford: Oxford University Press.
Ozerov, Pavel. 2018. Tracing the sources of information structure: Towards the study of interactional management of information. *Journal of Pragmatics* 138. 77-97.
Ozerov, Pavel. 2021. Multifactorial Information Management (MIM): summing up the emerging alternative to Information Structure. *Linguistics Vanguard* 7(1).
Peterson, Tyler. 2016. Mirativity as surprise: evidentiality, information, and deixis. *Journal of Psycholinguistic Research* 45. 1327-1357.
Peterson, Tyler. 2017. Problematizing mirativity. *Review of Cognitive Linguistics* 15(2). 312-342.
Pinker, Steven. 1994. *The language instinct: How the mind creates language*. New York: Harper Perennial.
Pinker, Steven. 2007. *The stuff of thought: Language as a window into human* nature. New York: Viking.

Rooryck, Johan. 2021. 'Recycling' evidentiality: a research program. In Metin Bağrıaçık, Anne Breitbarth & Karen De Clercq, *Mapping linguistic data. Essays in honour of Liliane Haegeman*, 242-261.
Roth, Tim 2018. Aspect in Ikoma and Ngoreme: A comparison and analysis of two western Serengeti Bantu languages. PhD dissertation, University of Helsinki.
Rzymski, Cristoph, Tiago Tresoldi, Simon J. Greenhill, et al. 2020. The database of cross-linguistic colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific Data* 7, 13.
Sandman, Erika & Karolina Grzech. 2022. Egophoricity and evidentiality: Different categories, similar discourse functions - Insights on conversational data from the Tibetan Plateau and the Amazonian Foothills. *Interactional Linguistics* 2(3), 79-109.
Sasse, Hans-Jürgen. 1996. *Theticity*. Arbeitspapiere 27. Cologne: Institut für Sprachwissenschaft der Universität zu Köln.
Sasse, Hans-Jürgen. 2006. Theticity. In Giuliano Bernini & Marcia L. Schwarz (eds.), *Pragmatic organization of discourse in the languages of Europe*, 255-308. Berlin: Mouton de Gruyter.
Sikuku, Justine, Michael Diercks & Michael Marlo. 2018. Pragmatic effects of clitic doubling: two kinds of object markers in Lubukusu. *Linguistic Variation* 18.2: 359-429.
Sikuku, Justine and Michael Diercks. 2021. Object marking in Lubukusu: Information structure in the verb phrase. In Lee Bickmore, Michael Cahill, Michael Diercks, Laura J Downing, James Essegbey, Katie Franich, Laura McPherson & Sharon Rose (eds). *Celebrating 50 years of ACAL: Selected Papers from the 50th Annual Conference on African Linguistics*, 305-330. Berlin: Language Science Press.
Skopeteas, Stavros, Ines Fiedler, Sam Hellmuth, Anne Schwarz, Ruben Stoel, Gisbert Fanselow, Caroline Féry & Manfred Krifka. 2006. Questionnaire on Information Structure (QUIS). *Working Papers of the SFB632* (ISIS) 4.
Speas, Margaret. 2004. Evidentiality, logophoricity and the syntactic representation of pragmatic features. *Lingua* 114(3). 255-276.
Speas, Margaret. 2010. Evidentials as generalized functional heads. In Anna Maria Di Sciullo & Virginia Hill (eds.), *Edges, heads, and projections: Interface properties*, 127-150. Amsterdam: John Benjamins.
Speas, Margaret & Caro Tenny. 2003. Configurational properties of point of view roles. In Anna-Maria di Sciullio (ed.), *Asymmetry in Grammar*, 315-345. Amsterdam: John Benjamins.
Storch, Anne. 2009. Hone (Jukun). In Gerrit Dimmendaal (ed.), *Coding participant marking. Construction types in twelve African languages*, 123-140. Amsterdam: John Benjamins.
Storch, Anne 2018. Evidentiality and the expression of knowledge: An African perspective. In Alexandra Y. Aikhenvald (ed.), *The Oxford Handbook of Evidentiality*, 611-628. Oxford: Oxford University Press.
Storch, Anne & Jules J. Coly. 2014. The grammar of knowledge in Maaka (Western Chadic, Nigeria). In Alexandra Y. Aikhenvald & Robert M.W. Dixon (eds.), *The grammar of knowledge*, 190-208. Oxford: Oxford University Press.
van der Auwera, Johan. 2013. Semantic maps: for synchrony and diachrony. In Anna Giacalone Ramat, Caterina Mauri & Piera Molinelli (eds.) *Synchrony and diachrony: a dynamic interface*, 153-176. Amsterdam: John Benjamins.
van der Auwera, Johan, Peter Kehayov & Alice Vittrant. 2009. Acquisitive modals. In Lotte Hogeweg, Helen de Hoop, & Andrej Malchukov (eds.), *Cross-linguistic semantics of tense, aspect and modality*, 271-302. Amsterdam: John Benjamins.
van der Auwera, Johan & Vladimir A. Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2. 79-124.
Van der Wal, G.J. 2015. Bantu syntax. *Oxford Handbooks Online*.
Van der Wal, G.J. 2016. Diagnosing focus. *Studies in Language* 40(2), 259-301.
Waag, Christine. 2010. *The Fur verb and its context*. Cologne: Rüdiger Köppe Verlag.

Willett, Thomas. 1988. A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language* 12(1). 51-97.
Wiltschko, Martina. 2014. *The universal structure of categories. Towards a formal typology*. Cambridge: Cambridge University Press.
Wiltschko, Martina. 2021. *The grammar of interactional language*. Cambridge: Cambridge University Press.
Yliniemi, Yuha. 2023. Similarity of mirative and contrastive focus: three parameters for describing attention markers. *Linguistic Typology* 27(1). 77-111.
Zeisler, Bettina. 2016. Evidentiality, inferentiality, and speaker's attitude. Questionnaire or exemplary set. Online at: http://tulquest.huma-num.fr/sites/default/files/questionnaires/41/QuestionnaireEvidentiality.pdf