REGISTER
Aspects of
Language
in Situation

Register: Language Users' Knowledge of Situational-Functional Variation

Frame text of the Second Phase Proposal for the CRC 1412

Anke Lüdeling[a], Luka Szucsich[a], Lars Erik Zeige[a], Aria Adli[b], Artemis Alexiadou[a,c], Malte Belz[a], Miriam Bouzouita[a], Oliver Bunk[a], Malte Dreyer[a], Markus Egg[a], Anna Helene Feulner[a], Jürg Fleischer[a], Natalia Gagarina[c], Aron Hirsch[c], Stefanie Jannedy[c], Pia Knoeferle[a], Thomas Krause[a], Silvia Kutscher[a], Mingya Liu[a], Beate Lütke[a], Antonio Machicao y Priemer[a], Katja Maquate[a], Laura Merino Hernández[a], Roland Meyer[a], Christine Mooshammer[a], Stefan Müller[a], Uli Sauerland[c], Antje Sauermann[a], Viola Schmitt[a], Nicole Schumacher[a], Dina Serova[a], Stephanie Solt[c], Jozina Vander Klok[a], Elisabeth Verhoeven[a], Richard Waltereit[a], Melanie Weirich[d] & Heike Wiese[a]

[a]Humboldt-Universität zu Berlin [b]Universität zu Köln [c]Leibniz-Zentrum Allgemeine Sprachwissenschaft [d]Friedrich-Schiller-Universität Jena

The goal of the CRC 1412 "Register: Language Users' Knowledge of Situational-Functional Variation" is to answer the overarching research question "What constitutes a language user's register knowledge?". Our starting point is the observation that many situational and functional parameters – such as the relation between the interlocutors, the purpose of a conversation, the formality of the setting, etc. – influence the way speakers use language. Speakers seem to know which linguistic behavior is appropriate in which situation, typically without having been explicitly made aware of it. This is what we call **register knowledge** and The purpose of this CRC is to enhance our understanding of the role of register knowledge in language use, acquisition, processing, variation, and change; and develop a general theory of register knowledge to complement current linguistic models. In Phase I, we found that register is pervasive on all linguistic levels and is mostly non-categorial. In Phase II, we will be concerned with the integration of the findings into **models** of grammar, acquisition, variation, change, and processing.

Keywords: register, variation, intra-individual, situational-functional, modeling

BerlinUP Journals

Lüdeling et al. 2024. Register: Language Users' Knowledge of Situational-Functional Variation — Frame text of the Second Phase Proposal for the CRC 1412. *Register Aspects of Language in Situation (REALIS).* 3(1), pp. 1–62, doi=https://doi.org/10.18452/29131

ISSN 2750-9370

1 Introduction

The CRC 1412 "Register" initially defined register as those **aspects of intraindividual variation in linguistic behavior that are influenced by situational and functional settings**.* Specifically, it focuses on the linguistic knowledge speakers possess that underlies register variation. We call this register knowledge.[1] The central question the CRC Register pursues is:

***Q.* What constitutes a language user's register knowledge?**

Linguistic behavior is highly variable and much of this variation is (at least partly) influenced by situational and functional parameters. Register variation and register knowledge play a key role in the understanding of grammar, language acquisition, language change, or language processing – in brief, we are convinced that register is something all linguists need to know about. While the Phase I results have shed considerable light on register phenomena (see Section 2 for a detailed discussion), further interdisciplinary investigation will continue to be carried out in Phase II.

In the founding application, we defined three long-term research goals corresponding to the proposed successive funding periods of our CRC: In Phase I, we comprehensively identify and systematically describe register phenomena in all linguistic domains. Building on our findings, Phase II is devoted to how register can be captured with respect to existing models of grammar, acquisition, change, and processing and contribute to their further development and adaption. These will be the first steps towards a full integration of register into linguistic theories, which is the goal of Phase III.

In our understanding, registers are instantiated as individual linguistic behavior and they are recurrent and implicitly recognized and evaluated in speech communities. Following this notion, the CRC explores registers from several angles. First, register represents intraindividual variation and therefore we look

*A list of projects is included in the Appendix in Table 1. For the current status of our research and publications as well as further information about the CRC, we refer to our website at https://sfb1412.hu-berlin.de. Some administrative details of the research proposal were removed from the current publication of the frame text, indicated by […]. In addition to the project PIs listed here as authors, the following people also contributed to the text: Jordan Chark, Sophia Döring, Felix Golcher, Martin Klotz, Milena Kühnast, Nico Lehmann, Pia Linscheid, Anne Temme, Valentina Pescuma, Tonjes Veenstra.

[1]We extensively discussed different notions and traditions of situationally influenced linguistic behavior in the proposal of Phase I, which we published in the CRC's journal REALIS, see Lüdeling et al. (2022).

at linguistic behavior by the same individual in different settings. Second, as the intraindividual variation corresponds with the linguistic behavior of a group in comparable situational and functional settings, we also analyze group behavior. Third, we explore attitudes and awareness regarding registers. It will be the central contribution of the prospective models to capture the interaction among these three aspects.

In Phase II of the CRC, the focus of the work will thus shift from the analysis of data towards the development of **models** of register knowledge. The research during Phase I also led us to revise our working definition of register. For the task of modeling, we focus on the **relations** between situational-functional **parameters** and linguistic **phenomena** (see Section 3.1 for details). We conceptualize the total of all linguistic phenomena related to a particular set of parameters, i. e. the sum of the register-related phenomena, as **register**.[2]

During the past four years, we have demonstrated that register represents a challenging, but also rewarding, topic for linguistic research, especially because register unlocks a wealth of data that are frequently puzzling for existing analyses.

In the following section, we will briefly explain the structure of the CRC, and then describe the overall key results (Section 2) as the base for our focus **modeling** (Section 3.1). Section 3.2 introduces the types of models we will use in the CRC. Areas A, B, and C will then present some of their key results and the research program that follows from it. Some methodological considerations are presented in 5 before we show how we collaborate on the cross-project topics in Section 6.

2 Key results achieved in the first funding period

In Phase I of the CRC Register, we identified and classified **register parameters** and **register phenomena** on linguistic (lexical, morphological, syntactic, phonetic, semantic, pragmatic, text-structural) as well as some visual/graphic levels. Six **major general results** of Phase I are the following: 1) We are able to show that no level is exempt from register dependent variation. 2) Almost all of the register phenomena are non-categorial. 3) In multilingual settings, the choice of the language (language alternation) can be register-induced. 4) Regarding register variation, social meaning is (at least) as important as logical meaning. 5) A text

[2]Terms for the relevant concepts vary across theories and fields (see Lüdeling et al. 2022 for an overview). In the CRC, we decided on a consistent terminology.

can contain many registers. 6) Existing corpus and experimental methods have to be adapted to challenges imposed by the study of register-related phenomena.

Key Results 1 & 2 – Pervasiveness and Non-Categoriality:

We discuss the pervasiveness and non-categoriality of register phenomena in one section since they are both supported by the same data. Specifically, the evidence for pervasiveness and non-categoriality we discovered comes from register-related phenomena across linguistic levels, across contemporary and historical texts, across many different languages and cultures, as well as spoken and written texts, in production, acquisition, comprehension and perception.

In more detail, phenomena on all linguistic layers (including visual aspects of written texts) vary with situational-functional parameters. For instance, speech directed at the same interlocutor in two different attires exhibits **fine phonetic differences** (C02). Situational differences, but also properties of interlocutors like age and gender determine the **use of (non)honorific pronouns** in Persian (A06). B03 showed that the Ancient Egyptian **iconographic principles for the depiction of figures in daily-life scenes** varied for craftsmen in contrast to members of the elite.

Register knowledge is relevant to and takes effect on all linguistic layers but effects individual layers to varying degrees and according to specific situational-functional parameters. In some experiments, for example, we found clear phonetic differences but no syntactic differences (C02). In general, the lexicon seems to behave differently than grammar (syntax). In a controlled essay corpus, we observed that the interindividual variance in the lexicon is considerably higher than the grammatical variance (C04). And we saw that a large portion of the lexicon does not seem to have a strong register preference. On the other hand, some lexical and phrasal elements such as scientific terms or stance markers seem to be highly register specific (C05). For spoken language comprehension, **register effects interacted with lexical (verb) restrictions** on subsequent arguments suggesting a rapid interaction of standard language processing mechanisms with register (C03). Regarding syntax, only relatively **few syntactic alternations seemed indicative of registers**, but see, for instance, long topicalization in Czech. Its acceptability shows a stark contrast between formal written and informal spoken contexts (A03). Linking to situational and functional parameters, we observed **multimodal correlations** between text structure (sequence of text passages, etc.) or plot (narrative sequence vs. direct speech) and their respective visual/graphic representation (line vs. column, spacing, rubrum) on the textual carrier in Ancient Egyptian narrative texts (B03).

Key Result 3 – Language alternation:

In a multilingual setting, the choice of a specific language (or variety) is a potential register-related phenomenon (A02, A06, and C07), as is language mixing versus separation. This was observed in rather diverse contexts, e. g. in the context of Namibian German (C07, Sauermann et al. forthcoming) or in the context of Morisien vs. Bislama (A02, Meyerhoff et al. 2023, Veenstra 2021). In contexts where several languages or varieties are available to the interlocutors, the **distribution on functional sequences of communication** – like welcoming, narrating, or inquiring – can be understood as register differences (Creole Continuum; Latin/Vernacular; Ancient Egyptian/Nubian; multi-ethnic contexts). We find the same in historical texts (e. g. in magical spells from Ancient Egypt; B03): **Code-switching** can occur as a register marker.

Another interesting finding regarding multilingual contexts concerns **processes of register leveling**, as shown for certain heritage languages in Wiese & Bracke (2021). Grammatical phenomena that had before been attributed to language attrition in bilinguals and that are also evident in monolinguals are in fact register-related. Wiese et al. (2022) show that patterns of informal registers in monolingual speakers can be generalized to formal registers in heritage language use, leading to register leveling. Two such examples are non-canonical bare NPs in heritage German in the US, and relative clause formation in heritage Greek. Alexiadou & Rizou (2022) discuss a further case of register leveling in heritage Greek involving light verb constructions. The acquisition and development of registers in multilingual speakers is especially relevant for this CRC: language mixing is assumed to characterize spontaneous, informal speech of multilinguals. By contrast, formal registers in a minority language can only be acquired through formal education, which often leads to register leveling.

Key Result 4 – Significance of meaning differences:

Variationist work in sociolinguistics distinguishes between logical and social meaning and models the two as independent components of meaning to be studied separately (Eckert & Labov 2017). We found further evidence that both of these correlate with register variation. Certain phenomena involve logically equivalent variants which nonetheless differ in social meaning, examples being the phonetic variation investigated in project C02, the concord phenomena studied in A07, or the morphosyntactic alternations studied in A04. But in many (perhaps most) cases we are dealing with a difference in logical meaning as well (see especially the phenomena studied in A05; Mühlenbernd & Solt 2022, Sauerland 2022). Often very subtle meaning differences between alternative expressions (e. g. at-issue vs. non-at issue) are exploited for register purposes diachronically

and can subsequently spread and lead to a stable alternation (B01, Alexiadou et al. submitted, and Chark's dissertation project in A05). In some of the cases we were looking at, it is the most syntactically complex form that is recruited for the expression of non-at issue meaning, e. g. in project B01. Furthermore, a difference in logical meaning may causally underlie variation in social meaning (Beltrama et al. 2022), which in turn becomes a driver of register variation. Consequently, our results necessitate models of register that integrate social meaning, logical meaning, and their interaction. In some cases, we find register-related differences that do not seem to vary either with logical or with social meaning (see C06). These cases will be an interesting challenge for models of register.

Key Result 5 – Register variation within texts:

To a larger degree than expected, **texts are potentially heterogeneous** regarding registers. A text can contain multiple registers or sequences of registers and the level of granularity with which a text is studied may prove crucial for the analysis (for similar findings, see also Biber & Finegan 1994, Egbert & Gracheva 2023). In educational contexts, we find, for instance, that the asymmetric relationship between teacher and student can be interrupted by narrative passages with their own social role relationships and choice of grammatical phenomena (B04, C05). This has repercussions for corpus-based register research where a text is often equated with one register. We moreover notice **path dependencies** on several levels (C04). Structural priming, for example, and self-priming (as basic cognitive mechanisms) play a major role with respect to text production and therefore have to be taken into account when modeling register.

Key Result 6 – Research Methods:

We addressed the non-categorial nature of register phenomena employing either experimental methods or corpus data. For our investigations, we thus relied on a rich inventory of experimental and corpus-based methods, from frequency ratios to Latent Dirichlet Allocation, and from a range of experimental paradigms (such as matched and open guise, language situations, newspaper correction tasks, elicited production, acceptability, appropriateness, formality ratings, and eye-tracking). However, experimental investigation of register has frequently proven more challenging than anticipated. Thus, one of the main findings from Phase I concerns the methodological wealth and complexity characterizing an interdisciplinary approach to the empirical study of register. We have synthesized our efforts in a collective methodological publication (Pescuma et al. 2023, see also Section 5), where we show how (existing and novel) corpus-based and experimental methods can be adapted to the various challenges imposed by the study

of register-related phenomena, and how the various methodologies can inform different aspects of register research.

Previous literature and our own findings suggests that register can be construed as **multivariate** and **multidimensional**. It is certainly multidimensional. It is also multivariate because many parameters have an influence – one can investigate them separately, or in bundles. In Phase I, we found the following situational and functional parameters to be particularly significant: **social meaning** (A04, A05, A07, C02), **formality** (A03, C02, C03), (social-communicative) **function** (B02, B03, B04, C05), **time and space** (B03, C05), and **tenor** (B04). We also saw that even small changes in the parameters could have an influence on the phenomena.

These and other insights from Phase I will shape our research in Phase II, where the focus will be on modeling. Based on our findings, we have already started to model the mental representations implicated in the processing of register as well as add register knowledge to existing Bayesian pragmatic models of rational language use to develop and constrain hypotheses. These efforts are only first steps in modeling register and implicated mental representations, and they will be advanced in our research in Phase II.

3 Focus on modeling in Phase II

3.1 Relation between parameters and phenomena as base for models

All models to be adapted or developed in Phase II will have to specify which linguistic **phenomena** and extralinguistic **parameters** are assumed, and how the **relationship** between them is understood. Figure 1 illustrates the relation between parameters (P) and linguistic phenomena (L) by showing a) how situational-functional parameters exist in a multi-factorial constellation of inter-connected components, encompassing all the parameters that describe a particular situation, irrespective of whether they are related to the register (marked in blue) or not, and b) how an actualization of a given register with its clustering of register-related linguistic phenomena (marked in orange) will co-exist with unrelated linguistic phenomena in a multi-dimensional space. This multi-dimensional space, as schematically visualized in the right box, contains linguistic phenomena that can be looked at from different levels: take, for instance, referent identification as a higher dimension that groups lower dimension phenomena such as syntactic placement next to morphological form, i. e. whether the referent is referred to using a noun, pronoun or even left unspecified as with null arguments. Each of the lower dimension phenomena may then be investigated individually, yet

together they affect how the referent is specified, i. e. the encompassing dimension. At a very high dimension, switching from one language to another for register reasons (see Cross-Project Topic Multilingualism/Language Contact, Section 6.2) would encompass a very different constellation of linguistic phenomena altogether, highlighting the relevance of specifying linguistic dimensions. The left box portrays a snippet of all the possible situational-functional parameters where each P stands for a parameter such as place, time, channel, but also topic, number of speakers, etc. They are depicted as a network (without an intention to resemble neural networks). Some of the parameters will be connected to linguistic phenomena by a relation, i. e. they are register-related.

Parameters are inter-dependent, i. e. they build a complex network: Parliamentary debates, for instance, tend to have larger audiences so that setting the parameters for e. g. location, domain and purpose to parliamentary debate triggers the adjustment of the parameters for the presence of an audience and also audience size. In other words, changing one parameter causes traction on the strings to linked parameters. It will be vital in Phase II to strengthen our expertise on the constellation, interplay, and effect size of parameters, for example with respect to the culture-dependent parameter grouping associated with formality distinction (see Cross-Project Topic 6.1). The linguistic phenomena (L) on the right in Figure 1 represent a selection of possible linguistic means – understood in a broad sense – at play in a situation from the sphere of all communicative means. They are best conceptualized as variables on distinct dimensions, visualized by circles surrounding groups of phenomena. A linguistic phenomenon may be situated on any level of linguistic dimensions, i. e. larger or smaller encircled spaces in the phenomenon sphere. It is our goal in Phase II to get a clearer understanding of how register-related phenomena cluster on different levels of dimensions and what interactions between dimensions are at play.

Complexity is a linguistic dimension that has received ample attention with respect to register (Biber et al. 2022, Szmrecsanyi & Engel 2023: among many others). It perfectly exemplifies the need to grasp phenomena as elements in a multi-dimensional space, seen e. g. in the implementation of various complexity phenomena in the study by Weiss & Meurers (2019), who included over 300 complexity measures to grasp the overall complexity of a text. In sum, all the different complexity phenomena together form the dimension of complexity, but they also group into lower level dimensions such as syntactic complexity, lexical complexity, discourse complexity, or morphological complexity. At this lower dimension, a text may be syntactically more complex while being morphologically less complex compared to other texts. Biber & Gray (2010) illustrate how register effects differ depending on the dimension looked at; while academic writing is

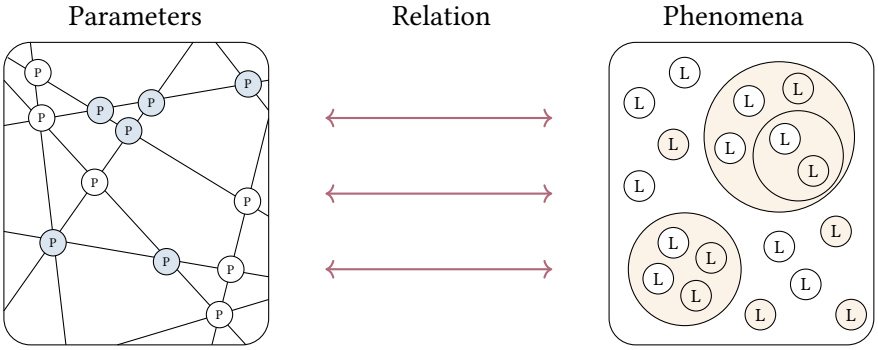

Figure 1: Register concept: A schematic depiction of the relation between parameters and phenomena which constitutes an abstract register. The multi-factorial constellation of parameters P (left-hand side) includes parameters that are register-related (blue) or not register-related (white). The multi-dimensional sphere of linguistic phenomena L (right-hand side, represented by circles) encompasses a cluster of frequently co-occurring register-related phenomena (orange) or phenomena not related to register (white) on various levels.

claimed to be syntactically more complex, an even lower dimension is shown to be much more informative: academic writing is more complex in terms of phrasal embedding but not more complex with respect to clausal embedding. At an even more fine-grained dimension, the modality of a discourse has been shown to be related to types of clausal embedding: written texts such as letters have more adverbial clause embeddings than spoken conversations (see results of A06, and similarly A03 regarding noun frequencies), but fewer complement clause embeddings (see also Verhoeven & Lehmann 2018 for differences between spoken registers). Awareness of the clustering of phenomena on different dimensions enables us to better compare register-related phenomena and gain deeper insights into their register impact. Defining the relation between parameters and phenomena will depend on looking at the relevant dimension in this relation.

Individual language users may have knowledge about this relation to different extents. One instance of language users' knowledge is schematically outlined as part of a register system by Figure 2. Language users are under constant impressions of parameter constellations while experiencing register instantiations, both when perceiving linguistic behavior and when producing it, as illustrated by the figure head juxtaposed with the parameters and phenomena. We depict the former as being linked symmetrically by an arrow pointing in both directions because parameters not only influence the individual but the individual also creates situational-functional parameters, e. g. by changing the topic or using hon-

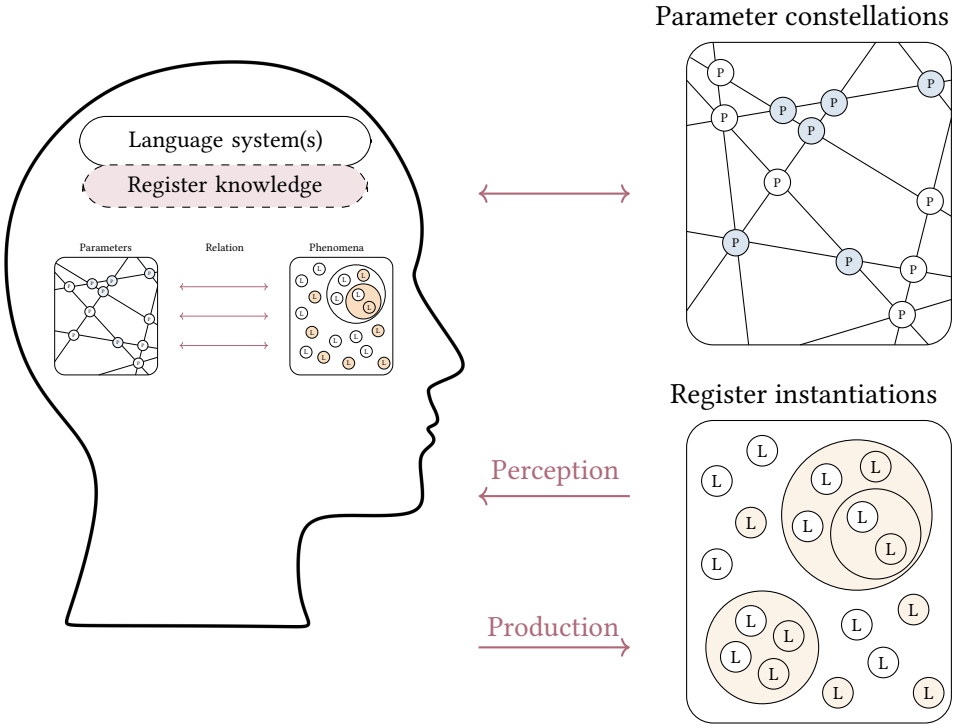

Figure 2: Register system: A schematic depiction of an individual's register knowledge for a given register, where external parameters (P) and perceived or produced co-occurring linguistic phenomena (L) in squared snippets on the right side influence the register knowledge representation (involving the three main parts of the register concept as presented in Figure 1) that a particular individual may have as seen inside the head on the left . (Dis)congruency between right side and representation displays how an individual's knowledge of the perceived register may be more or less accurate depending on exposure to instantiations of the register, i. e. familiarity with the register.

orifics to *force* a formal situation. By acquiring a register to some degree, a language user implements a representation of the register concept in their register knowledge, i. e. the register-related parameters and phenomena and the relation between them, yet this representation will not be completely identical to the perceived parameter constellations and register instantiations for various reasons. The results of Phase I indicate that register knowledge is non-categorial in nature and demonstrates a large degree of variation (see Key Findings from Phase I, Section 2). Thus, individuals may not recognize all relevant parameters and phenomena with respect to syntactic complexity. A speaker may, for instance, under-

stand that a high number of complement clauses embedded by verba dicendi or verba sentiendi is less appropriate in formal written registers while at the same time not realizing that written registers are syntactically and morphologically more complex than other registers. Figure 2 also sets language system(s) and register knowledge into relation by juxtaposing them with a shared line in between, thereby leaving open the manner and extent of their relation. One main question of Phase II is how register knowledge is connected to grammatical knowledge, and how these can be integrated.

3.2 Modeling options

As we stated in our founding application, register is and should be central to linguistic theory building. In Phase II of this CRC, we build on the findings of Phase I to formulate models of the relations between linguistic phenomena and situational-functional parameters as defined in Section 3.1. Moreover, it is our aim to exploit and revise current models in order to capture and integrate register knowledge: One way to exploit existing models is to approach the modeling of register choices along the lines of choices in other linguistic areas, such as allomorphy, the management of implicatures, or code-switching. All models represent parts of reality for the purpose of understanding by reducing complexity. Different models may thus highlight different aspects of the same object, depending on which properties are considered fundamental, emphasized, or disregarded. They illustrate, visualize, and demonstrate the knowledge of the fundamental properties of their objects, their relations, and operations (McClelland 2009, Stachowiak 1973: 131ff.). Different types of models are common in the respective subfields of linguistics represented in the CRC. Our goal of integrating register knowledge into existing models may thus require different steps for the different (types of) models: While situationally and functionally conditioned variation is already a component in some models, others may only need to be expanded by new concepts, and still others may have to be changed fundamentally. A crucial focus for the CRC-internal cooperation will thus be to explore and discuss the insights into register phenomena contributed by the models built in the different linguistic subdisciplines. This will constitute the empirical and theoretical foundation for the main goal of the CRC in Phase III which pertains to weaving together different strands into a comprehensive tapestry of register knowledge.

We agreed to distinguish between three very abstract model types: verbal models, formal-theoretical models, and statistical/probabilistic models.[3] The concrete models we use (to explain acquisition, change, grammar, etc.) fall into one

[3]We considered the vast and diverse body of literature on modeling from linguistics, Natural

or more of these types. Verbal models play the primary role in areas where data are scarce such as many domains of historical linguistics. Compared to formal models, they allow the representation of more complex sections of reality. Formal-theoretical and statistical/probabilistic models are used more in areas where data are readily available. While many formal linguistic models such as classical phrase structure grammar are purely discrete, the models we will employ in the CRC will mostly contain some probabilistic aspects because we have already found that register phenomena tend to be of a probabilistic nature. Different models are employed within and across areas, and often projects combine modeling options. In the following, we will briefly introduce the three model types:

In **verbal models**, knowledge is represented in the form of narratives. Verbal models are the dominant type of models in e. g. historical linguistics or variationist linguistics, and are an important component of neural and cognitive linguistics, covering, among other things, language variation and change, language processing (both production, and perception), and language acquisition. In these disciplines, verbal models can be based on both the interpretation of documentary materials, a hermeneutic approach, and on corpus or experimental data. Especially in scarce-data situations arising due to a lack of historical sources, restricted access to informants, or inaccessible cognitive processes, the explanatory value of verbal models is based on the ability to interpret poor data and to faithfully communicate the resulting limitations (Jenset & McGillivray 2017: 48). Projects A06, A09, B03, B04, C05, C06, and C07 will primarily employ verbal models to approach cognitive representations and processes as well as diachronic processes of linguistic registers. For language variation and change, verbal models are rooted in diverse theories of language change, from grammaticalization (Heine & Narrog 2012), or naturalness (Wurzel 1994) to the invisible-hand theory (Keller 1994), social functionalism (Eckert 2000), or dynamic systems theory (Lass 1997). They regularly include the use of poetic tropes like metaphors, e. g. 'bridging context' in grammaticalization, to grasp relevant concepts. For language processing, verbal models have been expressed in the form of principles or strategies on how to attach a new phrase into an already-built structure, how to build syntactic structure (a 'parse') and assign thematic roles, and when to integrate syntactic or phrase structure with semantic and world knowledge,

Language Processing, and the philosophy of science and chose to adopt these three categories. The main aim was to foster cooperation on comparable issues of modeling across projects and areas. See e. g. Stachowiak (1973), Giere (2004), Braun & Saam (2015), Ritter & Gründer (1984), Wolters (2004), Chomsky (1962), Hockett (1954), Montague (1974), Cresswell (1973), McClelland (2009), Blochowiak et al. (2017).

e. g. two-step models (Frazier & Fodor 1978), unrestricted accounts (MacDonald et al. 1994), the alignment framework (Pickering & Garrod 2004), or the coordinated interplay account (Knoeferle & Crocker 2006, 2007).

Formal-theoretical models in linguistics are composed by abstract and explicit rules and their descriptive and explanatory power serves to build theoretical accounts which are subsequently empirically tested, validated and often adjusted. Ideally, a formal model should not only possess explanatory power, but also be able to generate predictions of new observations (see Shmueli 2010, Jäger 2019 for discussion). In the generative grammar tradition, the rule-based modeling of language predicts the set of grammatical expressions and, moreover, it does not over-generate, i. e. there are no ungrammatical expressions in the output. Specifically, in the domain of morpho-syntax, register variation can be modeled as the result of an individual's control of multiple grammars (or multiple sub-grammars or multiple settings for individual parameters), each of which is fixed and deterministic, but which differ from each other regarding relevant linguistic features, and compete with each other for use at any given time. In this case, there is competition between two or more grammars with different parametric values (Kroch 1989). A different modeling option of register variation is to allow multiple distinct outputs for a given grammatical specification, in effect assimilating variation to allomorphy, see Adger's combinatorial variability approach (Adger 2006). We are aware of the fact that it is far from trivial how to precisely associate variants with social meaning and incorporate such aspects of meaning into grammatical representations, thus a certain degree of adaptation of such models is required (see Sauerland & Alexiadou 2020 and our remarks below). This is in contrast with Wiese & Bracke (2021), following Jackendoff (1997), where communicative situations may be integrated into lexical representations: Each lexical item is assigned a communicative-situation representation. Moreover, especially relevant in Phase II is how these different perspectives have been extended to code-switching/language mixing (viewed in this CRC as a register phenomenon). On the one hand, multilingual speakers may be argued to have two or more systems of realization of a common structure: speakers have a rich lexicon and a wider array of choices to lexicalize these structures, taken from both their languages (e. g. Alexiadou 2020, López 2020). Under the grammar competition approach, multilingual speakers are switching between two or more grammars with different parametric values. For usage-based models, knowledge of the inventory of variants and alternations is integrated into the grammar, and thus we have a holistic model where knowledge of variation and its use is inseparable from knowledge of grammar. Key result 4 from Phase I (i. e. that meaning differences are crucial) entails that semantic modeling is as important to the CRC

as morphosyntactic models. Above we mentioned already that the difference between logical and social meaning is frequently bridged with register phenomena. We observed as well that it is often not-at-issue content that yields register sensitivity (Alexiadou et al. submitted, McCready 2019, and others). Items that encode speaker stance or perspective are a further case (Acton & Potts 2014), e. g. the alternation between non-perspectival *früher* and perspectival *eher* 'earlier', the latter of which tends towards informal registers (Umbach & Solt 2021). Also conforming to the generalization, implicatures arising from differences in levels of precision seem to underlie their recruitment into the register system (Beltrama et al. 2022). Formal models of morphosyntactic structure and of semantics are employed by projects A03, A04, A05, A06, A07, A08, A10, and C03, often in combination with probabilistic models. C07 will further develop the communicative situation model (Wiese 2021, 2020) in combination with the social Coordinated Interplay Account (C03, see below).

Under the label **statistical/probabilistic models**, we distinguish between statistical methods in hypothesis testing and inherently probabilistic models. Statistical methods such as linear or logistic regression are used in empirical research to test model predictions. The probabilistic aspect in these models is solely viewed as the effect of noise caused by unknown or uncontrolled factors that affect the measured variables. A variety of such statistical methods is used across the projects that work with quantitative data, but these statistical models are not different from those already common in the field. We also include in this category probabilistic methods of data classification that are not directly intended as models of a speaker's register knowledge. Project A04 pioneered the use of Latent Dirichlet Allocation (Blei et al. 2003) for this purpose (Schäfer et al. in preparation). The inherently probabilistic models used in the CRC are of two different types. Type 1 are models that represent grammatical knowledge probabilistically. This is especially popular in usage-based models of grammar, but also found in parametric models (Yang 2002). Within the CRC, many projects pursue probabilistic models of grammar, language acquisition, processing and change based on abstract linguistic variables and the occurrence patterns of their variants as e. g. in Biber's multi-dimensional register analysis (Biber 1999: passim) or Systemic Functional Linguistics (Neumann 2014b). Projects that apply, evaluate, and refine different aspects of probabilistic approaches include A01, A03, A04, A06, A07, A09, B06, C02, and to a certain extent C05, and C06. Type 2 models represent and update speakers' and hearers' beliefs about each other probabilistically within Bayesian pragmatics. In particular, speakers' behavior is modeled as maximizing speakers' expected utility across a probability distribution across the space of possible addressee characteristics, situations, cultural requirements,

and other relevant parameters. At this point, the working model developed by project A05 for imprecision shows how semantic differences can yield register differences on this type of model (Mühlenbernd & Solt 2022). In Phase II, this modeling approach will be further developed by A05 and also utilized by A01, which in addition makes use of language models known from Natural Language Processing tasks, e. g. neural networks.

Combined models. While some projects rely on one type of model, several projects in the CRC combine two (or all three) types of models (C03, C05, C06), e. g. verbal models or formal-theoretical models on the basis of advanced statistical knowledge gained from multivariate analyses or a statistical exploration of scarce-data situations. We collaborate across projects and areas to gain, evolve, and share our expertise on modeling. For instance, models of language production and reception may rely on probabilistic knowledge of the processing of individual linguistic units in situational-functional contexts, tested in experimental research, but express a coherent approach to speaker knowledge in verbal form. The latter can make the model amenable to meta-reflection in the context of broader linguistic theorizing across the projects and areas of the CRC. Verbal models, in turn, may draw on (semi-)formal conceptual representations to describe the construction of an interpretation and its integration with contextual information as language is processed. As one concrete example, in Phase I, C03 has begun to extend the social Coordinated Interplay Account (sCIA: Münster & Knoeferle 2018) with assumptions regarding register-related knowledge and situation formality. Probabilities come into play to capture the gradation of register-related expectations. As a next step, we aim to draw on representational formats developed with the CRC by C07 and include these in the sCIA.

The benefits of combined models are clear also for our goal for Phase III: They relate individual research results from different areas of the CRC, formulating an overall approach to linguistic registers. That approach will include aspects of grammar, pragmatic use, variation and change, conceptualization of situation types, mental representations established in L1 and L2 register acquisition, and the important influence of individual and conventional socialization paths.

4 Areas of the Collaborative Research Center

4.1 Area A: Register and grammar

The overarching question addressed in Area A is:

QA. **How does register knowledge relate to grammatical aspects of linguistic knowledge?**

This directly relates to the CRC's examination of the interplay and parallels between register knowledge and grammatical knowledge. During **Phase I** we investigated how the various grammatical alternatives available are recruited for register purposes, and what parallels can be identified with the mechanisms involved in other linguistic phenomena. Our starting point was the observation that language involves choice and alternation at all levels. That is, many phenomena in other domains of linguistic theory are – like register-related phenomena – also centered around selections from a set of alternatives, examples including allomorphy, focus, code-switching, and scalar implicatures. One central thrust of the projects in Area A was to probe such connections for underlying shared mechanisms. By doing so, projects in Area A found register-related variation at multiple linguistic levels, from phonological to morphosyntactic to semantic/pragmatic to language level. Drawing on these results, we can group the register-related phenomena under investigation into three classes: **grammatical alternation** (e. g. doubling, voice alternations, concord, word order/dislocation, pronoun realization), **semantic alternation** (e. g. semantic strength, presuppositional content, speaker perspective, (non-)literality) and **language alternation** (e. g. language mixing, code-switching in language contact / multilingualism).

A crucial part of the research in Phase I was the development of new methodologies and the adaptation of existing methods to investigate register, which was done in close interaction with members of Areas B and C. Nearly all of the projects in Area A made use of **linguistic corpora** as the empirical basis for their research (A01, A02, A03, A04, A05, A06), in doing so developing new methods of annotation and analysis. A04 innovatively applied Latent Dirichlet Allocation (LDA) to the rich annotation of the DECOW16 web corpus successfully inferring register candidates. A04 also developed an annotation scheme for situational-functional parameters and showed that LDA register candidates align with situational-functional parameters in a meaningful way, which makes LDA a suitable approach to the study of register. A03 in their MDA study developed new ways of normalization of linguistic features in a more variationist spirit (Meyer et al. submitted). In cooperation with A04, A03 also undertook first steps for an LDA study of Czech. The annotation of metaphor and metonymy in A01 breaks new ground in combining syntactic and semantic information, first, by including levels of syntactic as well as of semantic analysis, and, second, by explicitly annotating the syntactic contexts which trigger a metaphorical interpretation (Egg & Kordoni 2022). In addition, several projects have employed **experimental methodologies** (A03, A05, A07). To establish associations between register parameters and phenomena, we have conducted production studies (A05), ratings studies (A03, A07), and forced-choice studies (A05, A07), all of which have

necessitated new advances in the experimental simulation of situations of language use. To investigate social meaning and its interaction with register, A05 extended the matched guise technique to incorporate a situational component. Furthermore, several projects have pursued novel combinations of corpus-based and experimental approaches, e. g. by deriving experimental items from an MDA for Czech and Russian which were then rated in (pre-tested) situational contexts (A03). Similarly, A05 developed a corpus-based measure of the formality of lexical items which is being experimentally validated in C03. Regular meetings of members of Areas A and C on experimental methodologies have facilitated this work. Finally, A02 and A06 have conducted **linguistic fieldwork** to develop new situation-specific corpora, for which A06 developed a novel data collection design that involves intraindividual variation in situational settings that are cross-culturally comparable (Adli et al. 2023); A02 advanced corpora based on dialogues between carefully selected speakers based on age, gender and familiarity.

In **Phase II**, we continue the investigation of **grammatical**, **semantic** and **language** alternations, extending the range of specific phenomena to be studied as continuing projects shift or expand their focus and new projects A08, A09, and A10 join the CRC. Specifically, grammatical alternation is the focus in A04, A07, A09, and A10, and one of the foci in A03, A06, and A08, whereas semantic alternation is the main focus in A01 and A05, and also important for A10. Language alternation plays a major role in A03, A06, and A08.

In keeping with the overall focus of Phase II of the CRC, a central goal for Area A is to develop models of the relations between situational-functional parameters and register phenomena, one focus being on capturing their typically probabilistic nature, and integrating these models into more general models of grammar and language use. As described in Section 3.2, the individual projects variously make use of formal, probabilistic, and verbal models, and combinations of these. As is discussed in more detail below, projects in Area A are involved in the following cross-project topics: Formality (A01, A03, A04, A05, A06, A07, A08, A09), Multilingualism/Language Contact (A03, A06, A08, A09), Lexicon (A01, A04, A08), Acquisition/Education (A09), Path Dependency (A03, A07), and Complexity (A04, A06, A07, A08, A10).

In **Phase I** and continuing into **Phase II**, we pursue and extend the broad goals of Area A via three specific sub-questions.

QAi. **How are existing alternations recruited for register purposes?**

From the set of linguistic phenomena that show alternations, only a subset will actually be exploited in a given cultural context to express register differences.

An important question in **Phase I** was thus how to determine which alternations will be sensitive to register, to what extent the inventory and the association with particular register types is predictable, and what other properties of the alternations and the variants involved are relevant. A04 based its corpus analysis on a particularly broad set of lexico-grammatical features (1,200) to determine "pregisters" (for 'potential registers') using Latent Dirichlet Allocation (LDA) (Schäfer et al. in preparation). Starting with grammatical phenomena, and in particular syntactic complexity, A03 found in a corpus study for Russian that written texts are not more complex than spoken ones in general, but that the texts differ in the nature of complexity measures, e. g. frequency of clausal embeddings vs. noun frequency, thus potentially constituting different registers (Buchmüller et al. 2022, see also A06 for German and Persian). Grammar-related results based on formality manipulations via interlocutor relation from A07 indicate that the distribution of negative concord is register-sensitive in that it was rated less appropriate in formal than informal contexts, an effect found in American English but not in British English (Rotter & Liu submitted). Nevertheless, negative concord turned out to be more acceptable than is claimed in the prescriptive and descriptive literature. In the semantic domain, A01 found significant differences in metaphor and metonymy usage between different registers: while highly persuasive registers show a high degree of metaphoricity, registers in which there are brevity constraints (in A01's case, debates and newspaper commentaries) exhibit more metonymies. A05 focused on choice between logically non-equivalent alternatives – such as numerical expressions at different precision level – finding that it may be sensitive to situational parameters, and furthermore that such variation can carry social meaning, which itself is partially dependent on the situational context (Beltrama et al. 2022). The cross-linguistic projects A02 and A06 both used fieldwork methods to explore and compare different languages, with A02 being concerned with the creole languages Bislama and Morisien, and A06 focusing on German, Persian and Yucatec Maya. For Bislama, the project found clear variation for phonological and syntactic variables. Most strikingly, language choice itself is a register marker, e. g. in first encounters, French is at first employed before switching to Kreol Morisien (Veenstra 2021). Furthermore, there is code-switching of noun phrases (as evidenced by the process of liaison in the noun phrase). A06 found that syntactic variants are recruited for register purposes based on their language-specific grammatical nature: while German exhibits arguments with a discourse management function more frequently in informal registers, they are more frequent in formal registers in Persian where they are more tightly integrated into the clause than in German (Lehmann et al. 2024). A06 also found differences in forms of 2SG address (honorific vs. non-

honorific pronouns) between various situations of interaction, which could be explained by drawing on politeness theory.

We are going to extend the repertoire of existing alternations in **Phase II**, adding a more diverse set of phenomena to inform our models, in particular increasing our understanding of also higher levels of language use, e. g. code-switching and linguistic hybridization. The phenomena relevant for Phase II falling under the realm of **grammatical alternations** include forms of address (A04, A05, A06), TAM marking (A03, A06), number marking (A06), agreement (A04, A09), voice alternations (A03, A04, A08) and other aspects of argument structure (A03, A04, A08, A09), negative and modal concord (A07) and doubling phenomena (A10). With respect to **semantic alternation**, A01 extends its investigation beyond metaphor and metonymy to include other forms of non-literal expressions (NLE) such as hyperbole, litotes and rhetorical questions, while A05 investigates the deployment of non-equivalent alternatives for politeness and face management, looking in particular at evaluative language, polite requests and honorification.

A series of A projects are also looking at **language alternations** (A03, A06, A08) in different cultural contexts to investigate the question whether resources of different languages available to bilinguals are used for register purposes in a similar manner as those of a single language. To what extent are languages selected as register markers, e. g. Czech/Polish (Silesian) / *Po naszymu* (hybrid language) (A03), Ukrainian/Russian/*Suržyk* (hybrid language) (A03), Kurdish/Persian (A06), Yucatec Maya/Spanish (A06), Saamaka/Dutch (A08), and how does this relate to variation within one of the involved languages? One question will be what part of the register spectrum is available to speakers of hybrid languages, i. e. are there effects of register leveling (A03, A08)? Moreover, we will investigate how phenomena that have not existed can emerge in contact situations and be integrated into the grammatical system, such as new plurality and definiteness markers in Yucatec Maya (A06). Also, to what extent is language mixing, e. g. between Saamaka and regional/national languages in three different community and linguistic settings, relevant for differentiation between registers (A08)?

QAii. **How do we model the choice in register alternations by language users?** (Phase I: How do language users choose among register-related variants?)

Addressing QAii in **Phase I**, the Area A projects developed insights into the situational parameters underlying register variation and the mechanisms by which speakers choose between alternatives. A01 identified brevity constraints as guiding speakers' choice among variants. For A03, next to mode parameters (oral

vs. written), narrativity seem to drive types of complexity (Buchmüller et al. 2022). For A06, the results on right periphery (word order variation) and forms of address (subject pronoun variation) show that both language-internal and language-external parameters lead to differences between registers. In addition, speakers do not differ uniformly between situations. Rather, their social background (gender, age group) determines which forms of address they prefer in a given situation. For A07, interlocutor relation such as in talking to the customer vs. one's spouse (public vs. private) reliably predicts the 'formality' of the conversation (Rotter & Liu 2023) and in turn the acceptability of register alternatives. A04 established that variant choice is to a great extent probabilistic (a finding echoed by results from several other projects). Language users choose among register variants according to which types of meanings are felicitous in each situation. By contrast, in preliminary data from a small number of speakers, A02 found that speakers categorically select (exclude) certain variants in some text types. It might well be that with more data coded for A02's indicator variables this text type difference will become significant (Meyerhoff et al. 2023). A05 proposed that differences in logical meaning among alternatives can drive register variation. A05's starting point was the perspective that language users make rational choices that address best their (potentially conflicting) communicative goals, including (i) informativity, (ii) hearer expectations and needs, (iii) speaker-related goals (e. g. social meaning), and (iv) further potential parameters, including economy considerations. A05 validated this approach via a probabilistic game-theoretic model which shows that for a particular set of contexts, one can very closely predict probabilistic speaker choices (Mühlenbernd & Solt 2022). The model predicts that form- and meaning-driven variation interact differently with register when the meaning difference between two register variants is important to the communicative goals.

Work in **Phase II** will seek to deepen our understanding of register parameters and their relation to register phenomena. All of the Area A projects investigate aspects of formality as a factor in register variation, with individual projects investigating more specific formality-related parameters including mode (A01, A03, A06, A10), interlocutor relations (A01, A03, A05, A06, A07, A08, A09), and conversational goal or purpose (A01, A03, A08, A09, A10). Both A05 and A06 focus on the role that face management plays in accounting for politeness-related phenomena. Several projects (including A03, A04, A05, and A07) posit social meaning to play a crucial mediating role in explaining the distribution of register phenomena, with A04 in particular exploring the hypothesis that register knowledge can be modeled via grammatically encoded social meanings. Project A10 makes the novel proposal that the frequency of doubling phenomena is sen-

sitive to properties of the discourse, specifically if there is ambiguity in the Question under Discussion (QUD), which in turn is impacted by parameters such as number of discourse participants and conversational goals. A number of projects seek explicitly to untangle complex constellations of register parameters by examining the interaction of situational with sociodemographic, cultural, political, and geographic parameters (A03, A06, A08, A09) as well as aspects of the situation of language use, e. g. monolingual/multilingual, majority/minority status (A03, A06, A08), and individual differences (A07). A variety of approaches will be applied to modeling these relations. As described in more detail below, this includes both formal grammatical models and (probabilistic) pragmatic models. In addition, A01 makes use of Deep Learning methods to capture the interdependence of register and non-literal expressions; A07 plans to integrate register into expectation-driven models of language processing; and A10 approaches their topic from the perspective of the three-dimensional model of diasystematic variation of Koch & Oesterreicher (1985), which will be adapted as necessary to capture their empirical findings.

QAiii. **How do we integrate models of register with models of grammar?**
(Phase I: What can register-related phenomena tell us about the grammatical implementation of variation?)

Addressing QAiii in **Phase I**, we discovered that many register-related phenomena interact with grammatical phenomena in an intricate way, making it a challenging task to precisely integrate variants associated with situational-functional parameters into models of grammatical knowledge. A02 uncovered evidence of a surprising subject/object asymmetry in the syntactic variable in Bislama. This invites a more thorough investigation of other possible sources of subject/object asymmetries in the grammar of Bislama, which ultimately may help tease apart the question of where the locus of variation is. At present, one cannot rule out the possibility of multiple grammars accessed probabilistically, nor the idea that variation is inherent to a single grammar. A03 showed that situation-related choices made by language users are overwhelmingly not categorical, but rather probabilistic (Meyer et al. submitted). This raises questions for the modeling of register differences: What counts as a variable (also regarding variationist variables vs. pure relative frequencies within a Biberian approach; see also discussion in Biber et al. 2016, Szmrecsanyi 2019)? Is it feasible to include variationist variables into a multidimensional analysis, e. g. by normalizing to the frequency of occurrence of the variable? Similarly, A04 showed that morphosyntactic phenomena subject to register variation require a probabilistic component relating

external parameters with grammar. A05 developed a modeling approach, which suggests that (in at least some cases) the probabilistic nature of register variation involving non-equivalent alternatives can be accounted for in the same pragmatic linguistic module responsible for phenomena such as scalar implicatures, rather than in grammar proper. A06 demonstrated that intraindividual variation shows that categorical principles of speakers' competence are enriched by systematic and meaningful information on frequency of use. These probabilistic patterns carry essential information on register and belong to speakers' and hearers' competence.

Work in Area A in **Phase II** will develop models of register variation that integrate with grammatical implementations of the linguistic phenomena investigated. All projects study phenomena that have been observed to vary in grammar, language use or language choice, but in many cases existing models of the variation fail to account for the register component thereof. For example, the doubling phenomena A10 investigates have been described as cases of free variation, completely ignoring their register component (Geurts & Huitink 2006, Zeijlstra 2007). For A01 existing work focuses on the near-equivalence of non-literal expressions to corresponding literal expressions, but differences in register have not yet been considered in detail. Both examples show how the CRC work will contribute to answer research question QAiii – only by capturing the register component of variation, it is possible to model other aspects of variation.

Several projects approach the modeling of register knowledge from the perspective of formal models of morpho-syntax. In the area of **grammatical alternation**, A04 will work within the framework of Head-Driven Phrase Structure Grammar (HPSG) to develop a unified formal grammatical model which not only accounts for the range of morphosyntactic variation (i. e. which forms are grammatically licensed) but also is able to capture the probabilistic preferences between register variants. Relatedly, for **language alternation**, investigating the register component will provide insight into the formal grammatical modeling of language choice, and on the cognitive model of multilingualism. A03 investigates how bilinguals recruit different languages, among others hybrid languages between closely related Slavic languages, to express register-related phenomena. The emergence of hybrid languages/lects rather suggests a single grammar model in the spirit, for instance, of the combinatorial-variability approach (Adger 2006, Adger & Smith 2010). A08 investigates language choice and mixing in three Saamaka communities. The modeling to be explored is within the context of Distributed Morphology approaches to language mixing (e. g. Alexiadou & Lohndal 2018).

Other projects investigating **grammatical alternations** and **language alternations**, in common with those studying **semantic alternations**, explore the possibility that the phenomena in question can be modeled outside of core grammar. In such approaches, formal grammatical and semantic models may play a supporting role in determining which forms will come to serve as register alternatives, e. g. on the basis of syntactic structure or truth-conditional (non)equivalence (see e. g. A05, A06, A07, A10). Register knowledge can then be modeled via statistical and (probabilistic) pragmatic approaches. In particular, projects A01, A05, and A10 seek to base the modeling of register variation on the same mechanisms proposed to underlie pragmatic phenomena such as scalar and manner implicatures. A05 in particular continues the development of a probabilistic pragmatic model in the Rational Speech Acts framework in which the choice between competing alternatives is a function of multiple potentially conflicting speaker goals; in Phase II, A01 builds on this work. A06 likewise hypothesizes that register variation, including language alternation and language mixing, may be modeled in a component of language outside of core grammar. Furthermore the project will apply statistical modeling to investigate the relations between linguistic phenomena and cultural parameters. Statistical modeling also plays a central role in A03, A07, and A09.

Importantly, the Area A projects plan regular exchange on the topic of modeling, to contrast these different approaches in terms of their explanatory adequacy and to explore the potential for their integration (see especially A04, A10).

4.2 Area B: Register and change

The focus of Area B of the CRC Register is the relationship between register and language change. The overarching research question addressed in Area B is:

QB. **How is register knowledge established and diffused in language change?**

In the founding application of the CRC, it was further specified by the three research foci QBi–QBiii. These research questions were planned as heading the three prospective funding periods: First, investigate grammatical means and situational-functional parameters that are involved in historical registers; second, examine to which extent emergence and change of register can be modeled via existing models; third, integrate register into theories of language change. We will now present the accumulated results of Area B for Phase I and turn to the research program for Phase II below.

***QBi*. How do registers emerge?**

In **Phase I**, the projects of Area B have extensively shown that diverse grammatical phenomena on all levels of the linguistic system vary in relation to the situational-functional parameters of their use. These relations were found in historically and culturally wide-spread contexts ranging from as far back in time as the Pharaonic period (c. from 2500 BC), through the Early Medieval and Late Medieval period (Old High German, 750–1050; Old Swedish, 1250–1450), and the Early Modern period (Old Lithuanian, c. 1500–1700; Old Latvian, c. 1550–1700) to present-day language use (Old to Modern English, before the year 700 CE to present day). Thus, the projects of Area B contribute to the CRC the central finding that register knowledge and the context-dependent use of linguistic variables are **historically pervasive**.

In all projects, we were able to show that texts of a particular genre need to be broken down into finer text sequences in order to understand register variation (e. g. homilies in B02; religious revelations, or monastic translations and commentaries in B04). In particular, embedded narrative passages came into focus (B03, B04). The parameters under investigation were manifold, from a formal/informal distinction (B01) to the domains and subdomains of Systemic Functional Linguistic's complex model of 'field, tenor, and mode' (e. g. social role relationship between addressor and addressee; B03, B04). The Hieroglyphic Egyptian texts of project B03 also highlighted the importance of multimodal visual parameters of graphic communication as indicators of registers. We also gained a much clearer picture as to what extent our historical texts show effects of an original source (e. g. in B04, for Old High German the Latin Vulgate effected not only the translations but also the commentaries; for Old Swedish, versions of Bridget's Revelations re-translated from Latin were less effected) or of a reference work (e. g. the older pericope book by Vilentas on Bretkes pericopes in B02). In sum, the projects of Area B gathered comprehensive evidence that historical register variation is a **multivariate and multidimensional** phenomenon in historical sources. Our study of emergent registers in diverse historical circumstances supports the assumption that language users adapt to their specific socio-cultural settings. These findings are in line with socio- and psycholinguistic results from Area C, e. g. C03 and C02.

Methodologically, the projects of Area B were able to demonstrate that even in circumstances of data scarcity, research into register knowledge and the context-dependent use of linguistic variables is not only feasible, but can be exerted with reliable corpus-linguistic, quantitative and qualitative measures. Through the

lens of historical texts, it was possible to access both the emerging register competence of individual authors (Old High German: Notker of St. Gall, c. 950–1022; Old Swedish: Bridget of Sweden, 1303–1373; Old Lithuanian: Johannes Bretke, 1536–1602; Old Latvian: Georg Mancelius, 1593–1654) as well as the registers in the process of conventionalization within larger language periods (Old/Middle/Late Egyptian; Old/Middle/Modern English). For this, the projects made use of existing corpora (*Thesaurus Linguae Aegyptiae*, B03; *Old German Reference Corpus* 'ReA', B04), but more importantly published freely available historical resources and guidelines for register research, such as an annotated corpus of Mancelius's Old Latvian Postil, an annotated corpus of the pericopes in Bretke's Old Lithuanian Postil (B02), and the Birgitta-Notker-Korpus (B04). These new resources operationalize diverse register-related categories, like text-functional span annotations (e. g., narrative, argumentative, devotional), or social role relationship, in addition to deep annotations of grammatical structures. Thus, the projects were able to make a decisive contribution to the development of instruments and procedures for the reconstruction of historical registers.

During Phase I, Area B stood in constant exchange on data, structures, and models as well as methods with projects from Area A and C, and will further pursue this cooperation. For example, B01 conducted an experimental investigation of the comparative alternation in English, including double comparatives, together with A07, and found an effect of register in that periphrastic comparatives received a higher rating in informal contexts than in formal contexts. They also looked at Greek comparatives and superlatives together with Fenia Karkalatsou (CRC fellow), results suggesting a cross-linguistic bias towards the analytic formation with increasing phonological, lexical and syntactic complexity. Project B03 is in intensive exchange with C02 on the questions of situational setting in communication with a primary focus on space and other text-external factors (such as clothing, gesture, mimics) as register-constituting parameters. This collaboration contributed to a deeper theoretical understanding of register knowledge as a cross-cultural and cross-temporal phenomenon. The bottom-up approach of identifying "pregisters" (for 'potential registers'), established by project A04 (Phase I), was successfully used by project B04 to break the register-proxy 'genre' into a more fine-grained analysis of situational-functional sequences. In terms of **cross-project topics**, the projects of Area B contributed to the CRC-internal paper on register terminology by the reading circle, the CRC's methods paper (Pescuma et al. 2023) as well as the annotation guidelines of the narration group (Lehmann et al. 2023).

Phase II. Based on the grammatical, contextual and methodological findings of Phase I, we feel confident to address the task of modeling the relationship be-

tween phenomena and parameters, central to the second funding period of this CRC. The specific focus of Area B in this regard will be the **modeling of register emergence and change**. As the findings from Phase I confirm the assumptions on historical register knowledge and the context-dependent use of linguistic variables that were the basis of the first funding period, Area B as a whole continues its originally proposed work plan and research questions.

At large, the **central research aspects** within Area B are the following: The interaction with, and the interdependence of syntax (B03, B04), morphology (B03, B04), lexical choice/semantics (B03, B04, B06); the function and formality of (parts of) texts (B03, B04, B06); topic and/or ideational domain of the text (B03, B04); target audience (B03, B04, B06); textual embedding of another person's speech (B03, B04); transfer of register patterns (B04, B06); and multimodality (B03). The B-projects will collaborate in the Cross-Project Topics Formality (B03, B04, B06), Multilingualism/Language Contact (B03, B04, B06), Narration (B03, B04), and Path Dependencies (B04), described in Section 6, in addition to joint activities between individual projects. Here, the projects will draw on findings from Areas A and C with regard to causal relations when binding specific linguistic phenomena to external parameters. In general, the projects will contribute a historical linguistic perspective to the discussions led in these groups and projects, and to the register models entertained by the CRC.

Research in Area B will continue to be based on **corpora** of historical written sources to gain indirect access to historical register knowledge. Continued projects will further exploit their resources established in Phase I and expand them by additional annotation layers as well as by including new texts where necessary. Expansions will be annotated according to the guidelines published in Phase I. In pilot studies, we successfully tested procedures for Phase II: automatic parsing of texts with or without manual correction (B04, B06), and multimodal annotation of image-text-compositions (B03). In the context of the CRC's general theme on **modeling** (see Section 3.2) in its second funding period, the projects B03 and B04 will focus on verbal models of the relationship between register and language change on the basis of quantitative and qualitative analyses. Project B06 will make use of probabilistic modeling in a combined approach with verbal modeling. We expect strong synergy effects between projects B03, B04, and B06, as for example Bible translations are an important source of data for both B04 and B06, or as B04 and B06 share an interest in the grammatical phenomenon of coordination.

As described in the founding application, Phase II was intended to pursue research question QBii. In accordance with the CRC's focus in its second funding phase, we adapted the wording to include 'modeling':

QBii. **How do we model register change over time?**
(Phase I: How do established registers change over time?)

From the perspective of language-user knowledge, we regard registers as sets of conventional form-meaning pairings (cf. Koch 1997), used and re-used by individuals and thus subject to change over time. Based on our observation of different historical and cultural settings for emerging registers, it is plausible to assume that changes in the contexts of use will entail changes in register patterns, e. g. through the increase in literacy and the subsequent diversification of text genres.

The task of modeling register emergence and register change in Phase II can rely on some fundamental insights and descriptive groundwork that were laid in Phase I. B01 – working diachronically on periphrastic and double comparatives in alternation with synthetic forms, and periphrastic progressives in alternation with non-progressive verbal forms – developed a unified mechanism for the stabilization of periphrastic alternations with three key ingredients: (a) a mechanism of competition between structures that differ in syntactic complexity, (b) patterns of Gricean reasoning leading to the emergence of not-at-issue meaning, and (c) an understanding of register effects as an important class of non-at-issue meaning. The analyses of the synchronic data in B02, B03, and B04 suggest that, in the case of Ancient Egyptian, the building of relations between phenomena and parameters must be modeled as a truly autochthonous process while register emergence in Old High German, Old Swedish, Old Lithuanian and Old Latvian is fundamentally characterized as cultural transfer, based on the established textual traditions of Medieval Latin or the German Protestant Reformation respectively. In this context, we expect relevant insights from our new project B06 which will investigate, among others, the re-emergence of registers in the early Romance languages. Our models also need to account for the data asymmetry caused by the dominance of religious genres in the textual transmission. In particular, the comparative evidence from Ancient Egyptian, Medieval and Early Modern German as well as Lithuanian allows us to triangulate the influence of the interrelated parameters 'elite literacy' vs. 'mass literacy' and 'administrative' vs. 'religious' contexts of emergence. We will also include the structural influence of contact languages as an integral, rather than additional, factor. More clearly than originally assumed, the results of Phase I point to the multilingual abilities of text producers, and, wherever writers are identifiable, their biographies prove the multilingual contexts of their text production. Awareness of multilingualism on the part of the text producer can also be found within representation of speech in fictional narrative texts (e. g. in dialogues). This is also attested in

historical contexts, cf. the Ancient Egyptian story of 'Wenamun', dated c. 1000 BCE (B03). Luther (B04), deeply embedded in the Latin tradition, translated the Bible from Greek and Hebrew; Bretke was a trilingual speaker of German, Old Prussian, and Lithuanian, and studied in Wittenberg in a German-Lutheran context. Accordingly, the aspects of multilingualism will receive a more prominent position in the working schedules, and the B-projects will take part in the CRC's new Cross-Project Topic Multilingualism/Language Contact to further discuss notions like 'structural influence of contact languages' or 'code-switching' (see Section 6.2).

Our projects approach register change from different perspectives: In projects B03 and B04, we complement the findings of Phase I with textual material of younger developmental language stages. B03 will trace down which parameters have an impact on the stability or, on the contrary, on the modification of established registers, by taking into account Ancient Egyptian texts with a long textual transmission, e. g. wisdom texts and teachings from the Middle Kingdom (c. 1950–1750 BC) in their adaptations in the New Kingdom (c. 1550–1070 BC). B04 follows a double comparative strategy: We check our Old High German findings from Phase I against the Early New High German material to be collected in Phase II, preparing the ground for the comparison of German and Lithuanian in Early Modern texts from the period of Protestant Reformation. Thus we motivate (a) the transfer of register patterns from the beginnings of the German vernacular tradition to Early Modern German and to Old Lithuanian and (b) the changes, adaptions, divergence, and conversion along the way. Project B06 investigates register change quantitatively from a language comparative perspective with a focus on the Indo-European language family. Its main empirical focus is the replacement of one form to express conjunction to another form. Our preliminary work on Latin has shown that the two forms during the change and also the change itself are register-sensitive. The goal of project B06 is to model interaction between language change and register across the Indo-European language family initially focusing on the expression of conjunction. In sum, the projects of Area B address the following essential questions: Which phenomena prove to be stable and which are more volatile? Is there a renewal or cyclical change of phenomena in long-term stable registers? Do core grammatical phenomena behave differently on various grammatical layers and differently from lexical or graphical ones? Do registers change differently in the context of a long-term continuous development, or in transfer? Do registers change in a different way when their change is brought about by new or changing parameters, like communicative functions, or situation types, or means of communication (e. g. religious pamphlets)? As methodological contributions to the field, projects B03 and B04 will

develop a formal model of annotations that expands the 'field-tenor-mode' framework of Systemic Functional Linguistics (Halliday 1978, Halliday & Matthiessen 2004, Neumann 2014a) – which was primarily developed for contemporary English – to appropriately cover historical texts. Project B06 designs quantitative models to capture the interaction between language change and register in a broad view of an entire language family.

QBiii. **How do we model the role of register in language change?**
(Phase I: How is register involved in language change?)

Ultimately, all projects within Area B plan to integrate phenomena and parameters of register emergence and change into comprehensive models of language change. With respect to cultural background and time-frame, all projects contribute their individual perspectives to the role of register in language change. The CRC's focus on register knowledge will lead us to look at phenomena of change from the point of view of the historical agents of change and their linguistic choices, but also the emerging, coexisting, and competing structures, both indicating intra- and interindividual register knowledge. Finally, the historical projects will also adapt theories of language change to include the role of registers. The different theoretical foci are outlined below.

In projects B03 and B04, the 'field-tenor-mode' model of Systemic Functional Linguistics serves as an important starting point for the modeling of language in context. Systemic Functional Linguistics' notion of register is embedded in a functional theory of language use. We will use the formal model of annotations developed for QBii and compare stages in the development of registers in individual languages. Based on these data, we elaborate a formal representation of change in context, thus creating a 'dynamic field-tenor-mode model'. The dynamic aspect of the model is created by interpreting the formal representations according to the communication-based model of language change in Zeige (2011, 2014). Here, at every instance of communication individuals select forms according to the context and their needs, but also according to the previous course of communication. Variation can be fed into the process at every instance of selection, with repeated similar selection contributing to register formation and change.

Of particular importance is B03's theoretical focus on multimodality. Aspects of this comparatively young linguistic field became particularly relevant when analyzing Ancient Egyptian text-image compositions, but can have a sustainable influence also on register research on historical languages using alphabet writings as part of complex page organization. Project B03 will integrate multimodality into the mode dimension of the 'field-tenor-mode model', and B04 will test the applicability to its early modern texts.

B06's specific theoretical contribution lies in integrating register within the model of language change of Kauhanen & Walkden (2018). This framework models the propagation of a new form through a population (S-curve), conditioned by a (cognitively derived) production bias. B06 will combine this approach with the game-theoretic model of Ahern & Clark (2017) which identified the novelty of a form as its central advantageous parameter. However, both approaches will benefit from integrating the role of register variation into the propagation of new forms.

In general, the theoretical approaches to language change within Area B will take into account the CRC's findings from Area A and C. Experimental studies will identify cognitive and social parameters of processing and linguistic choice that can be checked against our data, while typological studies will help to identify linguistic variables that are cross-linguistically tied to register variation and will hint towards frequent trajectories of grammatical change.

4.3 Area C: Register and cognition

Area C complements the foci of Areas A (register knowledge and grammatical knowledge) and B (register and language change) with a focus on cognitive aspects. The main question addressed in Area C is:

QC. **How is register knowledge acquired, represented and accessed?**

In **Phase I**, this central question was divided into three sub-questions: We focused on the influence of situational-functional parameters on register-related phenomena (QCi), on the acquisition of register knowledge in relation to linguistic awareness and experience (QCii), and on the perception of register differences (QCiii). As it turns out, register phenomena are relatively **pervasive** (see Pescuma et al. 2023): We observed effects of situational-functional parameters on register phenomena in language use, first and second language acquisition (C04, C05, C06), production and perception (C02, C03, C05, C06, C07). Effects emerged for phonetic variation in production (C02, C06), morphosyntactic and lexical-semantic aspects in comprehension (C03), in the L1 production of lexical and syntactic constructions in written academic discourses (C05) among others. First results highlight the need for more fine-tuning in modeling the relation between situational-functional parameters and register-related phenomena (C02, C03, C04, C07, see sub-questions below).

Exploring the pervasiveness of register phenomena, we relied on **experimental methodologies** (C02, C03, C05, C06, C07) and **linguistic corpora** (C02, C04,

C05, C06, C07). With respect to some of the challenges foreshadowed before Phase I (Lüdeling et al. 2022: 32ff.), we observed **intraindividual register variation** (of interest in the CRC) and substantial **interindividual variation** (see also Section 1.2.2.3 of the Phase I proposal). As planned, we took measures to tease apart intraindividual variation from interindividual variation. For experiments, we controlled interindividual variability through within-participant designs, random sampling, and extensive pre-testing and piloting of stimuli. Meta-data on participants (e. g. their language background, socioeconomic status, or personality traits) also captured interindividual variability against which we compared the effects of formality. We varied the role (e. g. interviewer vs. interviewee in C06; different addressees in C05) of the language user and the formality of the context (e. g. C02, C03, C07). To investigate lexical structures and inter- versus intraindividual variability in their corpus, C02 used analysis methods developed by A04; C04 made the interesting discovery of within-text variability – register shifted within one and the same text (e. g. from argumentative to narrative in academic essays), posing challenges for the analysis (surface forms vs. probabilistic methods).

During **Phase I**, we also began to **model register**. Using formal models, C07 extended Jackendoff's notion of conceptual representation (Jackendoff 2002) to include 'comm-sits' as formalized communicative situations relevant for intra-speaker variation. In the same vein, C03 added register indices in the social Coordinated Interplay Account of language processing (Münster & Knoeferle 2018). In language acquisition, the results informed theories of the acquisition of register knowledge via pragmatic entrenchment of linguistic behavior (e. g. C05). These results from Area C feed back into research Area A, as envisaged in the Phase I proposal. In Phase II, in order to refine our insight into the relation between situational-functional parameters and register phenomena, we continue to address the original questions with more in-depth modeling of register knowledge for each sub-question (QCi–QCiii). The specific focus will be on the relation(s) and interactions between situational-functional parameters, the properties of speaker and addressee and the observable register phenomena with an emphasis on specifying and **modeling the cognitive processes underlying these relations**. The C-projects rely on bottom-up verbal models (C02, C03, C05, C06), formal grammar approaches (C03, C07), statistical (C02, C05) and probabilistic modeling (C03), and a combination of these (C03, C06; see Section 3.2).

We organized collaboration to support inter-connectivity in Area C during Phase I. In Phase II, the projects of Area C will participate in the following Cross-Project Topics (see Section 6): Formality (C02, C03, C05, C06, and C07), Multilingualism/Language Contact (C03, C05, C06, and C07), Lexicon (C03, C05, and

C07), and Acquisition/Education (C05, and C03). In terms of project structure, C04 merges with C05, motivated by merging L2 and L1 perspectives for the study of late linguistic development in academic contexts. All other C-projects will continue into Phase II. Below we detail the results of Phase I for each sub-question (QCi–QCiii) and outline and motivate the research plan for Phase II.

QCi. **How do we model the effects of situational-functional parameters on register in production?**
(Phase I: How do situational parameters influence register-related phenomena in production?)

Addressing QCi in **Phase I**, we observed consistent effects of situational and functional parameters on linguistic phenomena in language production (C02, C05, C06, and C07; see also Cross-Project Topics Formality 6.1, and Multilingualism/Language Contact 6.2). (Perceived) **properties of the interlocutor** and **the social-role relationship between interlocutors** and the appraisal of **situational formality** proved highly relevant. We learned that even small differences in this respect may have an observable effect: In C02, the perceived social relationship between interlocutor and addressee and the functional and situational requirements (contributing to the perceived formality of a situation) affected fine phonetic detail in speech production (we saw, for instance, more variable and higher f0, more dispersed vowels for formal than informal parameters). C06 investigated phonetic properties in non-native addressee register and found that speakers adapt in a very fine-grained way to the perceived fluency of the (non-native) interlocutor, thus accommodating the addressee's needs in communication. Examining young adults' academic language development, C05 found that situational parameters affected written language production, with institutional setting and social relationships showing distinct effects on register-related phenomena: While the use of some syntactic constructions like passive, adverbial and attributive clauses was associated with the general distinction between private/public discourses , other linguistic phenomena like stance markers, grammar terms contributed to a finer register tuning most likely associated with the social role of the addressee (see also C02). Extending work to register differences in the lexicon, C02 in cooperation with A05 found differences in the use of adjectival modifiers depending on the intended formality of the experimental setup. As for language production in multilingual contexts, C07 found out that language choice is related to formality.

Some of these differences seem to be **motivated**. By motivation, we mean here that the variance corresponds to often subtle meaning differences, social meaning differences or the (assumed) affordances of the interlocutor (like those in

clear-speech situations; Roche 1998). We will closely collaborate with projects from Area A in the modeling of such differences. However, the picture is more complicated: We also found variation that is not easily motivated. C06, for instance, saw interesting (subtle but clearly observable) morpho-phonetic variation in two registers that cannot easily be explained by factors like meaning differences, social-meaning differences, or phonetic context like speech rate, etc. These pose challenges for modeling.

Not all linguistic levels were affected equally by the situational and functional parameters. In a few cases, situational formality had surprisingly little effect on language use, e. g. in the distribution of lexical structures investigated in C02. We found functionally motivated within-text register shifts (C04), and observed that speakers used lexical (vs. grammatical) variables more often to mark informal register (C07).

In **Phase II**, we will further analyze these multifaceted results and try to integrate them into production models. Because situational factors seem to play such a large role, C02 will immerse participants in actual physical locations by means of augmented reality (e. g. theater vs. funeral, or bus vs. library), to increase control over situational parameters. This substantially expands the methods portfolio of the C-projects. It also examines the effect of such locations on the variation of global phonetic parameters (f0/amplitude) of produced speech. C02 will statistically model their results. Conceiving of formality as a bundle of situational and functional parameters (including the addressee), C05 will focus on modeling the effects of (a)symmetric social relationships between speaker (L1 or L2 student) and addressee (instructor or peer) in oral academic communication as compared to written production. C06 will explore seemingly non-motivated (morpho-)phonetic variation that nevertheless varies with situational and functional parameters, like filler particles and liminal signs. The focus in C06 is on modeling such variation. C07 will look at the impact of language ideologies and social pressures on linguistic choices in multilingual contexts (Germany and Namibia) using discourse analyses of media and policy papers, sociolinguistic interviews, and corpus studies on productions in formal and informal settings.

QCii. **How do we model the acquisition of register knowledge? What is the role of linguistic experience and awareness?**
(Phase I: How is register knowledge acquired? What is the role of linguistic experience and awareness?)

In **Phase I**, our research on QCii contributed insights into the **acquisition of register knowledge and the role of linguistic awareness and experience** (C04,

C05, and C07). Register acquisition involves the accumulation of a diversified linguistic repertoire (in the spoken and written domain) and high metalinguistic and situational awareness which guide the appropriate and flexible employment of register resources according to culturally shaped and conventionalized communicative practices. The establishment of register resources through pragmatic association between linguistic and situational information requires conventionalized situated input and frequent language use in similar communicative situations. As speakers learn to navigate different types of communicative goals and situations in various speaker communities, the relevance of some situational-functional parameters for register acquisition and use changes over time and with the diversification of possible contextual configurations and the amount of allocated attention.

We investigated register acquisition during late adolescence/young adulthood in learners of German as a Foreign Language in C04 and young university students in Germany in C05. In C04, we found (among other things) that the parameters that seem to be relevant for classifying a register can be subject to transfer from the L1. C05 builds on the notion of **register flexibility**, where register flexibility is understood as an individual productive skill which pertains to the capacity of the speaker for fast and controlled adjustments of language use based on sensitivity for changing communicative goals and circumstances (Kaplan & Berman 2015, Qin & Uccelli 2020). A longitudinal study on the acquisition of academic register knowledge, register flexibility, and situational awareness revealed that first-year students exhibited a higher degree of variability in the informal contexts (email to a pupil/peer) reflecting adaptation to both institutional setting and social relationships to the addressee as compared to the formal ones (exam and tutoring class). We found that the use of stance markers and linguistic terminology or of passive and adverbial clauses were affected differently by the situational-functional parameters. From a developmental perspective, C05's findings at the first testing time point suggest that the establishment of register-distinguishing resources on the lexical (see also Cross-Project Topic Lexicon 6.3) and syntactic level proceeds at different paces and in close relation to educational processes. (Data from the second and third testing point will enrich our insights on the developmental trajectory shaped by the accumulation of specialized resources and increasing metalinguistic awareness.) Further, we found effects of speakers' experience and familiarity with a particular situation and thus of speakers' awareness of the communicative affordances, supporting the view of the acquisition of register knowledge as entrenchment of pragmatic associations with situated language use (Schmid 2020). The Entrenchment and Conventionalization Model thus predicts certain asymmetries in the register repertoire

of L2 learners due to limited experience with a variety of culturally established communicative situations and conventionalized linguistic practices in the target language. C04 found that multiple register shifts (e. g. narrative passages in academic essays) are produced within one and the same text type in advanced learner language. Project C04 developed methods for assessing contemporary texts written in academic registers by second-language learners of German. It investigated the abstract variable of noun modification (how a noun is modified). It used different layers of structural annotation, including dependency and morphological information. Awareness of register was observed in studies by C07, demonstrating that the use of lexical borrowings is associated with in-group membership and local solidarity and informal situations. C05 expanded the investigation by including further speaker-specific variables such as personality traits and motivation with the goal of distinguishing their effects on the development of register flexibility in higher education (see also Cross-Project Topic Lexicon 6.3 and Acquisition/Education 6.4).

Regarding development and acquisition, **in Phase II**, the C-projects will contribute substantially to research on register and **multilingualism/language contact** (also see Section 6.2): C03 looks at adolescent heritage speakers' language comprehension against the background of differences in linguistic experience (use of heritage language with peers and of the majority language in formal situations like school) with a view to refining models of register processing. In order to address different aspects of the multilingual space, C04 and C05 will merge into C05 and continue to explore the acquisition of register knowledge in L1 and L2 German in the context of higher education. The new C05 will extend the currently constructed longitudinal written corpus (ReFlexAE) to include L1 and L2 oral interactional data. In modeling the development of register flexibility of young adults, C05 aims at refining the construct of metalinguistic awareness with regard to the dimensions of situational sensitivity, register awareness and metalinguistic knowledge. Similar to C03, C06 contributes to the language contact topic in modeling, too, by further exploring the non-native addressee register data collected in Phase I. C07 ties in with the topic via its focus on the impact of language ideologies on register distinctions in multilingual contexts.

QCiii. **How do we model register in perception?**
(Phase I: How are register differences perceived?)

In **Phase I**, we gained first insights into the perception of register differences which were systematic despite considerable interindividual variability. In C03, consistent ratings of stimuli corroborated the perception of formality differences.

For example, words and sentences classified as 'formal' were rated higher on a formality scale than their informal counterparts. In C02, stimuli perception of formal vs. informal situations was tested regarding the perceived personality of the interlocutor judging 15 personality attributes on a seven-point scale (e. g. incompetent-competent, dominant-submissive, formal-informal, relaxed-tensed), and on potential occupations the person might have. The results of the ratings showed that the different personae elicited clearly different formality perceptions. For C07, in an open guise study, participants identified the supposed addressee (teacher vs. friend) of texts spoken in Standard German compared to Namibian German (characterized by lexical borrowings or specific grammatical phenomena). The perception of formality was influenced by the social salience of linguistic forms: Listeners associated lexical Namibian-specific phenomena with an informal context (speaking to a friend) and standard German variants with formal contexts (speaking to a teacher); specific grammatical phenomena were associated with both kinds of addressee. Research of the A-projects (A05 and A07) provided further insight into question QCiii of Phase I by showcasing that the choice of precision level in context affects perceptions of the speaker: Precise forms elicited higher ratings on status attributes like 'intelligent' or 'confident'; less precise forms, by contrast, elicited higher ratings on solidarity attributes like 'likeable' or 'laidback', and lower ratings for antisolidarity attributes like 'uptight'. Linking back to QCi, these associations were modulated by situation parameters (e. g. a court setting vs. selling a car).

Building on the findings from Phase I, namely that register differences are perceived using a variety of grammatical and lexical means, as well as, for example, the social salience and formality of linguistic forms, we will shift to the question of how to model these register differences in perception in **Phase II**. We will assess the generalizability of our Phase I findings by exploring further parameters influencing the choice of words, aspects of the voice and phonetic detail. To assess register awareness and also to capture interindividual variability (with the goal of differentiating it from intraindividual variability), all C-projects together with A03 and A09 will develop a **Debriefing Questionnaire** which targets speakers' perceptions of different situational-functional parameters and how such perceptions relate to speaker's language use. C03 contributes to the topic of modeling formality in perception by examining the effects of register (in)congruence on heritage speakers' language comprehension with a focus on continuing the comparison of how grammatical subject-verb and semantic verb-argument congruence as part of standard language grammar is processed versus congruence between situation formality and register-related phenomena. C03 plans to collaborate on modeling of language processing with project C07, adding 'comm-

sits' to the social Coordinated Interplay Account (Münster & Knoeferle 2018) to see how predictions of parameter-register (in)congruence effects would change when representations are formal and enriched with more register knowledge.

5 Methodology

As described above, many parameters concurrently influence the production and comprehension of linguistic phenomena. During Phase I of the CRC, we employed a wide range of experimental and corpus-based methods to specifically investigate situational and functional parameters and their relation to linguistic phenomena. The different research questions, theoretical perspectives, and phenomena that characterize our investigation require different methods. Although the phenomena span many different research domains, from phonetics to pragmatics, and from historical linguistics to psycholinguistics, we find common register-related themes across different projects. This generates interesting discussions and fosters methodological synergies between the projects in our CRC across all areas. In a methodological article resulting from a CRC-wide effort (Pescuma et al. 2023), we examined situational-functional parameters, linguistic phenomena, and the relations between them. Specifically, we looked at formality, as indexed by social roles and relations (A05, B03, B04, C02, C05, C07) and other context-related parameters (A04, A05, B03); grammatical phenomena (A04, A07, B04, C05, C07); lexical choice (A05, C03, C04, C05, C07); pragmatic/rhetorical devices (A01, B04, C02); non-linguistic variation (C03, B03); social meaning/inferences (A05, C07); mental representations/models (A05, C03, C07).

In this article and more generally in the CRC, we furthermore identified various methodological challenges related to the multidisciplinary perspective that we adopted and the intrinsic complexity of register. As far as **experimental approaches** to comprehension and perception are concerned, we employed methods such as matched guise, open guise, rating studies, eye-tracking during visual scene perception and during reading, and newspaper correction tasks (A05, A07, C03, C07). Some of the major challenges we faced in this respect concerned the operationalization of different parameters. Consider, among others, the operationalization of formality – we tried to find lexical items representing different levels of formality to create a mismatch situation in C03 using rating tests for the items in question. C02 also considered formality but focused, in their rating tests, on the formality traits of an interlocutor (exemplified by clothing, hair style, etc.). Hence, even though both projects operationalized formality, the materials resulting from these ratings differ greatly. Formality can be a property

of either linguistic phenomena or situational parameters and is itself certainly multi-layered and complex. This leads to the fact that experiments – even if they seemingly deal with the same parameter – cannot easily be compared and results need to be interpreted very carefully. This may seem trivial once the problem is identified. A CRC with its many different approaches and wide range of expertise is the ideal environment for finding and, eventually, resolving such situations. In the case of formality, we tackle the problem in the Cross-Project Topic Formality (Section 6.1).

Many projects rely on **corpus resources** of some kind (spoken, written, multimodal; synchronic, diachronic, etc.; A01, A03, A04, A05, A06, B03, B04, B06, C02, C04, C05, C06, C07). Some of the corpora already exist and are re-used, others such as the longitudinal ReFlexAE corpus of situated variation in academic contexts (C05) are specifically constructed in the CRC. We annotate the data in different ways and use a wide variety of manual, automatic and statistical approaches to find and describe register variation. But it has become more and more obvious that the relationship between a text (as a product) and the register knowledge of the person that produces the text is not necessarily straightforward. Any given study can only consider a finite number of situational and functional parameters on some specific level of granularity. In essence, this means that for any variation encountered, (a) other (perhaps still unidentified) parameters might be at work and (b) a different level of granularity might be required. For example, several projects identified the shared problem of how to approach the fact that any spoken and written text can consist of several registers (see Section 2). A story can be told as anecdotal evidence in an argumentative essay (C04) or a monk may intersperse his retelling of the gospels with religious advice (B04). The boundaries between the different registers within one text are not always clear. This leads to a methodological problem in all corpus-based studies that equate a text with a (typically high-level) register. We have explored different ways of dealing with this issue: Gohar Schnelle (B04) uses functional cues in Otfrid of Wissembourg's Old High German gospel harmony to subdivide the text into meaningful components (Schnelle et al. 2023). A04 clusters Web texts into what they call pregisters (from 'potential registers'). The members of the Cross-Project Topic Narration (see Section 6.5) explore a multi-layer annotation scheme (Lehmann et al. 2023). Other projects take great care to control parameters when acquiring the texts themselves, thus trying to keep as many parameters as possible constant while varying only one parameter (such as the L1 of the interlocutor in C06, attire of the interlocutor in C02, see also Wiese 2020). Additionally, all experimental and corpus-based studies should likely take into account path dependencies. While path dependencies occur in many situations and for many

reasons (some of which have nothing to do with register), they play a large role in register and need to be taken into account. One of the cross-project topics will deal with path dependencies (Section 6.6).

[…]

We strongly believe that methodological decisions must be transparently documented. Because it is not typically possible to publish corpus descriptions, materials, or annotation guidelines, we decided to establish the white-paper series REALIS (Register Aspects of Language in Situation). This series developed into a peer-reviewed journal, which publishes research articles as well as guidelines, or register-related MA- or BA-theses.

6 Cross-project topics

During the course of Phase I, a range of topics emerged as subject of interest within and across Areas A, B, and C. They sparked cross-project initiatives – regular discussion group meetings, co-organized workshops, joint research, and publications. In Phase II, we will install cross-project topics as an intersecting layer of cooperation within the CRC. They institutionalize an integral benefit of the CRC structure: Researchers from different theoretical backgrounds and areas of the CRC, working on different languages, register phenomena, parameters, and their relations come together to work on common themes. There will be seven cross-project topics which are described individually in the following subsections.

[…]

6.1 Formality

Many studies on register (including studies performed in this CRC) distinguish in some way between formal and informal registers. Formality seems to be an intuitive concept that can be used without much explanation. However, the situation is considerably more complex than it appears: There is a fundamental distinction between formality as a situational parameter (the way a speaker is dressed as an indicator of situation formality, i. e. in formal vs. informal attire, see C02) and formality as a linguistic phenomenon pertaining to linguistic properties of (in)formal language usage, e. g. *to detain* vs. *to grab* (C03). Secondly, the notion of formality on each of these dimensions is in itself complex and multi-layered.

The distinction between formal situations and formal language is exemplified by recognizable mismatches between the two notions. Language users easily

identify register violations based on the perceived mismatch between informal language use and the linguistic expectations raised by the appraisal of a situation as a formal one and vice versa (C03). The concept of situational formality is based on conventionalized, culturally sanctioned and agreed upon forms and customs of social behavior on official occasions. At least in Western societies, the formality of situations is generally defined in terms of public and private spheres of social interaction. Still, there appear to be at least three aspects that all contribute to the multi-dimensional concept of formality: hierarchy (authority or power gradients in social groups; social relationships between interlocutors), institutional setting, and communicative intention (functional goal). Formality thus includes aspects related to social distance, proximity and acquaintance between interlocutors (Koch & Oesterreicher 1985), number and constellation of interlocutors (Bell 1984), and the level of staging of communication situations based on culturally distinct world knowledge (Steger et al. 1974, Schank & Schoenthal 1976).

In the CRC we conceive of **situational formality** as a complex construct involving appraisal and integration of information from perceptual, conceptual, and linguistic sources. In Phase I, several projects have investigated how information about institutional setting (public/private), or the social relationship between interlocutors provided by a linguistic context (A05, A07, B04, C03, C04, C05, C07) and/or visual information (B03, C02, C03) influences register adaptation and found effects on phonetic, lexical and syntactic levels. The research of A01, A05, B03, C04, and C05 revealed how the assessment of communicative goals, i. e. brevity, knowledge transfer, argumentation, persuasion and involvement relates to production and interpretation of register-sensitive expressions in different discourse types or within single texts.

In determining the **formality of language** we need to consider the functional association of language use with distinct types of social practices. Categorizing language as formal in this sense focuses on specific types of linguistic expressions which are used to achieve an interpretation that is maximally independent of contextual information by minimizing ambiguity and maximizing objectivity and precision (Heylighen & Dewaele 1999). This premise innocuously applies to formulaic usage in ritualized practices such as esoteric languages in Saamaka (A08), or to highly specialized resources for the purposes of scientific or professional communication (C05) but also to other domains of social interaction requiring particular types of knowledge, i. e. in the religious domain (A01, B04). Ambiguity reduction also appears to be at the core of politeness systems which comprise typologically diverse but univocal means of linguistic behavior acquired and used according to culturally established and highly conventionalized communicative practices (A06, A08).

Notwithstanding the different approaches to formality and its operationalization in the CRC projects, the question of the criteria guiding the categorization of situations and language choices remains a central one in register studies. The relation between formality of language (phenomena) and formality of situation (parameters) represents a key issue for modeling register knowledge in Phase II, specifically with respect to inter-cultural but also intra-cultural differences (socioeconomic status, multilingualism, generation/age) in the acquisition, representation and processing of register knowledge. The parameters relevant in constituting situational formality, and the phenomena encoding formal language, as well as the relation between them, can be subject to change (B03, B04, B06).

[…]

6.2 Multilingualism/Language Contact

While multilingualism is the historical and present norm for most of the world (Trudgill & Cheshire 1998, Edwards 2007), research on register variation typically focuses on monolingual speakers. In Phase II of the CRC, we will explore the role of language choice and language mixing as a marker of register in synchrony and diachrony and in diverse contact situations, including heritage speakers (A03, A06, A08, B03, B04, C03, C07). Moreover, C05 will look at L2 acquisition of register, thereby deepening the research focus of the first phase of the CRC (QCii: How is register knowledge acquired?).

With respect to language choice and mixing, we draw on observations from current work within the CRC. C07 has shown that language mixing is a marker of informal registers in Namibian German: Specifically, "language" separation acts as a marker of formal registers. By contrast, lexical borrowing from English and Afrikaans takes place in informal communicative situations. Similar effects were noted for language choice in first-encounter contexts in Mauritius, number marking in Bislama and Yucatec Maya (A02, A06), the latter also in Old High German, Old Swedish, and Old Lithuanian (B02, B04). Thus, language mixing (including hybridization of languages) will be one important scenario to be investigated in Phase II. Several projects (A03, A08, B04, C07) will explore the hypothesis that languages themselves or the use of language mixing or hybrid lects can function as registers, i. e. intraindividual choices of different languages or language mixing may be determined by different situational-functional parameters, see, for instance, the comm-sit model (Wiese 2021). In other words, language purism and language mixing can be viewed as register markers and are inter-connected to the parameter formality: Keeping within the borders of socially constructed "pure" languages marks formal, transcending such borders informal production.

In addition, language mixing or even hybrid lects and their role for register variation pose a challenge for formal models of linguistic knowledge. The projects A03, A06, A08 and C03 aim at integrating register variation in multilingual contexts into models of grammar, and will explore various modeling options as stated above.

With regard to L2 acquisition of registers, it is to be expected that learners initially have a more limited inventory of registers (Gilquin & Paquot 2007, Deshors 2015). This holds, for instance, for academic registers in German, if one compares L2 learners to students who have acquired German as their L1. Also with regard to the relation between register-related phenomena and situational-functional parameters, different socialization may lead to differences in register knowledge between students with L1 and L2 German. Some of the results from the first phase of project C05 show that parameters such as familiarity with communicative situations and primarily the allocation in the public versus private sphere have a direct effect on language use. What is interesting is the interdependence between register knowledge in the L1 and L2, which can show up linguistically in transfer phenomena (C04), but also in the accumulation and use of register resources in instructed contexts. In the CRC, we will continue to analyze register knowledge in advanced learner language with a special focus on differences between L1 and L2 development in written and spoken academic registers. A similar situation emerges with heritage language speakers who can be fluent in both their heritage and majority language, but different social situations likely determine their language use: Typically the majority language is learned and used across formal and informal situations (e. g. in schools, at the job, but also with friends and siblings), but the heritage language is mostly acquired and used in informal situations (e. g. with family and peers).

[…]

6.3 Lexicon

In Phase I, many projects across all three areas of the CRC (A06, A07, B01, B02, C03, C04, C05, C07) investigate situated lexical variation as a register-related phenomenon in production and comprehension in different languages or historical stages. Although approaching lexical phenomena from these very different perspectives, we find converging evidence on the intricate relations between lexical phenomena and situational-functional parameters within and between languages and identify focal topics of investigation concerning the acquisition, use and change of register knowledge in the lexical domain.

Register knowledge manifests itself most prominently on the level of lexical choice. The association of meanings and functions with the contexts of usage is reflected in the different types of conceptual knowledge underlying the abstract semantic representation of lexical items (Wu & Barsalou 2009, Jackendoff 2002). Situational and introspective knowledge concerns prototypical participants, location, time, manner, affect, positive or negative evaluation; taxonomic relations such as synonymy may also include register-related information, i. e. (near) synonyms such as *essen* 'eat', *tafeln* 'dine', *futtern* 'nosh' are associated with more formal or colloquial registers (Pescuma et al. 2023). Research on the processing of register violations (mismatches between lexical phenomena and situational-functional parameters of the context) shows that they are closely linked to semantic violations but differ from the effects of morphosyntactic violations which emerge more robustly (C03).

While a high proportion of lexical items seems to be neutral with respect to register distinctions (Ravid & Berman 2009), some lexical phenomena show strong associations with salient situational-functional parameters or configurations thereof (see A05 and C02 on adjectival modification in contexts of varying formality). C05's study on the acquisition of academic registers shows that the production of stance markers and grammar terms differs in contexts appraised as private communication or public academic activity, respectively. Additionally, stance markers exhibit high selectivity for the parameter social relationship in the informal contexts. From a developmental perspective, the higher situational adaptation in the use of stance markers exemplifies the difference between the availability of already established lexical markers of involved discourse and the still evolving resources of academic language during late linguistic development.

Specialized vocabularies (scientific, professional, religious), stance markers, politeness expressions but also code-mixing are associated with specific contexts of acquisition and use and thus constitute a salient part of register-specific resources identifiable by the members of different speaker communities. This issue spurs collaboration with the Cross-Project Topic Multilingualism/Language Contact (Section 6.2): While lexical borrowings are more associated with informal contexts, language separation seems to be indicative of formal communication. Moreover, while lexical phenomena seem to be susceptible to strong parameter distinctions, grammatical phenomena exhibit less register differentiation (C07).

Parameters that lead to specific lexical choices may be subject to variation across varieties of the same language. A07 shows that multiple modal constructions (e. g. *might possibly*) are perceived as more acceptable given certain situational-functional parameters in British but not American English. Importantly, the same parameters also seem to affect lexical phenomena in the domain of

comparatives and superlatives in different languages such as English and Greek (B01, Karkaletsou & Alexiadou forthcoming). Even within the same language and under equal production parameters, i. e. identical external settings and communicative goals we have seen high lexical variability, yet this variability emerges to varying degrees in different linguistic layers – while the morphological structure of nouns varies to a very high degree, syntactic categories such as the types of parts of speech used or the dependency types occurring were found to be much more consistent across parameters and speakers (C04, Shadrova et al. 2021; cf. C07).

These findings from Phase I raise relevant questions for Phase II concerning how to model register-related phenomena concerning different layers of the lexicon and with respect to the accumulation and use of lexical resources in L1 and L2 register acquisition (Section 6.4) and specifically in language contact situations (Section 6.2). Projects A01, A04, A08, C03 and C05 will, among others, address the following questions in Phase II: How do lexical variables and their variants compare to structural variables and variants in register production and comprehension? How do we model any such differences in grammar and lexicon? How can we bring together our theoretical and empirical findings to better understand and model the cognitive processes underlying the establishment, consolidation and change of register information in the mental representation of lexical items?

6.4 Acquisition/Education

In a multilingual world, the study of register development calls for a careful consideration of the developmental factors and situational parameters influencing and differentiating the accumulation of register knowledge in L1, L2 and any further languages. From the perspective of the Entrenchment and Conventionalization Model (ECM; Schmid 2020), register knowledge develops through pragmatic strengthening of context-dependent meanings and functions through conventionalized linguistic input and through frequent and uniform usage. The processes of associative learning are enhanced by the similarity of communicative contexts speakers experience (and represent as situation types) in the course of their socialization. Therefore, the availability of situated input and familiarity with a large array of communicative situations are considered to be crucial conditions for the establishment and flexible access to register resources in any acquired language. Linguistic experience will be investigated as a factor influencing L2 register acquisition in adolescence and young adulthood (cf. C03 on register perception of heritage speakers and C05 on advanced L2 learners of German). In Phase I, we scrutinize the effect of situational and functional parameters on the

acquisition of written academic discourses and find that lexical and syntactic phenomena are differently affected by formality parameters such as public/private setting, social relationship or communicative function in L1 (C05) and L2 (C04). In Phase II, the investigation will be expanded to oral academic discourses, the new C05 advancing an integrated model of the development of register flexibility in L1 and L2 young adults.

Models of register development (Ravid & Tolchinsky 2002, Berman 2018) highlight the role of written language acquisition for the establishment of diversified register repertoires in the course of school and higher education. During the instructed acquisition of new knowledge domains, L1 and L2 learners accumulate specialized linguistic resources associated with specific scientific topics and conditions of use. Explicit teaching and learning strategies thus play a crucial role in the acquisition of academic language. On the one hand, academic language as a primary form of a formal register acquired at school serves communicative requirements in institutional settings. On the other hand, it provides the foundation for the development of languages for special scientific or professional purposes during higher education and vocational training. The processes of register development during adolescence and young adulthood are thus associated with increasing metalinguistic awareness and register flexibility (Kaplan & Berman 2015). These skills enhance communicative efficiency in professional and informal discourses, i. e. through linguistic adjustment to different types of social relationships between speaker and addressee concerning in- and outgroup membership in speaker communities or social distance. These issues will be investigated by A09 in Canarian Spanish and by C05 for L1 and L2 academic German. Education-related parameters such as type and levels of education or (non)academic family background have been shown to be indicative of the socioeconomic status of interlocutors and of their social persona (A09, C05). Type and level of education are associated with the range and flexibility of linguistic adaptation (utilization of specialized resources) in production and comprehension of register variation. These aspects play an important role, albeit in different contexts, in projects A09, C03, and C05, and are also relevant for the assessment of register phenomena in historical texts associated with instructional purposes (B04). Since level of education is considered as a parameter and controlled for in the experimental designs of several CRC projects, collaborative efforts will strengthen our methodological expertise in the study of register acquisition and use.

6.5 Narration

Various register models take goal orientation or the ideational communicative function as one of the key parameters for characterizing registers due to their functional nature (Biber 1994, Neumann 2014a, Hasan 2014, Biber et al. 2021). One of the most widely used and omnipresent communicative functions and/or text strategies is *narration*. As a pan-human communicative device, it is found in oral, signed, and written form, across time, cultures, and languages. It therefore comes as no surprise that several projects in the CRC find narration in one way or another in their data (B03, B04, C05). In some texts, we observed that narration is the overarching communicative function while in others, narrative passages are embedded in and linked to further goals, see similarly Biber et al.'s (2021) functional analysis of more general communicative purposes in oral communication. This approach shows how description is one of the more frequently occurring purposes for speech in conversations at a very granular level, often only spanning a few utterances and interwoven with other purposes. In Phase I, a working group (A06, B02, B03, C04) has set out to define the common criteria found in narrative texts or text passages, including the multi-layered structure pointing towards mediacy, i. e. perspectivization, and sequentiality of events (Zeman 2016, 2018, Martínez 2011, Martínez & Scheffel 2009, Genette 2010, Grabes 2014, Hühn 2013, Lahn & Meister 2016). We devised an operationalization for recognizing and annotating narrative passages in narrative and non-narrative texts (Lehmann et al. 2023). In Phase II, we want to investigate the extent to which register variation and narration are independent or intertwined. Are there different types of narration? And do these types of narration mark different registers? Or is narration a universal register displaying similar properties across languages, times, contexts, and text types?

6.6 Path Dependency

One aspect of modeling register effects (among others) that has to be taken into account in any model working with production data is the fact that a given linguistic phenomenon is not always distributed evenly throughout a text. Many phenomena tend to occur in clusters, as has been described for specific words as well as for structures such as passive sentences or particle verbs (Szmrecsanyi 2006, Gries 2005). There are different causes for this phenomenon and it has been analyzed in different ways, e. g. in statistical accounts by dispersion measures, in corpus linguistic studies as persistence, in phonetics as alignment, or in psycholinguistics as effects of (self-)priming; in Systemic Functional Linguistics it

is subsumed in the textual dimension. Underlying all of the different accounts is the finding that the occurrence of a linguistic phenomenon may depend on earlier co-text (by the speaker or interlocutors, or other texts that were copied). We will use the term *path dependency* here.

We found path dependencies in (at least) two areas. Firstly, we see a thematic development in a text, i. e. what is said earlier influences/constrains what will be said later and how it will be said. For example, this happens frequently in the marking of politeness (McCready 2019). Imagine a situation in which two German-speaking people of roughly the same (middle) age meet for the first time and it is unclear whether the more formal *Sie* or the more informal *du* 'you' should be used to address each other. The first interaction may be crucial – the decision taken there will set not only the pronoun but also the tone for the rest of the conversation, influencing other linguistic structures. In another example, Baayen (2001) attributes the finding that the number of definite articles was higher in the second half of a text (than in the first) to the fact that the relevant referents were already introduced. Secondly, there is the general cognitive mechanism of (self-)priming: Specific structures are activated and will be used more frequently. For example, C04 found bursts of complex verbs, passive sentences, etc. (Shadrova et al. 2021).

Path effects are relevant for us in two ways: They have to be taken into account in methodology and modeling. In corpus studies, we cannot assume that every occurrence of the phenomenon under investigation can be counted in the same way. We will weigh the possible repercussions for statistical analysis. Path effects may also have effects for the acquisition of register – the association of specific structures with specific textual functions can be learned as indicative for a given register. Projects A03, A07, B04, C06 and INF will cooperate on the investigation of path effects and their relevance for register models.

6.7 Complexity

One key area of interest for this CRC is the relationship between complexity and register (see A04, A06, A07, A08, A10). Linguistic complexity is multi-layered. Following Kortmann & Szmrecsanyi (2012), we distinguish between global complexity (of a language/dialect) and local complexity (of a specific domain). Furthermore, there are several sub-types of complexity as well as numerous complexity measures. To date, and in spite of the extensive discussion on these issues (Lu 2011, McWhorter 2011, Miestamo 2004, Weiß 2017), many open questions about the theoretical models, cognitive processes, and influencing factors behind complexity have yet to be explored, such as discourse status (Arnold et al.

2000), medium (Biber & Gray 2010, Verhoeven & Lehmann 2018), and language development (Weiss & Meurers 2019), to name a few that we found particularly interesting. Complexity pertains to all areas of language analysis, from phonological weight and morphological structure to the manifold levels of syntactic embedding and information packaging (Weiss & Meurers 2019) and is intangibly related to aspects of processing (Gibson 1998, 2001) and acquisition. Complexity is also controversially discussed in work on creole languages (McWhorter 2012, Aboh & deGraff 2016) and heritage languages (McWhorter 2011, Scontras et al. 2018). Recent research has shown that a number of linguistic areas involve different types of complexity (e. g. for clause vs. phrase level, see Biber & Gray 2010, and for center vs. peripheral embedding, see Karlsson 2007 and Verhoeven & Lehmann 2018), for recursivity in particular see Sakel & Stapert (2010), Kornai (2014). This literature, as well as our own research shows clearly that different registers exhibit different degrees of complexity on the various levels and that we often observe trade-offs between the levels. It is, however, not always clear what the exact theoretical relationship is between register and the individual facets of complexity. How do the situation's communicative needs and intentions affect our choice of one complexity measure over another? How do surface complexity and processing relate to one another, and what difference does the type of complexity make to this relationship? Thus, in Phase I a working group (A04, A06, B02, C04) set out to understand what types of complexity are involved in register and organized a workshop on this topic in 2021 (see https://sfb1412.hu-berlin.de/complexity-and-register/). Our joint interest and focus for Phase II is to investigate the relationship between complexity and register.

Funding information

The research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – CRC 1412, 416591334.

References

Aboh, Enoch & Michel deGraff. 2016. 401A Null Theory of Creole Formation Based on Universal Grammar. In *The Oxford Handbook of Universal Grammar*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199573776.013.18.

Acton, Eric K. & Christopher Potts. 2014. That straight talk: Sarah Palin and the sociolinguistics of demonstratives. *Journal of Sociolinguistics* 18(1). 3–31. DOI: 10.1111/josl.12062.

Adger, David. 2006. Combinatorial variability. *Journal of Linguistics* 42(3). 503–530. DOI: 10.1017/s002222670600418x.

Adger, David & Jennifer Smith. 2010. Variation in agreement: A lexical feature-based approach. *Lingua* 120(5). 1109–1134. DOI: 10.1016/j.lingua.2008.05.007.

Adli, Aria, Elisabeth Verhoeven, Nico Lehmann, Vahid Mortezapour & Jozina Vander Klok. 2023. *Lang*Reg: A multi-lingual corpus of intra-speaker variation across situations*. Version 0.1.0. Zenodo. DOI: 10.5281/zenodo.7646320.

Ahern, Christopher & Robin Clark. 2017. Conflict, cheap talk, and Jespersen's cycle. *Semantics & Pragmatics* 10, 11. DOI: 10.3765/sp.10.11.

Alexiadou, Artemis. 2020. Compound formation in language mixing. *Frontiers in Psychology* 11, 1021. DOI: 10.3389/fpsyg.2020.01021.

Alexiadou, Artemis, Foteini Karkaletsou, Izabella Rosa Malta, Dan Morris, Thomas McFadden & Giorgos Spathas. Submitted. Towards a unified mechanism for the diachronic development of periphrastic alternations.

Alexiadou, Artemis & Terje Lohndal. 2018. Units of language mixing: A cross-linguistic perspective. *Frontiers in Psychology* 9, 1719. DOI: 10.3389/fpsyg.2018.01719.

Alexiadou, Artemis & Vasiliki Rizou. 2022. The use of periphrasis for the expression of aspect by Greek heritage speakers: A case study of register variation narrowing. *Register Studies* 5(1). 82–110. DOI: 10.1075/rs.20022.ale.

Arnold, Jennifer E., Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1). 28–55. DOI: 10.2307/417392.

Baayen, R. Harald. 2001. *Word frequency distributions*. Dordrecht, Boston & London: Kluwer. DOI: 10.1007/978-94-010-0844-0.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2). 145–204. DOI: 10.1017/S004740450001037X.

Beltrama, Andrea, Stephanie Solt & Heather Burnett. 2022. Context, precision, and social perception: A sociopragmatic study. *Language in Society* 52(5). 805–835. DOI: 10.1017/S0047404522000240.

Berman, Ruth A. 2018. Language development and literacy. In Roger J. R. Levesque (ed.), *Encyclopedia of Adolescence*, 2093–2103. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-33228-4_19.

Biber, Douglas. 1994. An analytical framework for register studies. In Douglas Biber & Edward Finegan (eds.), *Sociolinguistic perspectives on register* (Oxford Studies in Sociolinguistics), 31–58. Oxford: Oxford University Press.

Biber, Douglas. 1999. A register perspective on grammar and discourse: Variability in the form and use of English complement clauses. *Discourse Studies* 1(2). 131–150. DOI: 10.1177/1461445699001002001.

Biber, Douglas, Jesse Egbert, Bethany Gray, Rahel Oppliger & Benedikt Szmrecsanyi. 2016. Variationist versus text-linguistic approaches to grammatical change in English: Nominal modifiers of head nouns. In Merja Kytö & Päivi Pahta (eds.), *Cambridge Handbook of English Historical Linguistics*, 351–375. Cambridge: Cambridge University Press.

Biber, Douglas, Jesse Egbert, Daniel Keller & Stacey Wizner. 2021. Towards a taxonomy of conversational discourse types: An empirical corpus-based analysis. *of Pragmatics* 171. 20–35. DOI: 10.1016/j.pragma.2020.09.018.

Biber, Douglas & Edward Finegan (eds.). 1994. *Sociolinguistic perspectives on register* (Oxford Studies in Sociolinguistics). Oxford: Oxford University Press.

Biber, Douglas & Bethany Gray. 2010. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes* 9(1). 2–20. DOI: 10.1016/j.jeap.2010.01.001.

Biber, Douglas, Bethany Gray, Shelley Staples & Jesse Egbert. 2022. *The register-functional approach to grammatical complexity: Theoretical foundation, descriptive research findings, application* (Routledge advances in corpus linguistics). New York: Routledge. DOI: 10.4324/9781003087991.

Blei, David M., Andrew Y. Ng & Michael I. Jordan. 2003. Latent Dirichlet allocation. *of Machine Learning Research* 3. 993–1022. https://dl.acm.org/doi/10.5555/944919.944937 (13 December, 2022).

Blochowiak, Joanna, Cristina Grisot, Stéphanie Durrleman & Christopher Laenzlinger. 2017. Introduction. In Joanna Blochowiak, Cristina Grisot, Stéphanie Durrleman & Christopher Laenzlinger (eds.), *Formal models in the study of language: Applications in interdisciplinary contexts*, 1–8. Cham: Springer.

Braun, Norman & Nicole J. Saam (eds.). 2015. *Handbuch Modellbildung und Simulation in den Sozialwissenschaften*. Wiesbaden: Springer. DOI: 10.1007/978-3-658-01164-2.

Buchmüller, Olga, Christoph Demian, Roland Meyer & Luka Szucsich. 2022. Syntactic complexity across registers in Russian. Submitted.

Chomsky, Noam. 1962. Explanatory models in linguistics. In Ernest Nagel, Patrick Suppes & Alfred Tarski (eds.), *Logic, methodology, and philosophy of science: Proceedings of the 1960 international congress*, 528–550. Stanford: Stanford University Press.

Cresswell, Max. 1973. *Logics and languages*. London: Methuen.

Deshors, Sandra C. 2015. A multifactorial approach to linguistic structure in L2 spoken and written registers. *Corpus Linguistics and Linguistic Theory* 11(1). 19–50.

Eckert, Penelope. 2000. *Linguistic variation as social practice*. Oxford: Blackwell.

Eckert, Penelope & William Labov. 2017. Phonetics, phonology and social meaning. *Journal of Sociolinguistics* 21(4). 467–496. DOI: 10.1111/josl.12244.

Edwards, John. 2007. Societal multilingualism: Reality, recognition and response. In Peter Auer & Li Wei (eds.), *Handbook of Multilingualism and Multilingual Communication*, 447–467. De Gruyter Mouton.

Egbert, Jesse & Marianna Gracheva. 2023. Linguistic variation within registers: Granularity in textual units and situational parameters. *Corpus Linguistics and Linguistic Theory* 19(1). 115–143. DOI: 10.1515/cllt-2022-0034.

Egg, Markus & Valia Kordoni. 2022. Metaphor annotation for German. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of LREC 2022*, 2556–2562. Marseille: European Language Resources Association.

Frazier, Lyn & Janet Dean Fodor. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6(4). 291–325. DOI: http://dx.doi.org/10.1016/0010-0277(78)90002-1.

Genette, Gérard. 2010. *Die Erzählung*. 3rd edn. Stuttgart: Fink. DOI: 10.36198/9783838580838.

Geurts, Bart & Janneke Huitink. 2006. Modal concord. In *Concord Phenomena and the Syntax Semantics Interface*, 15–20. Malaga: ESSLLI.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76. DOI: 10.1016/s0010-0277(98)00034-1.

Gibson, Edward. 2001. The Dependency Locality Theory: Papers from the first mind articulation project symposium. In Alec P. Marantz, Yasushi Miyashita & Wayne O'Neil (eds.), *Image, language, brain.* Cambridge, MA: MIT Press. DOI: 10.7551/mitpress/3654.003.0008.

Giere, Ronald N. 2004. How models are used to represent reality. *Philosophy of Science* 71(5). 742–752. DOI: 10.1086/425063.

Gilquin, Gaëtanelle & Magali Paquot. 2007. Spoken features in learner academic writing: Identification, explanation and solution. In Matthew Davies, Paul Rayson, Susan Hunston & Pernilla Danielsson (eds.), *Corpus linguistics conference CL2007*, 27–30. Birmingham: University of Birmingham.

Grabes, Herbert. 2014. *Sequentiality*. Peter Hühn, Jan Christoph Meister, John Pier & Wolf Schmid (eds.). The Living Handbook of Narratology. https://www.lhn.uni-hamburg.de/node/91.html (18 April, 2023).

Gries, Stefan Th. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34(4). 365–399. DOI: 10.1007/s10936-005-6139-3.

Halliday, Michael Alexander Kirkwood. 1978. *Language as social semiotic: The social interpretation of language and meaning*. London: University Park Press.

Halliday, Michael Alexander Kirkwood & Christian Matthiessen. 2004. *An introduction to Functional Grammar*. London, New York: Routledge.

Hasan, Ruqaiya. 2014. Towards a paradigmatic description of context: systems, metafunctions, and semantics. *Functional Linguistics* 1, 9. 1–54. DOI: 10.1186/s40554-014-0009-y.

Heine, Bernd & Heiko Narrog (eds.). 2012. *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press.

Heylighen, Francis & Jean-Marc Dewaele. 1999. *Formality of language: Definition, measurement and behavioral determinants*. Internal Report. Center "Leo Apostel", Vrije Universiteit Brüssel.

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10(2–3). 210–234. DOI: 10.1080/00437956.1954.11659524.

Hühn, Peter. 2013. *Event and Eventfulness*. Peter Hühn, Jan Christoph Meister, John Pier & Wolf Schmid (eds.). The Living Handbook of Narratology. https://www.lhn.uni-hamburg.de/node/39.html (18 April, 2023).

Jackendoff, Ray. 1997. *The architecture of the language faculty* (Linguistic Inquiry Monographs 28). Cambridge, MA & London, England: MIT Press.

Jackendoff, Ray. 2002. *Foundations of language*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198270126.001.0001.

Jäger, Gerhard. 2019. Model evaluation in computational historical linguistics. *Theoretical Linguistics* 45(3-4). 299–307.

Jenset, Gard B & Barbara McGillivray. 2017. *Quantitative Historical Linguistics: A Corpus Framework*, vol. 26. Oxford: Oxford University Press.

Kaplan, Dafna & Ruth A. Berman. 2015. Developing linguistic flexibility across the school years. *First Language* 35(1). 27–53. DOI: 10.1177/0142723714566335.

Karkaletsou, Foteini & Artemis Alexiadou. Forthcoming. Synthetic-analytic variation in the formation of Greek comparatives and relative superlatives. *Journal of Greek Linguistics*.

Karlsson, Fred. 2007. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43(2). 365–392. DOI: 10.1017/S0022226707004616.

Kauhanen, Henri & George Walkden. 2018. Deriving the constant rate effect. *Natural Language & Linguistic Theory* 36. 483–521. DOI: 10.1007/s11049-017-9380-1.

Keller, Rudi. 1994. *Sprachwandel: Von der unsichtbaren Hand in der Sprache*. 2nd edn. Tübingen: Francke.

Knoeferle, Pia & Matthew W. Crocker. 2006. The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science* 30(3). 481–529. DOI: 10.1207/s15516709cog0000_65.

Knoeferle, Pia & Matthew W. Crocker. 2007. The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language* 57(4). 519–543. DOI: 10.1016/j.jml.2007.01.003.

Koch, Peter. 1997. Diskurstraditionen: Zu ihrem sprachtheoretischen Status und ihrer Dynamik. In Barbara Frank, Thomas Haye & Doris Tophinke (eds.), *Gattungen mittelalterlicher Schriftlichkeit* (Script Oralia 99), 43–79. Tübingen: Gunter Narr.

Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36(1). 15–43. DOI: 10.1515/9783110244922.15.

Kornai, András. 2014. Resolving the infinitude controversy. *Journal of Logic, Language and Information* 23(4). 481–492. DOI: 10.1007/s10849-014-9203-2.

Kortmann, Bernd & Benedikt Szmrecsanyi (eds.). 2012. *Linguistic Complexity*. Berlin, Boston: De Gruyter. DOI: 10.1515/9783110229226.

Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3). 199–244. DOI: 10.1017/s0954394500000168.

Lahn, Silke & Jan Christoph Meister. 2016. *Einführung in die Erzähltextanalyse*. 3rd edn. Heidelberg: Metzler. DOI: 10.1007/978-3-476-05415-9.

Lass, Roger. 1997. *Historical linguistics and language change* (Cambridge studies in linguistics 81). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620928.

Lehmann, Nico, Motahareh Sameri, Vahid Mortezapour, Elisabeth Verhoeven & Aria Adli. 2024. The formal and functional distribution of right-peripheral arguments in German and Persian across registers. Submitted.

Lehmann, Nico, Dina Serova, Julia Lukassek, Sophia Döring, Frank Goymann, Anke Lüdeling & Roodabeh Akbari. 2023. Guidelines for the annotation of parameters of narration. *REALIS: Register Aspects of Language in Situation* 2(5). 1–41. DOI: 10.18452/26437.

López, Luis. 2020. *Bilingual grammar: Toward an integrated model*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108756181.

Lu, Xiaofei. 2011. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly* 45(1). 1–27.

Lüdeling, Anke, Artemis Alexiadou, Aria Adli, Karin Donhauser, Malte Dreyer, Markus Egg, Anna Helene Feulner, Natalia Gagarina, Wolfgang Hock, Stefanie Jannedy, Frank Kammerzell, Pia Knoeferle, Thomas Krause, Manfred Krifka, Silvia Kutscher, Beate Lütke, Thomas McFadden, Roland Meyer, Christine Mooshammer, Stefan Müller, Katja Maquate, Muriel Norde, Uli Sauerland,

Stephanie Solt, Luka Szucsich, Elisabeth Verhoeven, Richard Waltereit, Anne Wolfsgruber & Lars Erik Zeige. 2022. Register: Language users' knowledge of situational-functional variation: Frame text of the first phase proposal for the CRC 1412. *REALIS: Register Aspects of Language in Situation*. 1–58. DOI: 10.18452/24901.

MacDonald, Maryellen C., Neal J. Pearlmutter & Mark S. Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101(4). 676–703. DOI: 10.1037/0033-295x.101.4.676.

Martínez, Matías & Michael Scheffel. 2009. *Einführung in die Erzähltheorie*. 8th edn. München: C. H. Beck.

Martínez, Matías. 2011. Erzählen. In Matías Martínez (ed.), *Handbuch Erzählliteratur: Theorie, Analyse, Geschichte*, chap. AI1, 1–12. Stuttgart & Weimar: Metzler. DOI: 10.1007/978-3-476-05317-6_1.

McClelland, James L. 2009. The place of modeling in cognitive science. *Topics in Cognitive Science* 1(1). 11–38.

McCready, Elin. 2019. *The semantics and pragmatics of honorification: Register and social meaning*, vol. 11 (Oxford Studies in Semantics and Pragmatics). Oxford: Oxford University Press. DOI: 10.1093/oso/9780198821366.001.0001.

McWhorter, John H. 2011. *Why do languages undress?* Berlin, New York: De Gruyter Mouton. DOI: 10.1515/9781934078402.

McWhorter, John H. 2012. Complexity hotspot: The copula in Saramaccan and its implications. In Bernd Kortmann & Benedikt Szmrecsanyi (eds.), *Linguistic complexity: Second language acquisition, indigenization, contact*, vol. 13 (linguae & litterae), 243–266. Berlin, Boston: de Gruyter. DOI: 10.1515/9783110229226.243.

Meyer, Roland, Christoph Demian, Olga Buchmüller & Luka Szucsich. Submitted. Registers in Russian. A multidimensional analysis. 2022.

Meyerhoff, Miriam, Carol Aru, Manfred Krifka & Tonjes Veenstra. 2023. A multivariable analysis of variation in plural marking in Bislama. In Michol F. Hoffmann, Roland Beline Mendes & Rafael Orozco (eds.), *Coherence in linguistic systems*. London: Routledge. In preparation.

Miestamo, Matti. 2004. On the feasibility of complexity metrics. In *Finest linguistics: Proceedings of the annual Finnish and Estonian conference of linguistics, Tallinn*, 11–26. Tallinn: Tallinn University Press.

Montague, Richard. 1974. *Formal philosophy: Selected papers of Richard Montague*. New Haven: Yale University Press.

Mühlenbernd, Roland & Stephanie Solt. 2022. Modeling (im)precision in context. *Linguistics Vanguard* 8(1). 113–127. DOI: 10.1515/lingvan-2022-0035.

Münster, Katja & Pia Knoeferle. 2018. Extending situated language comprehension (accounts) with speaker and comprehender characteristics: Towards socially situated interpretation. *Frontiers in Psychology* 8, 2267. DOI: 10.3389/fpsyg.2017.02267.

Neumann, Stella. 2014a. *Contrastive register variation: A quantitative approach to the comparison of English and German*, vol. 251 (Trends in Linguistics. Studies and Monographs [TiLSM]). Berlin & Boston: de Gruyter. DOI: 10.1515/9783110238594.

Neumann, Stella. 2014b. Cross-linguistic register studies: Theoretical and methodological considerations. *Languages in Contrast* 14(1). 35–57. DOI: 10.1075/lic.14.1.03neu.

Pescuma, Valentina N., Dina Serova, Julia Lukassek, Antje Sauermann, Roland Schäfer, Aria Adli, Felix Bildhauer, Markus Egg, Kristina Hülk, Aine Ito, Stefanie Jannedy, Valia Kordoni, Milena Kühnast, Silvia Kutscher, Robert Lange, Nico Lehmann, Mingya Liu, Beate Lütke, Katja Maquate, Christine Mooshammer, Vahid Mortezapour, Stefan Müller, Muriel Norde, Elizabeth Pankratz, Angela G. Patarroyo, Ana-Maria Pleșca, Camilo R. Ronderos, Stephanie Rotter, Uli Sauerland, Gohar Schnelle, Britta Schulte, Gediminas Schüppenhauer, Bianca Maria Sell, Stephanie Solt, Megumi Terada, Dimitra Tsiapou, Elisabeth Verhoeven, Melanie Weirich, Heike Wiese, Kathy Zaruba, Lars Erik Zeige, Anke Lüdeling & Pia Knoeferle. 2023. Situating language register across the ages, languages, modalities, and cultural aspects: Evidence from complementary methods. *Frontiers in Psychology* 13. 1–31. DOI: 10.3389/fpsyg.2022.964658.

Pickering, Martin J. & Simon Garrod. 2004. The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences* 27(2). 212–225. DOI: 10.1017/s0140525x04450055.

Qin, Wenjuan & Paola Uccelli. 2020. Beyond linguistic complexity: Assessing register flexibility in EFL writing across contexts. *Assessing Writing* 45. DOI: 10.1016/j.asw.2020.100465.

Ravid, Dorit & Ruth Berman. 2009. Developing linguistic register across text types: The case of modern Hebrew. *Pragmatics & Cognition* 17(1). 108–145. DOI: 10.1075/pc.17.1.04rav.

Ravid, Dorit & Liliana Tolchinsky. 2002. Developing linguistic literacy: A comprehensive model. *Journal of Child Language* 29(2). 417–447. DOI: 10.1017/S0305000902005111.

Ritter, Joachim & Karlfried Gründer (eds.). 1984. *Historisches Wörterbuch der Philosophie*. Vol. 6: *Mo–O*, 13 vols. Basel: Schwabe.

Roche, Jörg. 1998. Variation in xenolects (foreigner talk). *Sociolinguistica* 12(1). DOI: 10.1515/9783110245172.117.

Rotter, Stephanie & Mingya Liu. 2023. Interlocutor relation predicts the formality of the conversation: An experiment in American and British English. *REALIS: Register Aspects of Language in Situation*. 1–27. DOI: 10.18452/26192.

Rotter, Stephanie & Mingya Liu. Submitted. A register approach to negative concord vs. negative polarity items in English.

Sakel, Jeanette & Eugenie Stapert. 2010. Pirahã – in need of recursive syntax? In Harry van der Hulst, Jeanette Sakel & Eugenie Stapert (eds.), *Recursion and human language*, 3–16. Berlin & New York: De Gruyter Mouton. DOI: 10.1515/9783110219258.1.

Sauerland, Uli. 2022. Quantifying the register of German quantificational expressions: A corpus-based study. In *Measurements, numerals and scales*, 261–273. Springer.

Sauerland, Uli & Artemis Alexiadou. 2020. Generative grammar: A meaning first approach. *Frontiers in Psychology* 11. DOI: 10.3389/fpsyg.2020.571295.

Sauermann, Antje, Britta Schulte & Heike Wiese. Forthcoming. Sprachkontakt in Namibia: Registerdifferenzierung im Namdeutschen. In Barbara Hans-Bianchi & Barbara Vogt (eds.), *Deutsch im Kontakt. Neue empirische Studien zu Kontaktphänomenen und -szenarien in der Gegenwart*. Germanistische Linguistik.

Schäfer, Roland, Felix Bildhauer, Elizabeth Pankratz & Stefan Müller. In preparation. *Modelling registers*.

Schank, Gerd & Gisela Schoenthal. 1976. *Gesprochene Sprache: Eine Einführung in Forschungsansätze und Analysemethoden*. Otmar Werner & Franz Hundsnurscher (eds.), vol. 18 (Germanistische Arbeitshefte). Tübingen: Max Niemeyer. DOI: 10.1515/9783111335971.

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198814771.001.0001.

Schnelle, Gohar, Mathilde Hennig, Carolin Odebrecht & Anke Lüdeling. 2023. Historische Korpora in sprachhistorisch orientierter germanistischer Hochschullehre. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 145(2). Forthcoming.

Scontras, Gregory, Maria Polinsky & Zuzanna Fuchs. 2018. In support of representational economy: Agreement in heritage spanish. *Glossa: a journal of general linguistics* 3(1). 108–145. DOI: 10.5334/gjgl.164.

Shadrova, Anna, Pia Linscheid, Julia Lukassek, Anke Lüdeling & Sarah Schneider. 2021. A challenge for contrastive L1/L2 corpus studies: Large inter- and intra-individual variation across morphological, but not global syntactic categories in task-based corpus data of a homogeneous L1 German group. *Frontiers in Psychology* 12. 1–29. DOI: 10.3389/fpsyg.2021.716485.

Shmueli, Galit. 2010. To explain or to predict? *Statistical Science* 25(3). 289–310.
Stachowiak, Herbert. 1973. *Allgemeine Modelltheorie*. Wien & New York: Springer.
Steger, Hugo, Karl Helge Deutrich, Gerd Schank & Eva Schütz. 1974. Redekonstellation, Redekonstellationstyp, Textexemplar, Textsorte im Rahmen eines Sprachverhaltensmodell: Begründung einer Forschungshypothese. In Hugo Moser (ed.), *Gesprochene Sprache: Jahrbuch 1972* (Sprache der Gegenwart 26), 39–97. Düsseldorf: Schwann.
Szmrecsanyi, Benedikt. 2006. *Morphosyntactic persistence in spoken English: A corpus study*. Berlin & New York: De Gruyter Mouton. DOI: 10.1515/9783110197808.
Szmrecsanyi, Benedikt. 2019. Register in variationist linguistics. *Register Studies* 1. 76–99. DOI: 10.1075/rs.18006.szm.
Szmrecsanyi, Benedikt & Alexandra Engel. 2023. A variationist perspective on the comparative complexity of four registers at the intersection of mode and formality. 19(1). 79–113. DOI: 10.1515/cllt-2022-0031.
Trudgill, Peter & Jenny Cheshire (eds.). 1998. *The sociolinguistics reader*. Vol. 1: *Multilingualism and variation*. London: Hodder Education.
Umbach, Carla & Stephanie Solt. 2021. Comparison via *eher*. *Journal of Semantics* 39(1). 39–85. DOI: 10.1093/jos/ffab021.
Veenstra, Tonjes. 2021. *Speakers' choices in creole contexts: The Mauritian case*. Lecture at Forschungskolloquium Romanistik. Universität Hamburg.
Verhoeven, Elisabeth & Nico Lehmann. 2018. Self-embedding and complexity in oral registers. *Glossa: a journal of general linguistics* 3(1), 93. DOI: https://doi.org/10.5334/gjgl.592.
Weiß, Zarah Leonie. 2017. *Using measures of linguistic complexity to assess German L2 proficiency in learner corpora under consideration of task-effects*. Eberhard Karls Universität Tübingen. (MA thesis). DOI: 10.13140/RG.2.2.12713.34401.
Weiss, Zarah & Detmar Meurers. 2019. Analyzing linguistic complexity and accuracy in academic language development of German across elementary and secondary school. In *Proceedings of the fourteenth workshop on innovative use of NLP for building educational applications*, 380–393. Association for Computational Linguistics. DOI: 10.18653/v1/W19-4440.
Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.). Found. by Wolfgang Viereck. Cont. by Manfred Krug & Heinrich Ramisch, *Proceedings of methods in dialectology xvi: Papers from the sixteenth international conference on methods in dialectology, 2017* (Bamberger Beiträge zur Englischen Sprachwissenschaft / Bamberg Studies in English Linguistics 59). Berlin et al.: Peter Lang.

Wiese, Heike. 2021. Communicative situations as a basis for linguistic systems: Integrating linguistic multi-competence with grammatical structure. *Working Papers in Urban Language and Literacies* 287.

Wiese, Heike, Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gargarina, Kateryna Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehaye, Sabine Zerbian & Yulia Zuban. 2022. Heritage speakers as part of the native language continuum. *Frontiers in Psychology* 12. DOI: 10.3389/fpsyg.2021.717973.

Wiese, Heike & Yannic Bracke. 2021. Registerdifferenzierung im Namdeutschen: Informeller und formeller Sprachgebrauch in einer vitalen Sprechergemeinschaft. In Csaba Földes (ed.), *Kontaktvarietäten des Deutschen im Ausland* (Beiträge zur Interkulturellen Germanistik Sammelbände), 273–293. Tübingen: Narr Francke Attempto.

Wolters, Gereon. 2004. Modell. In Jürgen Mittelstraß (ed.), *Enzyklopädie Philosophie und Wissenschaftstheorie*. Vol. 2: *H–O*, 4 vols. Sp. 911–913. Stuttgart & Weimar: Metzler.

Wu, L. L. & Lawrence W. Barsalou. 2009. Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychologica* 132(2). 173–189.

Wurzel, Wolfgang Ulrich. 1994. Grammatisch initiierter Wandel. In Benedikt Jeßing (ed.), *Sprachdynamik: Auf dem Weg zu einer Typologie sprachlichen Wandels*, 8 vols. (Bochum-Essener Beiträge zur Sprachwandelforschung 23.1). Bochum: Brockmeyer.

Yang, Charles. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Zeige, Lars Erik. 2011. *Sprachwandel und soziale Systeme*, vol. 27 (Germanistische Linguistik Monographien). Hildesheim: Olms.

Zeige, Lars Erik. 2014. On cognition and communication in usage-based models of language change. In Evie Coussé & Ferdinand von Mengden (eds.), *Usage-based approaches to language change* (Studies in Functional and Structural Linguistics 69), 49–80. Amsterdam: John Benjamins. DOI: 10.1075/sfsl.69.03zei.

Zeijlstra, Hedde. 2007. Modal concord. In *Proceedings of semantics and linguistic theory*, 317–332. DOI: 10.3765/salt.v17i0.2961.

Zeman, Sonja. 2016. Nähe, Distanz und (historische) Pragmatik Oder: Wie „nah" ist ‚Nähesprache'? In Helmuth Feilke & Mathilde Hennig (eds.), *Zur Karriere von ›Nähe und Distanz‹*. Berlin: de Gruyter. DOI: 10.1515/9783110464061-010.

Zeman, Sonja. 2018. What is a *Narration* - And why does it matter? In Annika Hübl & Markus Steinbach (eds.), *Linguistic foundations of narration in spoken and sign languages*, 173–206. Amsterdam: John Benjamins. DOI: 10.1075/la.247.08zem.

Appendix

Table 1: List of projects in the CRC[4]

Project	Status	Title	Research area	Project leader(s)
		Area A: Register and grammar		
A01		*GeRMaN: German register marking by non-literal expressions*	Semantics, pragmatics	Markus Egg (IAA)
A02	E	*Speaker's choices in creole contexts: Bislama and Morisien*	Creole linguistics, syntax, morphology	Manfred Krifka (ZAS)
A03		*Slavic in a multilingual setting: Register and fused (hybrid) lects*	Syntax, corpus linguistics, multilingualism	Roland Meyer, Luka Szucsich (ISH)
A04		*Building register into the architecture of language – an HPSG account*	Constraint-based syntax and semantics, morphosyntactic variation	Antonio Machicao y Priemer, Stefan Müller (IdSL)
A05		*Modeling meaning-driven register variation: Politeness and face management*	Formal & experimental semantics, game-theoretic pragmatics	Uli Sauerland, Stephanie Solt (ZAS)
A06		*Modeling register variation across languages*	Morphosyntactic variation, language contact, sociolinguistics, typology	Aria Adli (RS), Jozina Vander Klok, Elisabeth Verhoeven (IdSL)
A07		*Register effects in discourse expectations: Negation and modality in English*	Semantics, pragmatics, psycholinguistics	Mingya Liu (IAA)
A08	N	*Speakers' choices in three Saamaka communities*	Morphosyntax, lexicon, heritage grammars	Artemis Alexiadou (ZAS)
A09	N	*On the interplay between register and socio-geographic variation in Canarian Spanish*	Morphosyntax, variation, dialectology, sociolinguistics, probabilistic grammar, linguistic norm	Miriam Bouzouita, Laura Merino Hernández (IfR)
A10	N	*Doubling and register variation*	Semantics, pragmatics, morphosyntax, morphology	Aron Hirsch, Viola Schmitt (IdSL)

		Area B: Register and change		
B01	E	*Register and the development of periphrasis in the history of English*	Historical morphosyntax	Artemis Alexiadou, Thomas McFadden (ZAS)
B02	E	*Emergence and change of registers: The case of Lithuanian and Latvian*	Baltic linguistics, syntax, lexicon	Anna Helene Feulner, Wolfgang Hock (IdSL)
B03		*Register variation and asymmetric communication in Ancient Egypt*	Ancient Egyptian language, diachronic grammar	Silvia Kutscher, Dina Serova (IA)
B04		*Emergence and transfer of register patterns: Situational-functional parameters of intraindividual variation in the writings of Martin Luther and Johannes Bretke*	Historical morphosyntax, historical lexical semantics	Anna Helene Feulner, Jürg Fleischer, Lars Erik Zeige (IdSL)
B06	N	*Register in the cross-linguistic diachrony of logical particles*	Historical functional semantics, logical vocabulary, constant rate	Uli Sauerland (ZAS), Richard Waltereit (IfR)
		Area C: Register and cognition		
C02		*Variation in Situated Interaction II: Modeling real life situations*	Sociophonetics, sociolinguistics, multimodality, language attitudes	Stefanie Jannedy (ZAS), Melanie Weirich (GSW)
C03		*Real-time register comprehension in adolescent heritage speakers' languages*	Psycholinguistics, real-time language comprehension, multilingualism, heritage language	Natalia Gagarina, Pia Knoeferle, Katja Maquate (IdSL)
C04	E	*Register knowledge in advanced learner language*	Second language acquisition, corpus linguistics	Anke Lüdeling (IdSL)
C05		*Young adults' specialized register knowledge: Modeling late linguistic development in L1 and L2*	L1 and L2 register acquisition, academic language	Anke Lüdeling, Beate Lütke, Nicole Schumacher (IdSL)
C06		*Seemingly free (morpho)phonetic variation*	Phonetics, phonology, morphology, syntax, corpus linguistics	Malte Belz, Anke Lüdeling, Christine Mooshammer (IdSL)

C07	*The impact of language ideologies on register distinctions in multilingual contexts*	Sociolinguistics, language attitudes, language ideologies, language contact	Oliver Bunk, Antje Sauermann, Heike Wiese (IdSL)
	Central projects		
INF	*Data management, modeling and exploration*	Corpus linguistics, statistics, modeling, service	Malte Dreyer (CMS), Thomas Krause, Anke Lüdeling (IdSL)
MGK	*Integrated Research Training Group*		Uli Sauerland (ZAS), Richard Waltereit (IfR)
Z	*Central Tasks of the CRC*		Anke Lüdeling (IdSL), Luka Szucsich (ISH), Lars Erik Zeige (IdSL)

[4]Abbreviations: IdSL = Institut für deutsche Sprache und Linguistik, IfR = Institut für Romanistik, CMS = Computer- und Medienservice, IAA = Institut für Anglistik und Amerikanistik, IA = Institut für Archäologie, ISH = Institut für Slawistik und Hungarologie, ZAS = Leibniz-Zentrum Allgemeine Sprachwissenschaft, RS = Romanisches Seminar der Universität zu Köln, GSW = Germanistische Sprachwissenschaft der Uni Jena, E = project ended, N = new project