

Register: Language Users' Knowledge of Situational-Functional Variation

Frame text of the First Phase Proposal for the CRC 1412

Anke Lüdeling^a, Artemis Alexiadou^b, Aria Adli^c, Karin Donhauser^a, Malte Dreyer^a, Markus Egg^a, Anna Helene Feulner^a, Natalia Gagarina^d, Wolfgang Hock^a, Stefanie Jannedy^d, Frank Kammerzell^a, Pia Knoeferle^a, Thomas Krause^a, Manfred Krifka^b, Silvia Kutscher^a, Beate Lütke^a, Thomas McFadden^d, Roland Meyer^a, Christine Mooshammer^a, Stefan Müller^a, Katja Maquate^a, Muriel Norde^a, Uli Sauerland^d, Stephanie Solt^d, Luka Szucsich^a, Elisabeth Verhoeven^a, Richard Waltereit^a, Anne Wolfsgruber^a & Lars Erik Zeige^a

^aHumboldt-Universität zu Berlin ^bHumboldt-Universität zu Berlin / Leibniz-ZAS ^cUniversität zu Köln ^dLeibniz-ZAS

The Collaborative Research Center 1412 “Register: Language Users’ Knowledge of Situational-Functional Variation” (CRC 1412) investigates the role of register in language, focusing in particular on what constitutes a language user’s register knowledge and which situational-functional factors determine a user’s choices. The following paper is an extract from the frame text of the proposal for the CRC 1412, which was submitted to the Deutsche Forschungsgemeinschaft in 2019, followed by a successful onsite evaluation that took place in 2019. The CRC 1412 then started its work on January 1, 2020. The theoretical part of the frame text gives an extensive overview of the theoretical and empirical perspectives on register knowledge from the viewpoint of 2019. Due to the high collaborative effort of all PIs involved, the frame text is unique in its scope on register research, encompassing register-relevant aspects from variationist approaches, psycholinguistics, grammatical theory, acquisition theory, historical linguistics, phonology, phonetics, typology, corpus linguistics, and computational linguistics, as well as qualitative and quantitative modeling. Although our positions and hypotheses since its submission have developed further, the frame text is still a vital resource as a compilation of state-of-the-art register research and a documentation of the start of the CRC 1412. The theoretical part without administrative components therefore presents an ideal starter publication to kick off the CRC’s publication series REALIS. For an overview of the projects and more information on the CRC, see <https://sfb1412.hu-berlin.de/>.

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

1 Introduction

The CRC *Register* aims to investigate aspects of **intra-individual variation** in linguistic behavior that are influenced by situational and functional settings — what we will here define as **register**. Specifically, we target the linguistic knowledge that a speaker possesses **about situational and functional variation**, which we will here refer to as **register knowledge**. The central question we aim to investigate in CRC *Register* is:

Q: What constitutes a language user’s register knowledge?

We will illustrate the relevance of register with three examples that we briefly introduce here and then return to at relevant points later. First, the sentences *Meine Mama ist sauer* ‘My mom is ticked off’ and *Meine Mutter ist verärgert* ‘My mother is angry’, even though they can describe the same state of affairs, are appropriate in different situations. These situations can be distinguished perhaps by level of formality. Second, the expressions *7:53 pm* and *(around) 8 o’clock*, could be used to report the time of the same event, but have different contexts of use. While here level of formality may play a role, other factors are also relevant, such that e. g. *7:53 pm* would be preferred even in informal conversation between close friends if the topic is the departure time of a train. The third example involves pronunciation, specifically the realization of /ç/ in German as either [ç] (the so-called “ich-Laut”) or [ʃ] (the final consonant of *fish*). In words like *ich* ‘I’, speakers of e. g. the urban multiethnolect Kiezdeutsch variety generally have both the Kiezdeutsch [ʃ] and the standard [ç] realizations at their disposal and deploy them in ways that depend on their interlocutor(s) and the social face they want to project. Level of formality, purpose, status of the interlocutor and social projections are situational or functional features that influence the choice of one variant over another. A language user’s ability to negotiate this space as both an author and addressee constitutes their register knowledge.¹

Register is a topic of great popular awareness, often in the guise of notions such as *formal*, *colloquial*, and *standard language*, and it is a ubiquitous property of languages. Although register-related phenomena have been described and

¹We use the term “author” to refer to a person in the role of language production, independent of modality, i. e. a speaker, signer or writer. Analogously, an “addressee” refers modality-independently to a person in the role of language perception. We use “language user” as a cover term for a person using language in either production or perception roles.

studied in numerous descriptive, dialectal, historical, sociolinguistic, and corpus-based accounts of language, they have not yet played the important role that they require in linguistic or psycholinguistic modeling. The purpose of CRC *Register* is twofold:

- a) to enhance our understanding of the role of register knowledge in language use, acquisition, processing, variation, and change
- b) to develop a general theory of register knowledge to complement current models in grammatical theory and psycholinguistics

We have reason to think that register knowledge constitutes a **multifaceted, quantitative phenomenon pertaining to all linguistic levels, and relevant to all aspects of linguistic theory**. In what follows, we will motivate our conviction that the time is right for a comprehensive and targeted investigation of register phenomena and register knowledge, and that our group combines the many different types of expertise that are crucial for this investigation. We propose to carry out such a comprehensive and targeted investigation in CRC *Register* by pursuing three long-term research goals:

1. We will contribute substantially to the inventory of well-described register phenomena, covering both the variants and alternatives involved and the facts that influence the choice among them.
2. We will describe the co-occurrence of different variants and alternatives and different factors in specific linguistic domains (such as phonology or syntax), yielding multifactorial and multidimensional models.
3. Viewing the findings in 1 and 2 as reflexes of register knowledge, we will aim to integrate register knowledge into theoretical models of linguistic knowledge more broadly.

2 Detailed presentation of the research program

2.1 Introducing register and register knowledge

We define **register** as those aspects of socially recurring **intra-individual variation** that are influenced by **situational and functional settings** (similar to the definition in [Biber & Conrad 2009](#)), adopting a broad understanding of situation and function. The term *register* has been used in many different ways, and other terms, such as *diaphasic variation*, *context(ual) style*, *functional style* and

tenor have sometimes been used in ways that are similar to or overlap with how we use register here (Halliday et al. 1964; Labov 1966; 2001; Trudgill 1974; 2002; Coupland 1980; Milroy 1987; Giles et al. 1991b; Giles & Coupland 1991; Eckert & Rickford 2001; Halliday & Matthiessen 2004; Auer 2007; Croft 2010; Maas 2010; Biber & Conrad 2009, and many more). For the purposes of this proposal, we make a distinction between *register* as just defined and *style* as those aspects of an author's production that are individual.² Especially when dealing with written language, the concept of *genre* is also closely related to register but must be kept distinct. We will use it to refer to types of text (spoken as well as written) defined in language-external, culturally recognizable terms, independent of their linguistic properties. Genre is thus part of the situational-functional setting, and so a given genre may favor or require particular register choices (see Sampson 1997: for some relevant discussion).

The variation that register describes is ubiquitous in human language. The fact that language users vary their linguistic behavior according to the situation is among the linguistic observations most readily available to a layperson. Indeed, register phenomena can have great practical relevance to language users, as formulating an utterance in a way that is inappropriate for a given situation can have negative consequences. Some of the obvious differences are, of course, lexical – as in our *sauer/verärgert* 'ticked off/angry' example – or textual. Others concern syntactic or morphological complexity (cf. e. g. Biber & Conrad 2009) or even (perhaps less obvious) morphological productivity (cf. Plag et al. 1999). Any serious language description with non-trivial coverage – be it a grammar, a dictionary or a work for language learners – must recognize that some expressions and forms of a language belong to specific settings and not to others.

The pervasiveness of register phenomena and their emergence in many, often unexpected ways, in various areas of linguistics suggests that register should play a more fundamental role in linguistic theory building than has been acknowledged in many circles. There is good reason to think that register knowledge, rather than being secondary to central aspects of grammar, is integrated at an essential level with linguistic knowledge in other domains, and that this integration requires further investigation. Hence there is a convergence in what is needed to better understand grammatical and register knowledge: Grammatical theory

²We are, of course, aware of the fact that this is a theoretical distinction that is difficult to operationalize and that work on register often also touches on style, and furthermore that individual styles may generate intra-individual registers. We will discuss this in more detail in Section 2.3. Also, much like register, the term style has been defined in very different ways and applied to phenomena distinct from what we are concerned with here.

must further incorporate register knowledge into its models, and the study of register variation must increasingly take grammatical theory into account, so that we can develop a more complete understanding of a language user's linguistic knowledge that can properly model register-sensitive variation.

In the past, the relevant phenomena and above all their relation to a language user's register knowledge have typically been difficult to study, as they are often probabilistic, subtle in their manifestation and subject to reduced awareness in a language user, compared e. g. to extracting the meaning of a message. But with progress in experimental paradigms and statistical methods, as well as the availability of larger corpora from different functional settings, crowd-sourcing methods for experiments, and the computational power to carry out new types of data analysis, register variation and register knowledge now can and should be taken into account where they were previously set aside due to methodological limitations. We expect that *CRC Register* will have a substantial impact on all subject areas of linguistics, as our group encompasses expertise on sociolinguistics, psycholinguistics, grammatical theory, acquisition theory, historical linguistics, laboratory phonology, phonetics, corpus linguistics, and computational linguistics, as well as qualitative and quantitative modeling. Bringing together the models, methods, and expertise at our disposal, we will work on register phenomena in all of these areas.

The three examples we introduced above are all relatively simple, but they can be used to illustrate several interesting linguistic questions relating to register.³ As we will discuss in detail in Section 2.2 below, we use here the specific term *variable* in its restricted sense from variationist sociolinguistics for sets of linguistic expressions that are “ways of saying the same thing”, and the broader term *alternative* for linguistic expressions that stand in defined linguistic relationships to each other and typically have related meanings but need not be identical at the truth-conditional level.

- a) How does register variation interact with other contrasts between variants (i. e. the alternating forms that language users choose between)? How are existing contrasts in a language recruited to express register variation?

Since *sauer* and *verärgert* are interchangeable at some level, do they share (part of) their interpretation? Are the meanings of *sauer* and *verärgert* in

³The first example is adapted from the 2016 Mittlerer Schulabschluss exam in Berlin, an obligatory school exam for all 16-year olds (10th graders) in the city-state of Berlin. In one question, the students were asked to reformulate *Meine Mutter ist sauer* ‘My mother is ticked off’ in the standard language. One correct answer would have been *Meine Mutter ist verärgert* ‘My mother is angry’.

this context the same, or does one indicate a stronger degree of anger? How does this compare to the relationship between 7:53 pm and (around) 8 o'clock, which are also partly interchangeable but have a clearly definable difference in their interpretation relating to precision? How do other terms such as *genervt*, *empört* and *wütend* – with meanings shading from 'annoyed' to 'indignant' to 'enraged' – fit into the picture? How does this relate to the fact that attributive occurrences of the informal use of *sauer* (such as #*meine saure Mutter* 'my ticked-off mother') are distinctly odd?

- b) How can variants and alternatives be described and modeled? How should we integrate frequency into the models? Should register differences be modeled in terms of separate grammars, in terms of distinct variants or alternatives made available by a single grammar, or in terms of (non-modular) systems that incorporate register knowledge into the way that variability is produced?

How is the choice between *sauer* and *verärgert* implemented at the level of grammatical representation? E. g. does an utterance like *Meine Mutter ist sauer* involve a form of code-switching between distinct varieties of the language that differ in terms of properties like formality – since the formal form *Mutter* is used instead of the informal *Mama*? How can we understand in this context choices like that for the realization of /ç/ where, instead of a small number of discrete alternatives, we have a continuous range of possible articulations spanning the space from [ç] to [ʃ]?

- c) Which situational-functional factors influence the choice among variants or alternatives? Is there a finite inventory of registers that characterize individual situations in ways that are applicable to entire classes of variables or alternations? How do language users make use of their register knowledge to actively manipulate the situation and to interpret the behavior of their interlocutors?

What factors characterize the situations where *sauer* would be appropriate as a way to say 'angry' rather than *verärgert*? How are they distinct from the factors that are relevant for variation between [ç] and [ʃ]? It is well established e. g. that the latter alternation is associated with a specific set of social indexes in varieties like Kiezdeutsch. In such a context, how do authors go beyond simply reacting to the existing situation and actively deploy the different variants to signal group membership or convey social meaning?

- d) What factors affect the acquisition of register knowledge and linguistic flexibility? Which parts of register knowledge are acquired implicitly and which can be affected by explicit instruction? What is the role of linguistic experience?

When and how will the 16-year old students have acquired the knowledge about register needed in the exam (see again fn. 3)? How do learners recognize and then master the conditions for the choice between *7:53 pm* and (*around*) *8 o'clock*, given that there are situations where both are acceptable, and that the patterns of appropriateness here do not necessarily parallel those of many other register-sensitive alternations, e. g. ones that are sensitive to level of formality?

- e) How do register phenomena interact with language change?

The results of a sound change of /ç/ > /ʃ/ can be observed in several regional dialects of German, yielding a situation similar to that observed in Kiezdeutsch. When there is a diachronic asymmetry such that one variant of a variable is clearly innovative (in this case [ʃ] is innovative whereas [ç] is conservative), are there any predictable consequences for how the variants will be recruited for register purposes? Is it likely that *sauer* will lose its informal flavor in the meaning of 'angry' and be incorporated into more formal registers over time? Can we identify and model register-relevant factors that influence change?

- f) What can language comparison tell us about typological aspects of register phenomena?

We assume that mode of communication, level of formality, purpose, etc. play a role in influencing variation in all languages. But we also know that the phenomena that are influenced differ and that the parameters are not distributed in the same way in all languages. So while the choice between *7:53 pm* and (*around*) *8 o'clock* reflects a sensitivity to precision that we might expect to have clear parallels in other languages – German *19.53 Uhr* and (*gegen*) *8* show a similar distribution – we wouldn't necessarily predict the same cross-linguistic generality for something like the alternation between [ç] and [ʃ]. Which aspects of register variation are language-specific and which ones are found repeatedly across languages? E. g. does the *sauer/verärgert* alternation reflect any general principles of register-sensitivity to metaphorical usages?

g) How can register knowledge be integrated into linguistic theory?

What type of knowledge is the knowledge about the appropriateness of using *sauer* or [ç]? Is it linguistic knowledge? Is it perhaps more general social knowledge? How does it relate to knowledge about grammatical distinctions? How can we model such (quantitative and usage-based) knowledge?

Among the triggers for linguistic variation, the situation and function of an utterance have been recognized to influence all linguistic aspects and lead to intra-individual variation interacting with stable language user-specific properties that influence inter-individual – e. g. dialectal or sociolectal – variation. It is difficult to precisely define this situation-specific language use. However, it is clear that all language users are sensitive to this type of variation in their dominant/native language (and, to a more limited extent, in their other languages) and possess some knowledge of how to tailor their linguistic behavior to the situation and/or accommodate to their addressee(s). This leads to a complex state of affairs in which the appropriate abstractions must be drawn from the behavior of other language users and relevant situational properties so that language users are able to display the appropriate non-random variability in their own linguistic behavior and thus must have knowledge of how to maneuver the register space by selecting from the linguistic repertoire available to them. As we will discuss in detail below, the interplay between intra- and inter-individual variation is crucial here. Situationally appropriate language choices instantiate the linguistic flexibility of language users, a pivotal aspect of the human communicative capacity. The development of this ability begins with the earliest phases of language acquisition and continues during adolescence and adulthood (including second language acquisition). We thus view register as a complex phenomenon that requires investigation and explanation.

Register has long been a topic of research in descriptive grammar and in parts of sociolinguistics. The crucial innovation of CRC *Register* is to study register as an integral and active part of language users' knowledge or competence (broadly understood). This focus on the register knowledge of individuals leads us to approach register not only on the basis of corpus and other usage data but also by working with data from different experimental settings and from different language-user groups (e. g. children, young adults, non-native speakers etc.). By varying groups and languages we will address the question of which parts of register knowledge are universal on an abstract level and which parts develop during acquisition and change with use. Our focus further reflects the crucial

recognition that the adjustments in linguistic register are and should be central to linguistic theory building. A related innovation of our approach is the modeling of register choices along the lines of choices in other linguistic areas, e. g. as parallel to allomorphy, the management of implicatures and code-switching.

The central question we aim to investigate in *CRC Register* is therefore:

Q: What constitutes a language user's register knowledge?

2.2 Background on relevant strands of prior research

The observation that there are situation-dependent ways of using language can be found in virtually any body of linguistic work, independent of grammatical tradition. Witness already Dionysius Thrax's Greek grammar from 100 BCE, which outlines as its domain only "the usages of language as generally current among poets and prose writers" (Kemp 1986: 343) – and not the language used by farmers or even that used by poets when they are at the market. In the same vein, many "traditional" grammars and dictionaries often qualify their statements about grammatical forms as appropriate for a given situation such as *colloquial*, *vulgar*, *learned*, etc. (e. g. Paul 1920; Behaghel 1923). These works typically use a small, finite inventory of situational descriptions that are not clearly delineated, and they often provide little or only anecdotal evidence for their ascriptions. Such descriptions of situational language use also frequently conflate different phenomena and fail to take into account the probabilistic nature of much of register variation. There is often neither enough data nor an attempt at explanation or modeling of the differences. Some accounts of language use do probe deeper into the different types of situation-dependent variation. For example, we find in-depth research along the continua of conceptually spoken – conceptually written; literal/literate – oral/orate; informal – formal; immediacy – distance (see Koch & Oesterreicher 1985; 1994; Dürscheid 2003; Ágel & Hennig 2006a,b; Maas 2010, and the papers in Feilke & Hennig 2016, among many others). This descriptively very careful and theoretically insightful work will form an important foundation for our explorations.

In **historical linguistics**, register-related aspects of variation and change have been discussed in studies on both the individual and the textual level. One main area of interest has traditionally been the emergence of vernacular standard languages in contrast to dialectal variation (see, e. g. Wegera 2007 for Standard German, or Holm 1967 for Swedish). Along the prescriptive/descriptive divide, a central point of focus has been the role of early grammarians, translations of the Bible, school grammars, and institutional standardization efforts, judged

against the changes in actual usage. Another influential strand of research investigates the interactions of orality and literacy (CRC 321 “Mündlichkeit und Schriftlichkeit”, Freiburg), including reflexes of spoken language in historical texts (Macha et al. 2005; Ágel & Hennig 2006a,b). Furthermore, historical linguistics concerns itself with tracing the development of certain text types (e. g. Habermann 2001; Klein 2011; Taavitsainen et al. 2014). There has also been a continuous interest in languages for special purposes, e. g. that used by miners or students, often with the aim of documenting their functional lexical variation (Schirmer 1911). The advances in sociolinguistics spawned the subfield of historical sociolinguistics, demonstrating the existence of social strata in historical texts (Romaine 1982; Milroy 1987; Trudgill 2002; Tagliamonte 2016; Säily et al. 2017). Yet other studies have focused on the style of individual authors and/or texts (for example Siegroth-Nellesen 1979 for Meister Eckhardt, or Schenker 1977 for Luther or Adams 2015 for Birgitta of Sweden). However, historical linguistics does not analyze text types or seminal Bible translations as complex bundles of functionally distributed register phenomena, as proposed in CRC *Register*.

Early **variationist sociolinguistics** as initiated by William Labov emphasized the speech community but also recognized and carefully distinguished between different registers (typically referred to as “contextual styles”). In his early work, Labov was particularly focused on exploring the vernacular language, i. e. informal registers, because he believed that less “attention paid to speech” would yield more colloquial speech forms compared to more formal registers, which show more resemblance to written language. Labov’s work on methods for elicitation of different registers will crucially inform the data collection of many of the projects in CRC *Register*. A further crucial component to our research comes from Labov’s modeling of variation in terms of **variable** and **variant**. The fundamental idea is that many cases of variation can be understood in terms of a single abstract variable which can be expressed by different functionally equivalent concrete variants (see Labov 1966; 1969; 2001; 2004, etc., Bayley 2013; Tagliamonte 2016, among many others). For example, in most varieties of English we can recognize a variable (ING), which encompasses different pronunciations of the verbal suffix, orthographically *-ing*, including the variants [ɪŋ] and [ɪn]. While Labov’s main focus was on inter-individual variation, already in his earliest studies he showed that there is intra-individual variation with respect to such variables according to situation, in addition to grammatical and other factors. We will adopt here the notions of variable and variant as highly useful for the discussion of linguistic variation of all kinds, including that sensitive to register. Nevertheless, we note that it has proven rather difficult to define the term *variable* satisfactorily, as has been discussed since the beginnings of variationism

(Lavandera 1978; Labov 1978; Romaine 1981). Even phonetic variants may not be completely functionally equivalent, and it remains controversial how the notion can be extended to lexical and syntactic variants (cf. the discussions in Croft 2010; Adli 2013; Eckert & Labov 2017, among many others). But even against the background of these difficulties, we believe that variationism is one of the most important approaches for us to build on as it allows a careful consideration of the equivalence of variants (the envelope of variation, see e. g. Labov 1982; Cheshire 1999). It is also a good model for computing the effects of language-internal and language-external factors and their interaction, and how they govern the choice between variants. Moreover, Labov's work shows that linguistic variants can index group identification independent of social class, something which ultimately leads the way to research on the social meaning of variation, as we will discuss below.

To avoid some of the difficulties specifically with the concept of variable, we will adopt the term **alternative** as a more general concept for instances where variable would be less appropriate. We will understand alternatives as sets of linguistic expressions that stand in a defined linguistic relationship to each other, such as belonging to a single morphological paradigm or a pragmatic scale (in the sense of Horn 1972). Alternatives typically have related meanings but need not be strictly identical at a truth-conditional level, or even functionally equivalent, as in our example of *7:53 pm* vs. *(around) 8 o'clock*. This distinction should be understood not as embodying any strong claims but as an initial, working terminology to foster clarity and explicitness. In the framework of CRC *Register*, this topic will be a constant matter of discussion and investigation, and we will work on sharpening our understanding of the anatomy of variation, register-based and otherwise. It is thus to be expected that how we draw the relevant distinctions and refer to things will evolve and improve in the course of our work.

One theoretical tradition where register phenomena have received a great deal of attention is **Systemic Functional Linguistics** (Halliday et al. 1964; Halliday 1989; Trudgill 1983; Halliday & Matthiessen 2004, among others, and now Lukin et al. 2008 for an updated overview of Halliday's model of register). It proposes to model register theoretically as a particular **setting of systemic probabilities**. In this tradition, three components of register, field, tenor and mode, are distinguished (Halliday & Hasan 1985). *Field* refers to what is happening, to the nature of an interaction that involves language, including the respective subject matter and the purpose of the interaction. *Tenor* targets the participants, in particular, their statuses and social relationships in the interaction and in general. Here notions of power and solidarity play a central role in the analyses. Finally, *mode* is about the role of text or discourse in the interaction (e. g. how central is it to the

activity, and is it accompanied by other, non-linguistic activities), its organization and properties (such as its oral or written presentation or its monologic or dialogic nature).

Especially relevant for CRC *Register* is the body of work that establishes frequencies of different types of grammatical features across text corpora (Biber 1995; 1999; 2006; 2009, etc., Biber & Conrad 2009; Grieve et al. 2011; Neumann 2014; Kunz & Lapshinova-Koltunski 2015; Teich et al. 2016, and many others). This method is characterized by multidimensional and multivariate analyses (some of Biber’s studies consider more than 70 phenomena), and there are interesting ongoing methodological discussions about suitable statistical models (Grieve et al. 2011; Ruetter et al. 2014; Wieling & Nerbonne 2015; Szmrecsanyi 2017). Among the main insights of this strand of research is a better understanding of how linguistic features on all levels co-occur in different situations and how fine grained the distinctions are that language users are able to make systematically. The research in Biber’s tradition is absolutely foundational for the study of register and we will draw on it extensively. We do, however, see a number of places where CRC *Register* can extend and improve upon it. Rather than looking at functionally equivalent variants, the analyses in that tradition use corpus counts of surface forms, which sometimes makes it difficult to interpret the findings (see Lüdeling 2017 for a discussion, as well as Biber 2012; Biber et al. 2016 for some relevant background). Furthermore, Biber’s work is primarily agnostic about language-user identity, while our research interest specifically in register *knowledge* again implies a focus on intra-individual variation which we think will lead to new insights. See also Section 2.5 for discussion of how we will work on extending Biber’s statistical methodologies.

Cognitive linguistics treats situational knowledge, and thus social information, as an integral part of linguistic meaning (Langacker 1999). The theoretical framework of “situated cognition” considers situationally appropriate linguistic choices as an aspect of goal-oriented assessment of communicative situations. The representation of situational categories (scripts) evolves through experiences, subsuming multimodal perceptual information, affective appraisals, patterns of action responses and corresponding socio-cultural evaluations (Barsalou 2008; 2016; Santos et al. 2011; Wu & Barsalou 2009). The general mechanism of knowledge organization is also operative during language acquisition, consolidation and adjustment of linguistic structures. The inherently situated nature of language knowledge allows fast access to patterns of language use stored in the script as fitting with the communicative requirements, thus enhancing the probability of goal achievement. The conceptual structure underlying the meaning of content lexical items consists of three main knowledge types: taxonomic,

situational and affective knowledge. Situational knowledge elements comprise information about prototypical events: locations, time, causal relations, participants and their social roles, properties and actions (Santos et al. 2011; Wu & Barsalou 2009). Salient elements of situational knowledge such as the representation of prototypical participants and the appraisals of their social roles and relations focus attention on recurrent patterns of linguistic behavior in specific communicative situations and foster the association of linguistic expressions with external circumstances. Communicative goals encompass the addressee's cooperation and therefore foster addressee-directed optimization of language output. However, communicative goals are the goals of the author. Therefore, they involve author-centered aims such as the establishment or maintenance of social status (image) and social relations (in- and out-group-membership). Along with language proficiency in the native or a foreign language, speakers' ability to achieve communicative goals through functionally situated language choices might be associated with more general individual properties such as personal traits, empathy and motivation, age, gender and cultural background.

Finally, work within the **third wave of sociolinguistics** exploring individuals' language use (called speaking styles within that tradition) has recognized that language users choose a particular variant not as a result of a mechanistic process, but with a communicative intention (see Eckert 2008; 2012; Campbell-Kibler 2010 and others). Third-wave studies emphasize the individual stylistic perspective, in which language users have a more active role in placing themselves linguistically within a social landscape, and whereby the continuously shifting **social meaning** of variation is an essential feature of language. As will be discussed in detail below, this will be a crucial ingredient for our investigations, but an understanding of register variation must consider the overall situational landscape in addition to social aspects, and must also approach the relationship between inter- and intra-individual variation from a somewhat different perspective.

Further relevant approaches that look at intra-individual variation – not necessarily always triggered by register differences – which we will also draw upon are **Communication Accommodation Theory**, **Interactive Alignment Account**, and **Audience Design** (Giles et al. 1973; Bell 1984; Giles et al. 1991a,b; Soliz & Giles 2014; Pickering & Garrod 2004).

2.3 Core issues for the study of register knowledge

Among the many important questions for research on register, we can identify three core issues which are of central importance for our register knowledge perspective and thus will play a crucial role in the research of CRC *Register*.

The first issue is the **relationship between intra- and inter-individual variation** and the evidentiary roles of individual language users and speech communities for the study of register. As discussed above, under our definitions register is distinct from style in that the latter characterizes patterns specific to individuals, whereas the former characterizes behavior patterns that are recurrent – and perceived as recurrent – in a speech community. At the same time, register variation differs from regional or sociolectal variation in that it can only be identified on the basis of differences in the linguistic behavior of a single individual across situations. That is, register variation is intra-individual variation, but of a particular kind that is shared in its essentials across the language users of a community. This presents a number of theoretical questions and methodological challenges. One of the major outstanding challenges in research on language variation in general is that the factors and linguistic features that are relevant for intra-individual variation show considerable overlap with the ones that are involved in inter-individual variation. E. g. the variable realization of German /ç/ as [ç] or [ʃ] characterizes both social and regional inter-individual (i. e. dialectal) variation, and intra-individual register variation. This has been noted and addressed repeatedly, especially in variationist sociolinguistics (Labov 1966; Trudgill 1974; Bell 1984; Jannedy & Weirich 2014; Eckert & Labov 2017). Nonetheless, it presents consistent methodological challenges for analyzing specific cases of variation, as it requires careful work to disentangle the intra-individual from the inter-individual in order to identify the specific properties of each. It also raises interesting questions for how individual language users sift through the different types of evidence in their environment so that they can first acquire their knowledge of the different types of variation and then apply it in their own production and in interpreting the behavior of other language users. Given the extensive overlap, it is thus no small task for the learner, the language user, and also the analyst to abstract an understanding of the different contributions of inter- and intra-individual factors from the patterns of variation they observe.

The way that points of inter-individual variation can find their way into (register-sensitive) intra-individual variation has been discussed in terms of first-order vs. second-order indexes (Silverstein 2003; Eckert 2008; 2012; Wolfram & Schilling-Estes 2015) and enregisterment (Agha 2007; Snell 2017). According to Silverstein, a first-order index simply indicates membership in a population, what Labov (1971) calls a social indicator. Once a variant has been recognized as belonging to a specific context or social group, it can be used to index membership in this group. It is then a second-order index, or what Labov (1971) calls a social marker. For example, the already mentioned palatalization of /ç/ to [ʃ] (Jannedy & Weirich 2014) is a salient phonetic feature or indicator of Kiezdeutsch,

which can be employed by language users associating with this group and which then serves as a social marker. Speakers not directly associated with language users of Kiezdeutsch copy or cite (Rampton 1995; Auer 2003) this feature to invoke various kinds of social meanings ranging from positive (toughness and coolness) to negative (unsophisticated and uneducated). Once this indexicality has been conventionalized, it becomes available for further indexicalization over time and contexts. Eckert (2008: 453) calls this an “indexical field” — a range of potential meanings, which can be “activated in the situated use of the variable”. Register can thus, at least to a certain extent, be reconstructed from the indexical values of variants, e. g. as indexation of the value formal-informal or polite-impolite. One of our aims then is to understand ‘enregisterment’: “processes and practices whereby performable signs become recognized (and regrouped) as belonging to distinct, differentially valorized semiotic registers by a population” (Agha 2007: 81). To do so, we need to better understand the association between values and variants. For these purposes, work in sociolinguistics, sociophonetics, and psycholinguistics, as well as computational sociolinguistics (e. g. Nguyen et al. 2016) will clearly be very relevant to CRC Register, and the exploration of register knowledge will greatly benefit from embedding them in a broader approach.

These questions regarding the relationship between intra- and inter-individual variation are also extremely important for the investigation of the acquisition of register knowledge. The language learner must, after all, extract from the complex of different types of variation they are exposed to a similarly complex understanding of the constitutive elements of different registers and develop a complex understanding of their situated nature so that they can acquire a linguistic system with an appropriate component of variation. The acquisition of register knowledge comprises accumulation of diversified linguistic repertoires and the evolving ability of the speaker to shift smoothly and effectively between them in order to cope with changing situational and functional affordances. Thus we will devote significant attention to this topic, also known under the name **linguistic flexibility** (Berman & Nir-Sagiv 2007; Kaplan & Berman 2015). Linguistic flexibility is the domain-specific capacity of the language user for fast and flexible adjustments of language use to communicative goals, with respect to the affordances of different communicative situations (Berman 2015; 2018; Berman & Nir-Sagiv 2007; Kaplan & Berman 2015).⁴ Although linguistic flexibility is clearly related to the concept of register knowledge, there is a basic distinction between

⁴Linguistic flexibility is similar, but not identical, to cognitive flexibility, as they do not follow parallel developmental trajectories. Cognitive flexibility pertains to the capacity to understand tasks and their specific requirements, to adjust to alternations in task parameters, or to switch between tasks quickly and successfully (Deák 2003; Deák & Wiseheart 2015; Legare et al. 2018).

skill/ability and knowledge, both assessed by the outcome – the situated variation of linguistic expressions. During language development, there are phases in which linguistic behavior does not fully reflect register knowledge, as the ability for appropriately situated production might be hampered through ongoing processes of accumulation, consolidation and reorganization of linguistic means (Karmiloff-Smith 1997), often leading to linguistic entrenchment.

The second issue is how to **model the integration of knowledge of variation, register-sensitive and otherwise, with other types of linguistic knowledge, in particular grammatical knowledge**. In principle, we can distinguish two components of variation – the inventory of things that can vary with each other in a given variety (variables and their variants, as well as other types of alternations), and the factors that influence how a language user will choose among the variants or alternatives in a particular instance. The question of how models of grammar come into play can be answered differently for these two components. We can then distinguish three broad classes of models of grammatical variation according to those answers. The first split depends on whether a language user's knowledge of the details of how available variants are employed is integrated with other types of grammatical knowledge. If this is the case, this implies that the basic knowledge of the inventory of variants and alternations must also be integrated into the grammar, and we arrive at a holistic model where knowledge of variation and its use is inseparable from knowledge of grammar. Versions of this idea, which we will refer to as **Usage Based**, have been adopted by a number of approaches, where the idea is that knowledge of variation is extracted from usage entirely in tandem with knowledge of grammar, and in fact there is no principled, clear-cut distinction between the two (see for instance Goldberg 2006; Zeldes 2012; Pierrehumbert 2008). The alternative is to assume a form of modularity or autonomy of some core grammatical knowledge, which is distinct from the knowledge of how the available variants and alternatives are deployed. Among models of this type, a further split can be observed, depending on where they locate the inventory of variants or alternatives with respect to the modular divide. One possibility is to propose that individual grammars fix relevant properties one way or another and thus cannot produce variation for those properties. Rather, when variation arises, this is because an individual language user controls 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 on the relevant point, and compete with each other for use at any given time – even within a single utterance.⁵ These so-called **Grammar**

⁵A common confusion about this approach is that what is in competition are complete gram-

Competition models thus assimilate intra-individual variation to language/code-switching (Kroch 1989; 2001; Yang 2002; 2010). Thus, both the inventory of variants and alternatives, and the way in which they are used in variation, are external to the grammar proper. The other type of model maintains this modularity, but puts the inventory of available structures on the grammar side. Only the knowledge of how those structures are put to use is kept out of the grammar. In other words, a theory of grammar is adopted that can produce variation, in the sense that it can make multiple distinct outputs available for a given grammatical specification, in effect assimilating variation to allomorphy. An approach of this type, **Combinatorial Variability**, has been explicitly proposed and explored by Adger & Smith (2005); Adger (2006); Adger & Smith (2010); Adger (2014); Alexiadou (2017). The distinction between Grammar Competition and Combinatorial Variability at least can be seen as a debate over whether variation involves macro- or micro-choice in the sense of Coulmas (2007).

It remains very much an open question which of these approaches can best model intra-individual variation, and even whether the answer might be different for different instances or types of variation. One could imagine, for example, that the theoretical treatment that would be most appropriate for instances of variation involving semantically/functionally equivalent variants of a variable would differ from the one that is best for alternations between non-equivalent forms. The three approaches described above attempt to tie observed variation to grammatical systems at different levels, and thus they should be distinguishable in terms of their predictions about how certain types of variation might relate to other observable properties of a language. As with all questions at the level of theoretical frameworks, however, it is far from simple in practice to find clear empirical tests to decide among them. It will thus be an important focus of several projects in CRC *Register* to explore the consequences of different models of grammatical variation for the register phenomena being considered and to develop specific tests that can tease them apart.

The third core issue regards **our understanding of the meaning associated with register-sensitive variation**. Again, register phenomena constitute intra-individual linguistic variation in which language users choose among available variants according to the situation (as they perceive it to be). But as noted above, language users do not simply react to a situation, but rather take an active role in shaping it. That is, their choices are not determined solely by the situation,

mars, which would imply an explosion in the amount of grammatical information that a language user would have to carry around. What is actually intended (as is made clear in e.g. Kroch 2001) is rather that just those parts of the grammar that are in fact in variation are involved in the competition, with the rest of the grammar shared across the competitors.

but are also affected by communicative intent. Viewed from the perspective of the addressee, the register-sensitive choices that the author makes can thus carry a great deal of meaning about their intent and their estimation of the situation. Since language users will have some awareness of how addressees extract such meaning from the author's register choices, this will affect their own choices as authors, as they aim to convey particular meanings to the addressee. Register variation and the meaning attached to it thus subsumes a complex interaction between the communicative intent of the author, the author's and addressee's current assessments of the situation, and reasoning by both author and addressee about each other's knowledge states and motivations. Our approach to register from the perspective of register knowledge is thus uniquely well suited to achieving a deeper understanding of these interactions.

Relevant here as starting points are the notion of indexical order of variants and the emphasis on a language user's active role from the third wave of sociolinguistics (Silverstein 2003; Eckert 2008), as well as the concept of social meaning. From his earliest works, Labov speaks of the social implications and indexes of linguistic behavior. Systemic-functional approaches likewise emphasize that language is a "social phenomenon" (an idea going back to Saussure) and explore the semiotics of linguistic signs (Hasan 1973; Halliday 1978 etc.). More recently, Campbell-Kibler (2010); Eckert & Labov (2017), and other authors speak directly of social meaning carried by or indexically linked to a variant. Campbell-Kibler's work shows that addressees have social associations (such as intelligent/educated) with different variants. Furthermore, these associations are context dependent, in that addressees can have variable interpretations of particular variants depending on their mood and social perception of the author (Campbell-Kibler 2008). As an illustration, variable interpretations of identical acoustical stimuli dependent on co-presented information have been found for Kiezdeutsch in Berlin (Jannedy & Weirich 2014). Variants can signal group membership, and their social meanings can be ideologically constructed ones, which then gives us a link between linguistic variables and macro-sociological categories. Because of their context dependency and constant mutability, some of these indexes are only interpretable to those who have refined perceptions, e. g. due to experience with the relevant speaker group or social stereotypes (Wells 1982). Other dimensions of social meaning, including some related to register phenomena, may be more widely accessible; the range of variation in this area remains to be further explored.

However, the social meaning conveyed by the choice of a particular variant or alternative is not independent from other sorts of linguistic meaning. Relevant register phenomena frequently involve instances of lexico-pragmatic variation

that do not fit the original concept of a Labovian variable, but rather alternations between forms with different grammatical properties or truth-conditional meanings. In the typology of meaning and inference types developed in semantics/pragmatics (entailment, presupposition, implicature, etc.), certain aspects of social meaning can be aligned to projective aspects of conventional meanings (Smith et al. 2010; Davis & Potts 2010). Other aspects appear to extend this typology, an example being the notion of “dog-whistle” invoked by Henderson et al. (2017). Dog-whistle language involves phrases typically used in political propaganda, which are innocuous to the general public, but have specific meanings for certain political groups. E. g. most Americans understand “inner-city residents” as being race-neutral, but in racist circles the phrase is synonymous with “blacks”. In our theoretical model, this means that certain variants may be chosen because they have the same meaning to some addressees, but different meanings to others. Finally, these different sorts of meaning have been shown to interact, examples being Beltrama’s (2016) finding of an interaction between conventional and social meaning in the domain of adjectival intensification, and findings by Feeney & Bonnefon (2012) and Mazarella (2015) that politeness interacts with implicature calculation. This suggests that the study of semantic/pragmatic meaning and of social meaning can inform one another, and should proceed in tandem (see also Gutzmann 2015: for a model of meaning that incorporates both traditional semantic truth conditions and use conditions, which may have interesting implications for the study of both social meaning and register meaning).

Semantics and pragmatics have developed explicit models of the process whereby conventional meanings are negotiated by the interactions of authors and addressees. In particular, game-theoretic semantics (Lewis 1969; Nowak et al. 2000; Benz et al. 2006) and Bayesian pragmatics (Frank & Goodman 2012; Goodman & Frank 2016; Franke & Jäger 2016) are related research programs that seek to account for linguistic phenomena via mathematical models of authors, addressees, and their expectations of each other. Such approaches have been productively applied to areas including emotive and vague language (McCready 2012; Lassiter & Goodman 2017; Schöller & Franke 2017). Recently, Burnett (2017; 2019) demonstrates that ideas of the game-theoretic model can be fruitfully extended to social meaning, allowing in particular the modeling of register phenomena. As social meaning will be a topic of interest for several projects within CRC *Register*, two workshops on the topic will be held in 2021 and 2023, following on prior successful workshops in 2017 and 2019, and co-organized by A05, and C02.

Our perspective on register as being part of the internalized and learned linguistic knowledge of a language user leads us to a distinctive and comprehensive approach: Register knowledge is not necessarily secondary to other types

of linguistic knowledge. It interacts with them at a basic level, and we must be prepared to consider it as integral to a comprehensive theory of language, along with e. g. grammatical and sociolinguistic knowledge. While there is some prior and current work in a similar vein, we feel that nowhere is the goal pursued as broadly as in our planned CRC *Register*. In this enterprise we bring together – in a novel way – expertise in formal models of different domains of grammatical theory (especially syntax, semantics, and pragmatics) and usage-based approaches to language in historical linguistics and language change, sociolinguistics, phonetics, psycholinguistics, language acquisition, language education, corpus, and computational linguistics. The researchers participating in CRC *Register* bring experience on all levels of language exploration (phonetics, phonology, morphology, syntax, semantics, pragmatics, text structure, etc.) and cover a broad range of languages and varieties as well as a full range of modern research methodologies, including corpus studies, fieldwork, sociolinguistic interviews and an array of experimental techniques. This puts us in a unique position to approach register knowledge, taking into account all ingredients that we think are involved in register-dependent variation within and across languages. Overall, we aim to elevate the status of register from its current marginal role to that of a mainstream concern in the field, similar to the effect that the successfully completed CRC 632 had for information structure, which involved some of the same PIs. Our specific goals are:

1. to expand on the inventory of well-described register-sensitive phenomena and classify them according to their grammatical properties and the situational parameters that they are sensitive to,
2. to model the described register phenomena in various domains and their interactions with situational and other factors, and
3. to integrate register knowledge into theoretical models of general linguistic knowledge.

2.4 Research areas

Recall now our central question Q from above:

Q: What constitutes a language user's register knowledge?

We plan to pursue Q in three research areas, A: Register and grammar, B: Register and change, and C: Register and cognition. These areas provide a useful structure

for bundling the projects of CRC *Register* based on common research questions that further specify Q, though of course the three areas and their projects are intimately connected to each other and will mutually inform each other in crucial ways. Here we will describe the three research areas in more detail, in terms of both what brings together the projects within each area and how the three areas and their projects are tied to each other.

2.4.1 Area A: Register and grammar

Area A of CRC *Register* will examine the interplay and parallels between register knowledge and grammatical knowledge. Language makes use of choice and alternation at all levels, perhaps most obviously in the paradigmatic aspects of grammar. Thus it is clear that a significant portion of a language user's knowledge will consist in relating particular alternatives to each other and regulating the choices among them. The matter to be investigated in CRC *Register* is how the various grammatical alternatives made available are recruited for register purposes, and what parallels can be identified with the mechanisms involved in other manifestations of choice and alternation in language. The unifying question for the projects in this area is thus:

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

The projects in Area A will use corpus analysis (A01, A02, A03, A04, A06), fieldwork (A02, A06), and several experimental methodologies (A01, A02, A03, A04, A05, A06) to investigate syntactic (A02, A03, A04, A06), lexical (A01, A02), morphological (A02, A04), semantic (A01, A05), pragmatic (A01, A05), and phonological and prosodic (A02, A03) phenomena in a wide range of languages and their interactions with register. The results from Area A will provide theoretical tools for the studies of language change in Area B and the work on acquisition and processing in Area C, and will in turn build on input from those areas to increase their explanatory depth.

All register phenomena involve choices, like that between *sauer* and *verärgert* or the one among the various gradations between [ç] and [ʃ] as realizations of /ç/. These choices rely on the language user's knowledge of the set of alternative items to be chosen from (be they variants of a variable in the sociolinguistic sense or something else) and an array of language-internal or language-external factors that condition the choice. These factors may in some cases be conventions attached to the items (cf. Irvine & Gal 2000 on iconicity) or relate to the concept of

markedness in one of its guises, but in many cases they seem to be directly related to situational and functional differences. Several phenomena in other domains of linguistic theory are also centered around choice from a set of alternatives. For instance, allomorphy, focus, code switching, and scalar implicatures have all been discussed in such terms. One central thrust of *CRC Register* and of *Area A* in particular will be to explore such connections and probe them for underlying shared mechanisms.

We can break down the issues that will be pursued here in terms of three questions that further specify aspects of QA.

QAI: How are existing alternations recruited for register purposes? (A01, A02, A03, A04, A05, A06)

As noted, a large number of phenomena at all levels of grammar can be usefully thought of in terms of alternations, where a language user makes a choice from a set of alternatives. Even restricting our attention to those alternations where the choice is not deterministically regulated by the grammatical context, only a subset of these will actually be exploited in a given language to express register differences. An important question is thus how it is determined 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. Project A04, e. g. will be heavily concerned with determining, for contemporary German, which phenomena are subject to variation across registers and how they can be categorized. A03 will do so for Czech and Russian. In matters of lexical choice, we could imagine that the origins of particular lexical variants could play a role, an issue that will be investigated in projects A01 and A02. The metaphorical origins of the ‘angry’ meaning of *sauer*, for instance, could play a role in its relative informality, and borrowings often carry particular register associations due to the relative status of the source language. An important check on any explanations we might propose for why particular alternants show a particular type of register sensitivity can come from language comparisons, a central focus of project A02, A06. If the use of a certain kind of alternant in a certain kind of register has a principled explanation, we would expect to find it again and again across languages, perhaps even universally. On the other hand, register associations that are arbitrary should be language specific. One expectation is that pre-existing asymmetries between variants that are potentially in alternation with each other will have a clear tendency to be exploited for register purposes when two registers are sensitive to those asymmetries. A specific case of this that is the focus of project

A05 is when two alternatives have clear differences in their semantic content, e. g. two time expressions that describe the same point in time at different levels of precision (*7:53 pm* vs. *around 8 o'clock*). The project seeks to understand what are the properties that such cases of non-equivalent alternatives share, and what underlies these. Finally, project A03 will explore the recruitment in certain Slavic languages of word-order variants involving varying degrees of dislocation of material and how this might correlate with baseline acceptability and register effects.

QAii: How do language users choose among register variants? (A01, A02, A04, A05, A06)

Given that a particular alternation is associated with a particular register or some feature shared by a number of registers, it remains for us to determine how an author will choose a given variant in a particular situation and how an addressee will react to and interpret that choice. The questions that arise here include not only why one of the variants is associated with one register and not another, but also how register interacts with other factors that might contribute to the choices made. Furthermore, how do language users go beyond simply reacting to a given situation with their choices, and actively manipulate the situation or communicate special types of meaning via the choices they make? These issues will be a particular interest of project A05, which will investigate and develop explicit models of how semantic content and pragmatic reasoning interact and are manipulated by authors to express social meaning via the register-related choices that they make, focusing to start on the domain of numerical (im)precision discussed above. The consciousness of the register association of certain alternative structures will be studied by projects A04 and A06. A01 will consider how register factors play into language users' choices regarding use of metaphor and metonymy, and furthermore how these relate to grammatical environment and functional considerations. A02 will look at what factors play into how individual speakers of two Creole languages choose among available varieties given particular situations.

QAiii: What can register phenomena tell us about the grammatical implementation of variation? (A02, A03, A04, A06)

An important issue for any theoretical work on intra-individual variation – register-sensitive or otherwise – is how to integrate that variation into broader linguistic models. As noted above, we can contrast here approaches that incorporate

the language user's knowledge of variation with their other grammatical knowledge into an integrated model of their language (Goldberg 2006; Pierrehumbert 2008; Croft 2010; Hilpert 2013), with approaches that put a modular divide between knowledge of variation and knowledge of the core grammar. And within this latter divide we can distinguish those approaches which at least include the inventory of available variants within the grammar, as in the Combinatorial Variability of Adger (2006), from those approaches where distinct variants imply distinct grammars, and variation amounts to competition between grammars rather than variability within a single one (Kroch 1989; Yang 2002). Precisely because the projects in Area A are concerned with register-sensitive grammatical phenomena, they will be especially well positioned to shed light on this question. Project A02 envisages the co-existence of distinct types of intra-individual variation in the kinds of multilingual linguistic communities where the creoles they investigate are spoken. Here a familiar creole continuum exists alongside varieties of the lexifier language(s) as well as distinct local or colonial languages. This leads to a situation that is rich in variation within the creole language and in code-switching between the creole and other languages. A03 will specifically target the question of whether register-sensitive "expressive" word-order patterns in Czech and Russian provide evidence for variation within a single grammar or the co-existence of distinct grammars. An important part of project A04 will be exploring ways in which the various logical options for modeling grammatical variation can be implemented in the framework of Head-Driven Phrase Structure Grammar, and investigating to what extent register-related phenomena might provide evidence for the various approaches. A06 will explore this question from the angle of how to integrate cross-linguistic variation in register phenomena with language-specific register variation.

2.4.2 Area B: Register and change

The focus of Area B of the proposed CRC *Register* is the relationship between register and language change. Here again, the perspective of the language-user's knowledge provides a novel direction for considering register phenomena. Since languages differ not only in the specific linguistic phenomena that are sensitive to register, but also in the inventory of register distinctions they make, we must recognize an important place for change in any complete model of register. Registers themselves emerge and change diachronically, and there is also reason to think that register variation can play a causal role in changes to other aspects of a language. The central research question for the projects in Area B will thus be:

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

By virtue of the central role played by data from historical languages, which by definition have no living native speakers, the research here will be based primarily on corpora of historical texts. The focus will be on how the function and formality of (the sub-part of) the text (B01, B02, B03, B04, B05), topic (B02, B04, B05), target audience (B02, B03), attempts to render the speech of others in written form (B01, B03, B04), and register-related interactions with other written languages (B01, B02, B04, B05) interact with syntax (B01, B02, B03, B04, B05), morphology (B01, B02, B03, B04, B05), lexical choice (B02, B03, B04, B05), and semantics (B01, B03, B05). Work from Area B will yield additional comparative data and generalizations to inform the projects in Area A and will serve as a testing ground for ideas coming from Area C, given the central role that the acquisition and use of register knowledge must play in register-related aspects of change. Area B will also have an important role in the methodological investigations of CRC *Register*. All of the B projects work with corpora of historical languages and thus have to confront the issue of only having indirect access to register knowledge, in addition to the standard difficulties of dealing with texts written in dead languages. They will thus participate heavily in and make crucial contributions to the development of tools and procedures for reconstructing register information on the basis of limited written texts, i. e. for pursuing QMi, as will be described in more detail in Section 2.5.

The relationship between register and language change will be addressed under three perspectives that can be considered both separately and in conjunction:

QBi: How do registers emerge? (B01, B02, B03, B04, B05, also A02, A06)

Focusing in part on the earliest written attestations of individual languages, the projects addressing this aspect will identify individual authors with a variety of preserved text types or larger single texts showing internal functional diversity. These will be treated as expressions of emerging register knowledge, implying that language users adapt to changes in their historical circumstances, resulting in new patterns of intra-individual variation. When this variation becomes conventionalized, we may speak of emergent registers. The central aim is to identify the features that characterize the emerging registers, properties that make them susceptible for register-specific use and the circumstances that facilitate the necessity of a new register. In large parts of Medieval and Early Modern Europe, for example, emergent registers have been strongly associated with Latin/vernacular bilingualism. Initially, this resulted in a situation where Latin represented the formal and the vernaculars the informal varieties. At a later stage, Latin was replaced by the vernacular as the main written language, but it continued to be of

profound influence. For instance, Old Swedish copied the Latin system of relative pronouns that are inflected for gender, number, and case (Höder 2009). An additional factor in the emergence and spread of literacy was the religious nature of many of the early texts, which contributed to the divergence between formal written registers. Nevertheless, also within religious texts, several registers may be employed by the same author (narrative, argumentative, devotional, etc.). Highlighting the earliest stages of the appearance of specific registers in the vernacular can thus yield highly valuable insights into the accumulation of individual register knowledge over time.

QBii: How do established registers change over time? (B01, B02, B03, B04, B05)

From the perspective of language-user knowledge, we consider registers to be sets of conventional form-meaning pairings (cf. Koch 1997), used and re-used by individuals and thus subject to change over time. Accordingly, it is feasible to identify the types of change that can affect registers as well as the changes in circumstances that trigger these developments. Regarding the types of changes, a generalization is desirable on two levels: First, we aim to establish a typology of broad categories like register divergence, convergence, and alternation. Second, we must come to terms with the fact that an individual's register competence includes both conscious and unconscious knowledge of features that identify a specific register (see also QCii below). It is therefore relevant to identify bundles of features with particular behaviors in register change: Which features prove to be stable and which are more volatile? Do core grammatical features behave differently from lexical ones? How do established registers reprocess language change that affects their core features in contrast to more peripheral features? Do core features undergo a "life cycle" of erosion and reinforcement? With respect to the driving forces of register change, one might also work towards a generalized typology of causes. Do register changes brought about by establishing new societal functional systems (law, higher education, sciences) differ from changes, say, by the emergence of new means of communication (manuscript > print, print > visual and audio, offline > online)? Methodologically, the typology of changes and the typology of causes can be addressed by comparing the structural and extra-linguistic features of exemplary historical registers. Interactions with Areas A and C will be critical for investigating these questions. Area A will be especially relevant for issues of how different types of linguistic features interact with register, and Area C for questions of conscious vs. unconscious knowledge of register.

QBiii: How is register involved in language change? (B01, B03, B05, also A06)

The notion that register plays a role in language change is recognized in the variationist paradigm, in that varieties associated with certain speech styles may lead change, either “from below” or “from above”. Since in Labov’s work this contrast refers to the level of the language user’s awareness of their intra-individual variation, it naturally implies a focus on the speaker’s knowledge. We hypothesize that change from below is led by informal registers, while change from above is led by more formal registers, since the latter are associated with more attention paid by the speaker and hence more awareness. Naturally, informal language is rarely reflected in historical texts.⁶ One way of addressing this methodological limitation arises from Labov’s (2013) observation that the most vernacular form of speech is found in oral narrative of events of which the speaker has personal knowledge, most especially when these events revolve around three “universal centers of interest”: (i) death and danger of death (including violence, illness, etc.); (ii) sex (including marriage, affairs, etc.); (iii) moral indignation (blame, injustice, social norms etc.). This is of clear interest to diachronic linguistics. The implication of Labov’s findings is that discourse contexts relating to these centers of interest reflect a more vernacular register, and that their features may then spread through the community by ordinary linguistic diffusion. A prediction of this hypothesis is that new variants of a variable will have a higher concentration in discourse contexts that relate to the three centers of interest referred to above than elsewhere. Turning to change from above, we will study it mainly in the historical context of opening up vernacular European languages to registers previously reserved for prestige languages (e. g. Latin, French). In the process, the erstwhile vernacular languages acquired morphological alternatives, syntactic constructions, and vocabulary borrowed from the prestige languages. In some cases the register distribution of a novel construction may be able to serve as a diagnostic of whether it was initially a calque on a construction in a prestige language – and thus characteristic of highly formal, especially written registers – or a native innovation (and thus more widely distributed across registers (see Smith 2007: for relevant discussion on the origins of the English progressive). Our focus on register knowledge and competence will lead us to look at these phenomena from the point of view of the historical agents of these changes as well as their choices.

⁶Increasingly, however, researchers are making efforts to search out and study instances of informal language recorded in specific text types, e. g. trial proceedings and collections of correspondence (Ágel & Hennig 2006a,b; Szczepaniak & Barteld 2016; Säily et al. 2017).

The projects that deal with these three historical facets of register knowledge will feed CRC *Registers* endeavor to create a generalized theory of register knowledge. The findings from the historical projects are furthermore expected to have interesting implications for established theoretical frameworks that have been developed to model language variation and change. Drawing on statistical methods proposed in, among others, Gries & Stefanowitsch (2004); Hilpert (2013), these projects aim to identify distributional differences and collostructional preferences that are salient for specific registers. A language user's register knowledge may constrain the historical development of a language, so that accounts of one may imply predictions for the other. As already noted, Kroch's Competing Grammars theory e. g. models the co-existence of distinct varieties in the same individual's mind (Kroch 1989: et seq.). Yang (2000: et seq.), building on Kroch's work, develops specific models of how inter- and intra-individual variation influence acquisition and thus effect change. According to Adger (2014), in contrast, change takes place where there is variability with respect to choices that one could make within a single grammar to realize a particular structure, rather than choices between distinct grammars. Construction Grammar models of grammatical change (e. g. Gries & Hilpert 2008; Hilpert 2013) assume that in multi-layered form-meaning pairings at various levels of abstraction some of the form-meaning associations change over time. They also assume that constructions are distinct entities in the speaker's competence, as opposed to being merely analytical distinctions. Finally, the diasystematic approach of Höder (2012: et seq.) is based on the individual's singular linguistic system, which can embed lectal variants of any construction. All of these models put the language user's knowledge of variation — for register or otherwise — into sharp focus, along with how this knowledge is deployed to choose one of a series of candidate elements in a given context. Nonetheless, they make distinct predictions on points of detail, and thus an important focus of the work in Area B will be to tease them apart and test them on the historical data that will be examined. All of these approaches relate diachronic change to synchronic variation, for which the lessons learned from the projects in Area A will be indispensable. In this regard, the cross-linguistic perspective (e. g. in A06) will lead to interesting generalizations. The identification of linguistic variables that are cross-linguistically tied to register variation could hint towards typologically frequent patterns and trajectories of grammatical change. And of course any conclusions we draw about how knowledge of variation is structured must be critically examined based on the acquisition and processing evidence from Area C.

2.4.3 Area C: Register and cognition

This area of CRC *Register* is concerned with the role that register knowledge plays in the acquisition and processing of linguistic knowledge as a whole. The central research question for the projects in this area is:

QC: How is register knowledge acquired, represented and accessed?

Specifically, we will use different experimental designs and corpus evidence to study how registers are acquired and processed by kindergarten and school children (C01, C03), young adults (C03, C04, C05), and learners of a foreign language (C04), targeting the role of formality (C01, C03, C04), topic (C05), mode (C04, C05), or the interlocutor in production (C02, C05, C06) and perception (C02, C03, C06). We will study register effects in phonetics (C02, C06), the lexicon (C01, C02, C03, C04, C05, C06), and grammar (C01, C02, C03, C04, C05, C06). The hypotheses tested in Area C will be informed by theoretical considerations connected to those in Area A, and the results from Area C will feed into the models in Area A and will shed light on the mechanisms of language change in Area B.

We can break down QC in terms of three more specific questions:

QC_i: How do situational parameters influence register features in production, perception, and acquisition? (C01,C02, C03, C05, C06, also A06)

As stated above, we assume a broad definition of situational and functional influences on register knowledge, evidenced here by linguistic behavior. This means that we will have to test *whether* and *how* specific situational parameters influence production and perception. Some of the projects in Area C will do this in a fairly direct fashion, using experimentally set-up elicitations, which will subsequently be stored as spoken corpora and analyzed using corpus-analysis techniques as well as (phonetic) measurements (C02, C06). These projects will investigate linguistic behavior on all levels in situations that vary only in one parameter. The parameters that will be experimentally varied here are formality (expressed by clothing style and appearance, another social index, C02) and the linguistic background of the interlocutor (C06). Using different experimental techniques, project C03 will examine how both situational-functional variation and register knowledge impact language comprehension. We examine the effects of visual and linguistic register-inducing context on register-related mental representations in language comprehension using eye-tracking and event-related brain potentials. The theoretical goal of C03 is to enrich an existing processing

account of language with suitable representations to accommodate such register-related context effects including intra-individual variation and variation across the lifespan. In a similar vein, project C05 investigates the effects of social role and informational needs of the addressee as tenor-related situational variables on the probabilistic configuration of linguistic choices in written explanations of grammatical issues produced by students enrolled in primary teacher programs.

QCii: How is register knowledge acquired? What is the role of linguistic experience and awareness? (C01, C04, C05, C06)

Three projects in Area C deal with different aspects of the acquisition of register knowledge (C01, C04, C05). We assume that insights into the acquisition of register will help us understand what kinds of knowledge are necessary, which role the input plays, and which aspects of register knowledge are explicitly accessible. It has long been established that even very young children are able to distinguish between registers based on the social role of participants in familiar situations (Andersen 1986; O'Neill 1996; Wagner et al. 2010; 2014; Abbot-Smith et al. 2016). Project C01 will study the early acquisition of register phenomena in three different languages (monolingual German and bilingual Japanese- and Russian-German) with a focus on formality and politeness, and looking at register awareness in addition to the implicit, purely experience-based acquisition of register knowledge. The growth of linguistic awareness is associated with the development of cognitive academic linguistic proficiency (Cummins 2008). During school, children and adolescents undergo processes of enregisterment, wiring together novel situation types and appropriate linguistic behavior, which ensure participation in more sophisticated scientific and socio-cultural discourses. Project C05 will study the development of languages for specific purposes in university students. The project targets the effects of register-determining factors acquired earlier (social status, informational needs) or later (literacy, educational contexts) on the linguistic flexibility of young adults as a vocationally relevant communicative skill (Abbot-Smith et al. 2016; Alamillo et al. 2013; Jisa 2004; Kaplan & Berman 2015) and will investigate how salient, deliberately used linguistic features interact with less salient, implicit elements of register knowledge in the acquisition and use of scientific languages (cf. Biber 2006; Bhatia et al. 2011; Feilke 2012). C05 expands the common model of situational features by including further speaker-specific variables such as personal traits and motivation in order to distinguish the effect of instructed acquisition of scientific registers on linguistic flexibility from other individual properties influencing goal-directed behavior. In a final acquisition project (C04), we will study register knowledge

in late L2-learners — language users who have register knowledge in their first language but cannot have the same amount and type of linguistic input in their L2 that native speakers have. It has been hypothesized that learners make fewer register distinctions than native speakers (Gilquin & Paquot 2008; Deshors 2015), but there are not enough studies to show under which circumstances this holds and whether there might be registers that are more easily acquired than others. Closely related to the study of implicit and explicit acquisition of register knowledge and salience of register features is the question of whether speakers modify their linguistic behaviour automatically or how much awareness and control they have over subtle register differences in production. This will be investigated in C06 using an experimental paradigm in which speakers are distracted with additional memory tasks. A related question has been what mental representations are implicated in bi- and mono-lingual language processing (see Alexiadou & Lohndal 2016; 2018) and across languages (e. g. Yatsushiro et al. 2017). Regarding acquisition, the debate has centered on how children acquire and process knowledge of their language (e. g. Gleitman 1990; Pinker 1984; Tomasello 1992; Knoeferle 2015; Özge et al. 2016 for a review).

QCiii: How are register differences perceived? (C02, C03, C06, also A02)

Most existing register studies as well as most studies in CRC *Register* look at language production. But language production can only be understood if comprehension is also understood — acquisition (and salience) crucially rely on linguistic experience and its interpretation. This phase of the CRC therefore includes four projects that study register comprehension (we hope to include more comprehension studies in the following phases). The time is ripe for such studies because research on language acquisition, use, and processing has — much like linguistic theory — focused its efforts on phonetic, phonological, morphological, syntactic, semantic and pragmatic knowledge of standard language varieties and has come to understand that ‘standard’ varieties only cover a small part of linguistic behavior.

Experiments on standard language have, for example, shown that discourse context rapidly modulates language comprehension (Altmann & Steedman 1988; Kaiser & Trueswell 2004; van Berkum et al. 1999; Spivey & Tanenhaus 1998). The debate about context influences has, for a long time, centered on assessing whether discourse and referential context modulates initial parsing decisions (Frazier & Fodor 1978; Friederici 2002; Garnsey et al. 1997; McRae et al. 1998; Tanenhaus et al. 1995; Spivey et al. 2002; Trueswell & Tanenhaus 1994). Ensuing

research has tested context effects in standard language further, and has corroborated rapid context effects on language perception and processing (e. g. Knoeferle & Guerra 2016; Huettig et al. 2011). In brief, linguistic, world, and immediate-situation-based knowledge all rapidly modulate standard language processing. EEG and eye-tracking studies have added to this state of the art by revealing more about the nature of these situation-sensitive processes (Amsel 2011; Anderson et al. 2011; Holler et al. 2009; Kelly et al. 2015; Knoeferle et al. 2008; Mishra et al. 2012). Project C06 will explicitly test whether register shift that aims at enhancing and simplifying speech does indeed improve intelligibility for addressees of varying levels of non-native listeners.

The insights into standard language processing mechanisms and associated context effects offer a solid baseline for examining situation-specific register comprehension (e. g. comprehension of language register in a formal vs. informal situation). Research in Area C exploits empirical and experimental methods to address QCiii since these have proven successful for investigating real-time language processing, language use, and acquisition, not only in purely linguistic and visual contexts but also in socially situated contexts (Van Berkum et al. 2008; Carminati & Knoeferle 2013; Staum Casasanto 2008). Using a range of empirical measures, C03 examines the processing of situation-register congruence and compares this to the processing of congruence between language and grammatical knowledge. A02, C02 and C06 also address register comprehension in parts of their projects.

2.5 Methodologies

As register variation can only be observed and studied in language use, all of the projects in CRC *Register* will use experimental or corpus-based methods, and – with the assistance of the INF project – we will share and combine methods as much as possible (as has been advocated in many recent articles, see e. g. Wulff 2010; Gilquin & Gries 2009 or the chapters in Krug & Schlüter 2013). In the following we will briefly describe the main methods we will be using, outlining the potential benefits and possible challenges. We will then turn to a discussion of some of the particular methodological challenges raised by work on register and how we intend to meet them, not just by adopting existing tools, but also by extending them and developing and testing new ones, such that the methodology of the study of register will be an active area of research itself. The methods that we adopt are not independent of each other, and we will try to use common methods for storage and analysis wherever this may be useful.

Corpus studies. Several projects will make use of written or spoken corpus data (sometimes with videos) – some will analyze existing corpora (A01, A03, A04, A05, A06 all Area B projects, C01 and C04), often adding supplemental annotations, and some will (additionally) collect corpus data (A02, A06, C05, C06). For example, A02 will acquire audio-visual corpus data by observing the linguistic behaviour of a small number of speakers in different settings, like in public speeches and private conversations of different nature. In addition to careful qualitative and mono-variate investigations of single variables, we will carry out multidimensional studies (cf. e. g. Biber & Conrad 2009; Neumann 2014) where several linguistic variables will be analyzed to find co-occurring features, which then will be interpreted with regard to register. These studies will help us understand how situational and functional properties influence the different linguistic levels. We will evaluate the data using multivariate studies and mixed-effect models (Gries 2015; Szmrecsanyi 2013, see also Ruetten & Speelman 2012), where different extra-linguistic variables are aggregated and interpreted with regard to their influence on a given linguistic variable (see INF and A04 for discussion). A06 will also use Bayesian Network Modeling.

Corpora based on spontaneous, naturally occurring speech and writing can provide a large volume of useful data and give crucial insights into the combination of register features and factors that influence production in authentic settings. An important complication, however, is that it is often not possible, especially with pre-existing resources, to exclude or even identify and control for additional confounding factors. For this reason, corpus analysis is best complemented by data collected in more controlled settings, as we will now describe.

Fieldwork, sociolinguistic interviews, and controlled collection of (spoken) data. Some projects will conduct sociolinguistic interviews in the Labovian sense or experimentally based collections of spoken/video data. An example for such structured data collection is C06: A participant will interact with two previously instructed interlocutors (foils) in a diapix setting (i. e. a “spot the difference” task) in order to discover the differences in production triggered by one specific difference of the interlocutor (their L1 in this case). (Parts of) these interviews will be transcribed and stored, annotated and analyzed as corpora. Phonetic features of these interviews will be analyzed using PRAAT or similar programs. In contrast, C05 uses the same task and varies the level of three situational variables, thus eliciting spoken and written data structured accordingly to the factorial design. As a third example, A02 will work with a selected group of speakers that presumably have above-average experience with register variation, namely actors.

It will make use of storyboards that are not focused on the truth-conditional content of a story but on the way that different characters would act in the story or tell parts of the story.

Using such data collection methods, we will be able to target specific situational and functional differences while keeping other parameters stable. With these techniques one can find out *whether* a given situational parameter has an influence on production and *which specific influence* it triggers. But even in fairly structured collections, there will be many influencing factors that are beyond our control (e. g. even if we try to keep properties such as gender, age, background, etc. stable, the interlocutors in the C06 experiment will differ in many ways besides their L1). In order to get a complete picture, these findings will have to be supplemented by even more controlled production data and, crucially, also perception data.

Experimental techniques. The projects in CRC *Register* will thus also employ a range of experimental methods. Several projects will collect judgments using the matched guise technique, following Lambert et al. (1960) (A03, A05, A06, C01). The term matched guise technique refers to a setup where two target stimuli are presented in otherwise identical settings, and participants judge the stimuli on one or several relevant scalar qualities such as formality, politeness, intelligence and many others. This tightly controlled setup has made it possible to discover many interesting and subtle influences. But to uncover a classification of register phenomena that can be mapped to other competence phenomena and to verify theoretical accounts of register knowledge, we will need to make use of the full array of modern phonetic and psycholinguistic methodology, in addition to careful and controlled corpus analysis. Acceptability, intelligibility and assessment studies (using e. g. semantic differentials) and forced choice experiments such as IATs (Implicit Association Tasks) will be conducted in a number of projects (e. g. A03, A04, A06, C02, C06) in order to explore the register sensitivity of specific linguistic phenomena and to validate register features or functional and situational conditions in general. These experiments will take situational and linguistic contexts into account which have been singled out in previous corpus-linguistic analyses and/or matched guise tests. Furthermore, reaction time studies can clarify interpretive preferences in cases of ambiguity introduced by register phenomena (e. g. the metaphoric versus literal interpretation of *sauer* in our paradigm example, cf. Glucksberg 2003) (C03), and studies using EEG and eye-tracking can investigate similarities and differences between register processing and the processing of syntactic and semantic structure (C03).

Moreover, to delineate register flexibility from more general cognitive and social skills we need a better understanding of the interaction between situa-

tional factors and properties of language users such as personal traits, language knowledge, age, attitudes or achievement motivation that are expected to influence intra-individual linguistic variation (Adli 2004; Mercer et al. 2012; Kaplan & Berman 2015). As factors influencing goal-directed behavior, they bear directly on the strategic aspects of register variation that range from social indexing and identity construction to audience design. A better understanding and control of their influence on speakers' propensity for situationally optimised language is an important aspect in the currently proposed experimental approaches to the development of specialised linguistic repertoires, metalinguistic awareness and linguistic flexibility as constitutive elements of register knowledge (C05). The age and language knowledge of the language user are also relevant in C01 which looks at register behavior of mono- and bilingual children and C03 which considers L2 and age. This aspect will be even further emphasized in Phase 2 of CRC *Register*.

As noted above, the study of register is associated with several particular methodological challenges which have only partially been touched on by previous work. Thus we see it as important that CRC *Register* not simply adopt existing methodologies as a means to pursuing its theoretical research goals, but additionally extend existing tools and develop new ones. That is, the methodology for investigating register will be an area of active research itself, where we expect to achieve important advancements for the field, pursuing dedicated methodological research questions, as we will briefly lay out here. One central question, in particular for the several projects working with corpora, is the following:

QMi: How do we reconstruct intra-individual register variation on the basis of textual and experimental data?

As laid out in detail in the discussion of the first core issue in Section 2.3 above, CRC *Register* sees register variation as a type of intra-individual variation, which however bears a special relationship to inter-individual variation. That is, the patterns of register variation are shared across the language users of a community, such that it is a type of intra-individual variation that is more or less consistent inter-individually. The focus on intra-individual variation presents serious methodological challenges to any study based on a (limited) collection of texts. This is because it implies restrictions on the quantity and thematic breadth of text available from any given individual, especially in the case of the historical projects, to the point that we may have only one text from many or even most of the individuals whose output is contained in the corpus. This means that we

will in fact have very little data on true intra-individual variation of the relevant type.

A fairly direct way to deal with this issue is to focus on longer texts which are subdivided by purpose or other situational parameters (or multiple, situationally distinct texts) by single speakers. This allows a very fine-grained analysis of true intra-speaker variation and will thus be pursued by several projects (e. g. **B01**, **B02**, **B03**, **B04**, **B05** but also partly **A01**, **A02**, and similarly for structured data collections in the same vein see **C02**, **C04**, **C06**). That strategy is, however, limited in its applicability by the availability of such longer, register-diverse texts. To supplement it, we must thus (additionally) rely on collections of texts which differ in situational-functional parameters, but were written by several different authors. At a basic level, such data of course only provide direct evidence only for inter-speaker variation. Nonetheless, careful analysis of such collections of texts, supported by appropriate statistical methods, can yield at least indirect evidence for intra-speaker variation, precisely because register variation is inter-individually consistent. That is, we can often infer intra-individual variation of the relevant kind by comparing stretches of text reflecting distinct situational contexts even if they were written by different individuals. However, these inferences must be drawn carefully and subjected to sanity checks based on comparisons with those cases where we do have multiple types of text from a single author.

An even more basic methodological challenge shared by all of the projects in CRC *Register* that rely (at least in part) on existing text (corpora) is the question of how we can identify register distinctions, i. e. the situational and functional settings that steer the type of variation we are interested in, within the naturally occurring language recorded in texts. In experimental settings one can directly manipulate and document the context within which linguistic behavior occurs, but this is not the case with text that is produced for independent purposes and then collected and catalogued by the researcher. Depending on the circumstances, a certain amount of information may be available about the situation in which and for which the text was produced, but this information will always be limited to a greater or lesser extent. The problem is especially acute for the projects concerned with data on historical languages, since we generally have less information about the circumstances surrounding the creation of text the older it is, but the issue arises in principle even for carefully documented contemporary text. These issues are of course significantly mitigated when corpora are available that have been constructed and annotated with register explicitly in mind, e. g. the ARCHER and CORE corpora (see project **B01**), RuMorS (Morfologičeskij standart), Russian National Corpus, Czech National Corpus (see

project A03) as well as the RUEG corpus, once it is available, and so the projects of CRC *Register* will make use of such corpora wherever possible (see e. g. the discussion of the use of the different English historical corpora in B01).

However, in many cases, we must rely (additionally) on corpora that are not annotated for register. Even with those that are, we cannot simply take that information for granted and uncritically adopt the understanding of register that it is built on. After all, our overall goal in CRC *Register* is to advance the state of the art in the field's understanding of register and specifically register knowledge, which means re-examining existing models and methodologies at every step. Among other things, we must develop tools for identifying register categories for arbitrary texts, apply these to the corpora that lack register information and compare them to the results for the corpora that already include annotation of register. This should rely on a combination of text-external and text-internal types of information. As one starting point, e. g. we can take text genres as an approximate stand-in for register, but we must develop more sophisticated procedures that go beyond this. On the one hand, we know that the register adopted can shift, sometimes multiple times, within a single text, and on the other hand, in some cases there may well be multiple text genres that use the same register.

Again, these various issues associated with QMi represent general methodological challenges for CRC *Register*, and an important part of the work across projects in phase 1 will be to develop and test further methods for extracting useful information about register-sensitive intra-individual variation from corpora and from existing texts in general. This will thus be an important point of collaboration and interaction for all projects that rely on corpus data, and will also be the topic of the workshop "Data-driven analysis of registers (MDA and beyond)", to be organized by A04 in collaboration with A01, A03, B01, and INF. While this is an issue where we hope to come up with important novel ideas in the course of our research, at this point we can already briefly discuss three strategies that will be employed wherever possible: (a) hypothesis-led corpus construction, (b) text-internal segmentation with explicit annotation of the different registers in a single text, and (c) the combination of corpus-based methods with experimental methods.

As noted above, a number of the projects (A02, A06, C02, C05, C06) will collect novel corpus data, and this will allow them to make use of *hypothesis-led corpus construction*. Under this approach, only one parameter – the parameter of interest – is explicitly varied (the fact that there may or will be other parameters which vary in a non-controlled way of course cannot be ignored). In some cases, e. g. the same speakers are asked to produce different text types (similar to

what Wiese 2020 suggests). A06 systematically implements intra-speaker variation in this way by collecting production and perception data (several registers) with same speakers. In other cases, the same speaker is asked to produce the same type of text or perform the same task while the interlocutor is varied in a minimal way. In C02, the interlocutor is actually the same person but with different clothing that implies a distinct social standing and level of formality. In C06, two interlocutors are used, one a native speaker of German and the other an advanced learner of German as a Foreign Language.

In the cases where we work with existing corpora that have not already been (sufficiently) annotated for register, we will undertake a *text-internal segmentation*, i. e. we will identify and annotate distinct registers within a single text. Since the individual text in such cases typically was produced by a single author for a single audience, the parameter which can shift within the text and thus is to be investigated will be the purpose or goal. Text-internal categorization of this kind can be done using explicit structuring criteria such as headers or specific parts of an interview, as presented in Labov's sociolinguistic interviews (Labov 1984) or shown in the discussion of the different sections in academic research articles (Biber & Conrad 2009) or the chapters in Otfried's Evangelienharmonie (Schnelle 2018). Sometimes the relevant or interesting segments are more difficult to detect and occur within a stretch of text that is not explicitly marked off. Examples are the "narrative" (where the preacher retells a biblical story) and the adhortative (where the preacher addresses the audience with advice on how to live) parts of a sermon (cf. Hirschmann et al. 2012). In CRC Register, this would pertain to sermons (A01), stretches of rendered speech within narratives (B01, A06), early Romance epics (B05) and learner essays (C04). Such text-internal classification presents challenges and there will be segments where the appropriate division into distinct registers is not immediately clear. In these cases, we will have to define independent criteria for segmentation which, in turn, will have to be evaluated and documented. All projects that use text-internal segmentation will cooperate on this methodology, jointly developing improved procedures for the task. It will be especially important for the projects in Area B, as it is not possible with historical languages to collect new hypothesis-led corpus data or to use experimental techniques, due to the lack of living native speakers.

Finally, in accordance with recent developments (Gilquin & Gries 2009; Arppe & Järvikivi 2007), we will *combine methods* wherever possible (A03, A04, A06, C01, C06). Existing corpus data can be used to explore, understand and preliminarily model a phenomenon which makes it possible to formulate specific hypotheses that can then be tested by experimental methods and by the hypothesis-led construction of new corpora.

In addition to challenges when studying register phenomena in a (limited) collection of texts, we also expect methodological challenges in controlled experiments examining register effects. A first challenge is that register effects may be subtle. It is unclear whether the status of this sort of knowledge (and the mechanisms implicated in its acquisition and language use) is similar to standard lexical-semantic, syntactic, and pragmatic knowledge. For the latter aspects of linguistic knowledge we know robust effects emerge (e. g. reading *Die Vater* 'the.FEM father' is immediately perceived as incorrect). The flexibility of language use may be greater regarding register than, for instance, regarding grammatical phenomena or truth-conditional meaning. If so, we may see less robust register-related language use compared with the application of our standard knowledge of grammar and meaning. More variability presents challenges for statistical analyses. We embrace this challenge by conducting power analyses (based on pilot data) to assess the appropriate sample size given the register phenomena under investigation. We also use, for example, linear mixed effects models and minF statistics to accommodate variability by participants and by stimuli within one analysis (see also INF).

Different presentation formats can further provide insight into similarities of register and standard language processing. For standard grammatical and semantic knowledge we know that even when people face variable input (e. g. a grammatical sentence followed by ungrammatical input), they immediately respond to it overtly (e. g. when judging sentence acceptability) and covertly (e. g. in their eye movements and brain responses). We do not yet know whether such robust responses occur for register variants that may or may not fit situational-functional constraints. We address this, for instance, by presenting stimuli once in a single block (e. g. formal situations only) and once interleaved (formal and informal situation, see C03 for an example). Blocked compared with interleaved presentation of language variety can contribute to assessing the robustness and similarity of register processing compared with standard language processing.

Another way of assessing the status and variability of register competence relative to standard linguistic knowledge is to employ tasks and stimulus manipulations for which we already know what to expect for standard competence. Participants' responses to text or speech that draws on their standard linguistic competence can then be systematically compared to how they respond when they perceive or produce register variants. Among the tasks that the projects employ are truth-value judgements/sentence verification, sentence repetition, story elicitation, picture completion, and felicity judgement tasks (see A03, A04 and the Area C projects).

A further challenge that we meet is control over random variability in corpora. As noted above, corpus data based on naturally occurring production is never as clean as lab-produced experimental data. We will thus rely on field and laboratory experimentation to complement the important insights from the analyses of corpora. In-lab experiments using behavioral tasks and collecting behavioral and neuroscientific data can control variability. For instance, in **C03**, we create stimuli that vary in formality and assess situation-formality effects on the ensuing comprehension of standard language and register variant. **C01**, as another example, examines in a controlled setting the extent to which mechanisms implicated in the acquisition of register competence are shared or not and examines — in a controlled setting — the contribution of linguistic experience to this process.

QMii: What are the register specific statistical features of (corpus) data? Which methods are most appropriate to analyze them and what does this tell us about the nature of register? Can and should those methods be enhanced and expanded?

At the outset, we plan to use statistical methods that are long established in register studies (such as Biber's multidimensional analysis), apply them to new data, compare results between languages and corpora and, most importantly, address their limitations (e. g. **A01**, **A03**, **A04**, **B02**, **B03**, **C04**, **C05**, **C06**). As noted above, in several projects this will crucially include cross-checking the results provided by these initial statistical analyses by the targeted application of experiential methods. **A03** e. g. will confront the factors initially identified in a Biber-style MDA analysis with judgments from a matched-guise study.

From there, we will move on to explore novel methods for identifying and analyzing register based on sophisticated statistics. Crucially, *CRC Register* brings together researchers with proven skills and experience in the advanced statistical modeling of language data, and we will use this expertise to adapt and combine statistical methods in novel ways for use in the study of register phenomena. This major methodological focus will help us overcome limitations inherent in established approaches. In this context it is important to keep in mind that even Biber's MDA, one of the best-known data-intensive approaches in register research, is based on a standard statistical procedure (factor analysis), which is applied to language data in a specific way, followed by careful linguistic interpretation. *CRC Register* plans to make use of more recent and more sophisticated statistical approaches and to explore ways in which they can most profitably be applied in the specific context of register studies.

For instance, project A04 plans to use Latent Dirichlet Allocation (Bayesian generative modeling), which offers a way to uncover hidden variables controlling the language generation process. The method, first used in genome analysis, is applicable to many kinds of data and has become popular in the Digital Humanities for analyzing the topical structure of large collections of documents. In the context of register research, the hidden variables are not text topics but registers, and the observed variables are not only words, but linguistic information of all kinds, including typical Biber features and POS-ngrams. Applying such LDA techniques in the field of register studies is both promising and highly innovative. The methodological contribution includes the proof of concept, an assessment of which and how many linguistic features are necessary to obtain stable results, exploration of a number of important parameter settings, as well as an evaluation of the induced registers in terms of cognitive plausibility. The latter question will be tackled using psycholinguistic experiments.

The INF project will make central contributions to addressing QMii and will coordinate and facilitate work related to it in the various individual projects and communication about relevant results, issues and best practices across the CRC. See the INF project description for further discussion of relevant issues.

3 Conclusion

With this extract from the CRC 1412 frame text, we gave an extensive overview of the literature on register research and provided our own take on register with regards to the issue of register knowledge and its relation to situational-functional aspects of language. This paper highlights several research strands for register in linguistics, in particular the three research areas with the relation between register and a) grammar explored in Area A, b) historical aspects researched in Area B and c) cognition assessed in Area C. We further mapped out a series of open questions that have yet to be answered and posited hypotheses to be tested in the course of the CRC.

Funding information

Funded by the Deutsche Forschungsgemeinschaft (DFG) – SFB 1412, 416591334

Acknowledgements

A CRC proposal is always a strong collaborative effort with a long history – beginning with the first ideas, many discussions, stages, revisions, more discussions, and more revisions. In addition to all PIs, we would like to acknowledge the contribution by a number of researchers who were actively involved in the writing of the proposal: Felix Bildhauer, Sophia Döring, Felix Golcher, Milena Kühnast, Roland Schäfer.

References

- Abbot-Smith, Kirsten, Erika Nurmsoo, Rebecca Croll, Heather Ferguson & Michael Forrester. 2016. How children aged 2;6 tailor verbal expressions to interlocutor informational needs. *Journal of Child Language* 43(6). 1277–1291.
- Adams, Jonathan. 2015. *The revelations of St Birgitta: A study and edition of the Birgittine-Norwegian texts, Swedish National Archives, E 8902*. Leiden: Brill. 624 pp. DOI: doi.org/10.1163/9789004304666.
- Adger, David. 2006. Combinatorial variability. *Journal of Linguistics* 42(03). 503–530. DOI: [10.1017/s002222670600418x](https://doi.org/10.1017/s002222670600418x).
- Adger, David. 2014. Variability and grammatical architecture. In M. Carme Picallo (ed.), *Linguistic variation in the Minimalist framework*, 179–196. Oxford: Oxford UP. DOI: [10.1093/acprof:oso/9780198702894.003.0009](https://doi.org/10.1093/acprof:oso/9780198702894.003.0009).
- Adger, David & Jennifer Smith. 2005. Variation and the Minimalist Program. In Leonie Cornips & Karen P. Corrigan (eds.), *Syntax and variation: Reconciling the biological and the social*, 149–178. Amsterdam: John Benjamins.
- 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](https://doi.org/10.1016/j.lingua.2008.05.007).
- Adli, Aria. 2004. *Grammatische Variation und Sozialstruktur*. Berlin: Akademie Verlag.
- Adli, Aria. 2013. Syntactic variation in French wh-questions: A quantitative study from the angle of Bourdieu’s sociocultural theory. *Linguistics* 51(3). 473–515.
- Ágel, Vilmos & Mathilde Hennig. 2006a. Praxis des Nähe- und Distanzsprechens. In Vilmos Ágel & Mathilde Hennig (eds.), *Grammatik aus Nähe und Distanz. Theorie und Praxis am Beispiel von Nähetexten 1650-2000*, 33–74. Tübingen: Niemeyer.
- Ágel, Vilmos & Mathilde Hennig. 2006b. Theorie des Nähe- und Distanzsprechens. In Vilmos Ágel & Mathilde Hennig (eds.), *Grammatik aus Nähe und Distanz. Theorie und Praxis am Beispiel von Nähetexten 1650-2000*, 3–31. Tübingen: Niemeyer.

- Agha, Asif. 2007. *Language and social relations*. Cambridge: Cambridge UP. 446 pp.
- Alamillo, Asela Reig, Jean-Marc Colletta & Michèle Guidetti. 2013. Gesture and language in narratives and explanations: the effects of age and communicative activity on late multimodal discourse development. *Journal of Child Language* 40(3). 511–538.
- Alexiadou, Artemis. 2017. Language variation and change: A case study on the loss of genitive case in Heritage Greek. *Belgian Journal of Linguistics* 31. 54–72.
- Alexiadou, Artemis & Terje Lohndal. 2016. Editorial: The grammar of multilingualism. *Frontiers in Psychology* 7. 1397. DOI: [10.3389/fpsyg.2016.01397](https://doi.org/10.3389/fpsyg.2016.01397).
- Alexiadou, Artemis & Terje Lohndal. 2018. Units of language mixing: A cross-linguistic perspective. *Frontiers in Psychology* 9. 1719.
- Altmann, Gerry & Mark Steedman. 1988. Interaction with context during human sentence processing. *Cognition* 30(3). 191–238. DOI: [http://dx.doi.org/10.1016/0010-0277\(88\)90020-0](http://dx.doi.org/10.1016/0010-0277(88)90020-0). <http://www.sciencedirect.com/science/article/pii/S0010027788900200>.
- Amsel, B. D. 2011. Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia* 49(5). 970–83. DOI: [10.1016/j.neuropsychologia.2011.01.003](https://doi.org/10.1016/j.neuropsychologia.2011.01.003). <http://www.ncbi.nlm.nih.gov/pubmed/21219919>.
- Andersen, Elaine S. 1986. The acquisition of register variation by Anglo-American children. In Bambi B. Schieffelin & Elinor Ochs (eds.), *Language socialization across cultures*, 153–161. Cambridge: Cambridge UP.
- Anderson, Sarah E., Eric Chiu, Stephanie Huette & Michael J. Spivey. 2011. On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychologica* 137(2). 181–189. DOI: <http://dx.doi.org/10.1016/j.actpsy.2010.09.008>. <http://www.sciencedirect.com/science/article/pii/S0001691810001915>.
- Arppe, Anti & Juhani Järviö. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2). 131–159.
- Auer, Peter. 2003. Türkenslang: Ein jugendsprachlicher Ethnolekt des Deutschen und seine Transformationen. In Annelies Häcki Buhofer (ed.), *Spracherwerb und Lebensalter*. Tübingen, Basel: Francke.
- Auer, Peter. 2007. Style and social identity. In Peter Auer (ed.), *Style and social identities. Alternative approaches to linguistic heterogeneity*, 325–360. Berlin: de Gruyter.
- Barsalou, Lawrence W. 2008. Grounded cognition. *Annual Review of Psychology* 59(1). 617–645.

- Barsalou, Lawrence W. 2016. Situated conceptualization offers a theoretical account of social priming. *Current Opinion in Psychology* 11(Supplement C). 6–11.
- Bayley, Robert. 2013. Variationist sociolinguistics. In Robert Bayley, Richard Cameron & Ceil Lucas (eds.), *The Oxford handbook of sociolinguistics*, 11–30. Oxford: Oxford UP.
- Behaghel, Otto. 1923. *Deutsche Syntax. Eine geschichtliche Darstellung*. Heidelberg: Carl Winter.
- Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2). 145–204.
- Beltrama, Andrea. 2016. *Bridging the gap: Intensifiers between semantic and social meaning*. University of Chicago. (Doctoral dissertation).
- Benz, Anton, Gerhard Jäger & Robert van Rooij. 2006. An introduction to game theory for linguists. In Anton Benz, Gerhard Jäger & Robert van Rooij (eds.), *Game theory and pragmatics*, 1–82. Houndmills, Basingstoke: Palgrave Macmillan. DOI: [10.1057/9780230285897_1](https://doi.org/10.1057/9780230285897_1).
- Berman, Ruth A. 2015. Language development and use beyond the sentence. In Edith L. Bavin & Letitia R. Naigles (eds.), *The cambridge handbook of child language*, 2nd edn. (Cambridge Handbooks in Language and Linguistics), 458–480. Cambridge: Cambridge UP. DOI: [DOI:10.1017/CBO9781316095829.021](https://doi.org/10.1017/CBO9781316095829.021). <https://www.cambridge.org/core/books/cambridge-handbook-of-child-language/language-development-and-use-beyond-the-sentence/D90DDC19E2C353E1A170B8DA4FFF0031>.
- 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](https://doi.org/10.1007/978-3-319-33228-4_19). https://doi.org/10.1007/978-3-319-33228-4_19.
- Berman, Ruth A. & Bracha Nir-Sagiv. 2007. Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes* 43(2). 79–120. DOI: [10.1080/01638530709336894](https://doi.org/10.1080/01638530709336894).
- Bhatia, Vijay, Purificación Sánchez Hernández & Pascual Pérez-Paredes. 2011. *Researching specialized languages* (Studies in Corpus Linguistics 47). Amsterdam: John Benjamins.
- Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge UP. DOI: [10.1017/cbo9780511519871](https://doi.org/10.1017/cbo9780511519871).
- 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](https://doi.org/10.1177/1461445699001002001).

- Biber, Douglas. 2006. *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins. DOI: [10.1075/scl.23](https://doi.org/10.1075/scl.23).
- Biber, Douglas. 2009. Multi-dimensional approaches. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook* (Handbücher zur Sprach- und Kommunikationswissenschaft 2), 822–855. Berlin: de Gruyter Mouton.
- Biber, Douglas. 2012. Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory* 8. 9–37.
- Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style* (Cambridge textbooks in linguistics). Cambridge, UK: Cambridge UP. DOI: [10.1017/cbo9780511814358](https://doi.org/10.1017/cbo9780511814358).
- 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 UP.
- Burnett, Heather. 2017. Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of Sociolinguistics* 22. 238–271.
- Burnett, Heather. 2019. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy*. 1–32. DOI: [10.1007/s10988-018-9254-y](https://doi.org/10.1007/s10988-018-9254-y).
- Campbell-Kibler, Kathryn. 2008. I'll be the judge of that: Diversity in social perceptions of (ING). *Language in Society* 37(05). 637–659. DOI: [10.1017/s0047404508080974](https://doi.org/10.1017/s0047404508080974).
- Campbell-Kibler, Kathryn. 2010. The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change* 22(3). 423–441. DOI: [10.1017/s0954394510000177](https://doi.org/10.1017/s0954394510000177).
- Carminati, Maria Nella & Pia Knoeferle. 2013. Effects of speaker emotional facial expression and listener age on incremental sentence processing. *PLoS ONE* 8(9). e72559. DOI: [10.1371/journal.pone.0072559](https://doi.org/10.1371/journal.pone.0072559).
- Cheshire, Jenny. 1999. Taming the vernacular: Some repercussions for variation and spoken grammar. *Cuadernos de Filología Inglesa* 8. 59–80.
- Coulmas, Florian. 2007. *Sociolinguistics: The study of speakers' choices*. Cambridge: Cambridge UP.
- Coupland, Nikolas. 1980. Style-shifting in a Cardiff work-setting. *Language in Society* 9(1). 1–12. DOI: [10.1017/s0047404500007752](https://doi.org/10.1017/s0047404500007752).
- Croft, William. 2010. The origins of grammaticalization in the verbalization of experience. *Linguistics* 48(1). 1–48.

- Cummins, Jim. 2008. BICS and CALP: Empirical and theoretical status of the distinction. In Nancy H. Hornberger (ed.), *Encyclopedia of language and education*, 487–499. New York: Springer US.
- Davis, Christopher & Christopher Potts. 2010. Affective demonstratives and the division of pragmatic labor. In Maria Aloni, Harald Bastiaanse, Tikitu de Jager & Katrin Schulz (eds.), *Logic, language and meaning*, 42–52. Berlin, New York: Springer. DOI: [10.1007/978-3-642-14287-1_5](https://doi.org/10.1007/978-3-642-14287-1_5).
- Deák, Gedeon O. 2003. The development of cognitive flexibility and language abilities. In Robert V. Kail (ed.), *Advances in child development and behavior*, vol. 31. 271–327. San Diego, CA: Academic Press.
- Deák, Gedeon O. & Melody Wiseheart. 2015. Cognitive flexibility in young children: General or task-specific capacity? *Journal of Experimental Child Psychology* 138. 31–53. DOI: <https://doi.org/10.1016/j.jecp.2015.04.003>. <http://www.sciencedirect.com/science/article/pii/S0022096515000946>.
- 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.
- Dürscheid, Christa. 2003. Medienkommunikation im Kontinuum von Mündlichkeit und Schriftlichkeit: Theoretische und empirische Probleme. *Z f angewandte Linguistik* 38. 37–56.
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453–476. DOI: [10.1111/j.1467-9841.2008.00374.x](https://doi.org/10.1111/j.1467-9841.2008.00374.x).
- Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41(1). 87–100. DOI: [10.1146/annurev-anthro-092611-145828](https://doi.org/10.1146/annurev-anthro-092611-145828).
- Eckert, Penelope & William Labov. 2017. Phonetics, phonology and social meaning. *Journal of Sociolinguistics* 21(4). 467–496. DOI: [10.1111/josl.12244](https://doi.org/10.1111/josl.12244).
- Eckert, Penelope & John Rickford. 2001. *Style and sociolinguistic variation*. Cambridge: Cambridge UP. 360 pp.
- Feeney, Aidan & Jean-François Bonnefon. 2012. Politeness and honesty contribute additively to the interpretation of scalar expressions. *Journal of Language and Social Psychology* 32(2). 181–190. DOI: [10.1177/0261927x12456840](https://doi.org/10.1177/0261927x12456840).
- Feilke, Helmuth. 2012. Bildungssprachliche Kompetenzen – fördern und entwickeln. *Praxis Deutsch* 233. 4–13.
- Feilke, Helmuth & Mathilde Hennig (eds.). 2016. *Zur Karriere von “Nähe und Distanz”. Rezeption und Diskussion des Koch-Oesterreicher-Modells*. Berlin: de Gruyter.
- Frank, Michael C. & Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998. DOI: [10.1126/science.1218633](https://doi.org/10.1126/science.1218633).

- Franke, Michael & Gerhard Jäger. 2016. Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Z f Sprachwissenschaft* 35(1). 3–44. DOI: [10.1515/zfs-2016-0002](https://doi.org/10.1515/zfs-2016-0002).
- 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](http://dx.doi.org/10.1016/0010-0277(78)90002-1). <http://www.sciencedirect.com/science/article/pii/0010027778900021>.
- Friederici, A. 2002. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6. 78–84.
- Garnsey, Susan M., Neal J. Pearlmutter, Elisabeth M. Myers & Melanie A. Lotocky. 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37(1). 58–93.
- Giles, Howard, Justine Coupland & Nikolas Coupland (eds.). 1991a. *Contexts of accommodation*. Cambridge: Cambridge UP. DOI: [10.1017/cbo9780511663673](https://doi.org/10.1017/cbo9780511663673).
- Giles, Howard & Nikolas Coupland. 1991. *Language: Contexts and consequences* (Mapping Social Psychology). Pacific Grove: Brook/Cole.
- Giles, Howard, Nikolas Coupland & Justine Coupland. 1991b. Accommodation theory: Communication, context, and consequence. In Howard Giles, Justine Coupland & Nikolas Coupland (eds.), *Contexts of accommodation*, 1–68. Cambridge: Cambridge UP. DOI: [10.1017/cbo9780511663673.001](https://doi.org/10.1017/cbo9780511663673.001).
- Giles, Howard, Donald M. Taylor & Richard Bourhis. 1973. Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society* 2(02). 177–192. DOI: [10.1017/s0047404500000701](https://doi.org/10.1017/s0047404500000701).
- Gilquin, Gaëtanelle & Stefan Th. Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5(1). 1–26.
- Gilquin, Gaëtanelle & Magali Paquot. 2008. Too chatty: Learner academic writing and register variation. *English Text Construction* 1(1). 41–61.
- Gleitman, Lila. 1990. The structural sources of verb meanings. *Language Acquisition* 1(1). 3–55. DOI: [10.1207/s15327817la0101_2](https://doi.org/10.1207/s15327817la0101_2).
- Glucksberg, Sam. 2003. The psycholinguistics of metaphor. *Trends in Cognitive Sciences* 7(2). 92–96. DOI: [10.1016/s1364-6613\(02\)00040-2](https://doi.org/10.1016/s1364-6613(02)00040-2).
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford UP.
- Goodman, Noah D. & Michael C. Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences* 20(11). 818–829. DOI: [10.1016/j.tics.2016.08.005](https://doi.org/10.1016/j.tics.2016.08.005).
- Gries, Stefan Th. 2015. The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10(1). 95–125.

- Gries, Stefan Th. & Martin Hilpert. 2008. The identification of stages in diachronic data: Variability-based neighbor clustering. *Corpora* 3. 59–81.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97–129.
- Grieve, Jack, Dirk Speelman & Dirk Geeraerts. 2011. A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change* 23(02). 193–221. DOI: [10.1017/s095439451100007x](https://doi.org/10.1017/s095439451100007x).
- Gutzmann, Daniel. 2015. *Use-conditional meaning: Studies in multidimensional semantics*. Oxford: Oxford University Press.
- Habermann, Mechthild. 2001. *Deutsche Fachtexte der frühen Neuzeit*. Berlin: de Gruyter.
- Halliday, Michael A. K. 1978. *Language as social semiotic: The social interpretation of language and meaning*. University Park, PA: University Park Press.
- Halliday, Michael A. K. 1989. *Spoken and written language*. Oxford: Oxford UP.
- Halliday, Michael A. K. & Ruqaiya Hasan. 1985. *Language, context and text: a social semiotic perspective*. Victoria: Deakin UP.
- Halliday, Michael A. K. & Christian Matthiessen. 2004. *An introduction to Functional Grammar*. London: Routledge.
- Halliday, Michael A. K., Angus McIntosh & Peter Strevens (eds.). 1964. *The linguistic sciences and language teaching*. London: Longman.
- Hasan, Ruqaiya. 1973. Code, register, and social dialect. In Basil Bernstein (ed.), *Class, codes, and control*, 253–292. London: Routledge & Kegan Paul.
- Henderson, Robert, Nadine Bade & Eric McCready. 2017. Dog-whistles and the at-issue/non-at-issue distinction. In *39. jahrestagung der deutschen gesellschaft für sprachwissenschaft*, 141–142. <http://dgfs2017.uni-saarland.de/wordpress/>.
- Hilpert, Martin. 2013. *Constructional change in English. Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge UP.
- Hirschmann, Hagen, Anke Lüdeling & Amir Zeldes. 2012. Measuring and coding language change: An evolving study in a multilayer corpus architecture. *Journal on Computing and Cultural Heritage (JOCCH)* 5(1), 4. 1–16.
- Höder, Steffen. 2009. Converging languages, diverging varieties: Innovative relativisation patterns in Old Swedish. In Kurt Braunmüller & Juliane House (eds.), *Convergence and divergence in language contact situations*, 73–100. Amsterdam: John Benjamins.
- Höder, Steffen. 2012. Multilingual constructions: A diasystematic approach to common structures. In Christoph Gabriel & Kurt Braunmüller (eds.), *Multilingual individuals and multilingual societies*, 241–257. Amsterdam: John Benjamins.

- Holler, Judith, Heather Shovelton & Geoffrey Beattie. 2009. Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior* 33(2). 73–88. DOI: [10.1007/s10919-008-0063-9](https://doi.org/10.1007/s10919-008-0063-9).
- Holm, Gösta. 1967. *Epoker och prosastilar*. Studentlitteratur.
- Horn, Laurence. 1972. *On the semantic properties of logical operators in English*. UCLA. (Doctoral dissertation).
- Huettig, Falk, Joost Rommers & Antje S Meyer. 2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica* 137(2). 151–171.
- Irvine, Judith T. & Susan Gal. 2000. Language ideology and linguistic differentiation. In Paul Kroskrity (ed.), *Regimes of language: Ideologies, politics, and identities*, 35–84. School of American Research Press.
- Jannedy, Stefanie & Melanie Weirich. 2014. Sound change in an urban setting: Category instability of the palatal fricative in Berlin. *Laboratory Phonology* 5(1). 91–122. DOI: [10.1515/lp-2014-0005](https://doi.org/10.1515/lp-2014-0005).
- Jisa, Harriet. 2004. Growing into Academic French. In Ruth A. Berman (ed.), *Language development across childhood and adolescence* (Trends in Language Acquisition Research 3), 135–161. Amsterdam: John Benjamins. DOI: [10.1075/tilar.3.10jis](https://doi.org/10.1075/tilar.3.10jis).
- Kaiser, E. & J. C. Trueswell. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94(2). 113–47. DOI: [10.1016/j.cognition.2004.01.002](https://doi.org/10.1016/j.cognition.2004.01.002). <http://www.ncbi.nlm.nih.gov/pubmed/15582623>.
- Kaplan, Dafna & Ruth A. Berman. 2015. Developing linguistic flexibility across the school years. *First Language* 35(1). 27–53.
- Karmiloff-Smith, Annette. 1997. *Beyond modularity: a developmental perspective on cognitive science*. Cambridge, MA: MIT-Press.
- Kelly, S., M. Healey, A. Özyürek & J. Holler. 2015. The processing of speech, gesture and action during language comprehension. *Psychonomic Bulletin & Review* 22. 517–523. DOI: [10.3758/s13423-014-0681-7](https://doi.org/10.3758/s13423-014-0681-7).
- Kemp, J. Alan. 1986. The *Tekhne Grammatike* of Dionysius Thrax: Translated into English. *Historiographia Linguistica* 13. 343–363.
- Klein, Wolf Peter. 2011. Die deutsche Sprache in der Gelehrsamkeit der frühen Neuzeit: Von der lingua barbarica zur HautbSprache. In Herbert Jaumann (ed.), *Diskurse der Gelehrtenkultur in der Frühen Neuzeit: Ein Handbuch*, 465–516. Berlin: de Gruyter.
- Knoeferle, Pia. 2015. Language comprehension in rich non-linguistic contexts: Combining eye tracking and event-related brain potentials. In Roel M. Willems

- (ed.), *Cognitive neuroscience of natural language use*, 77–100. Cambridge: Cambridge UP.
- Knoeferle, Pia & Ernesto Guerra. 2016. Visually situated language comprehension. *Language and Linguistics Compass* 10(2). 66–82.
- Knoeferle, Pia, Boukje Habets, Matthew W. Crocker & Thomas F. Münte. 2008. Visual scenes trigger immediate syntactic reanalysis: Evidence from ERPs during situated spoken comprehension. *Cerebral Cortex* 18(4). 789–795. DOI: [10.1093/cercor/bhm121](https://doi.org/10.1093/cercor/bhm121).
- 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. 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. 15–43.
- Koch, Peter & Wulf Oesterreicher. 1994. Schriftlichkeit und Sprache. In Hartmut Günther, Otto Ludwig, Jürgen Baumann, Florian Coulmas, Konrad Ehlich, Peter Eisenberg, Heinz W. Giese, Helmut Glück, Klaus B. Günther, Ulrich Knoop, Bernd Pompino-Marschall, Eckart Scheerer & Rüdiger Weingarten (eds.), *Schrift und Schriftlichkeit. Writing and Its Use. Ein interdisziplinäres Handbuch internationaler Forschung. An Interdisciplinary Handbook of International Research*, vol. 1, 587–604. Berlin: de Gruyter. DOI: [10.1515/9783110111293.1.5.587](https://doi.org/10.1515/9783110111293.1.5.587).
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3). 199–244. DOI: [10.1017/s0954394500000168](https://doi.org/10.1017/s0954394500000168).
- Kroch, Anthony. 2001. Syntactic change. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*, 698–729. New York, Oxford: Blackwell. DOI: [10.1002/9780470756416.ch22](https://doi.org/10.1002/9780470756416.ch22).
- Krug, Manfred & Julia Schlüter (eds.). 2013. *Research methods in language variation and change*. Cambridge: Cambridge UP. <https://www.uni-bamberg.de/eng-ling/research-methods/>.
- Kunz, Kerstin & Ekaterina Lapshinova-Koltunski. 2015. Cross-linguistic analysis of discourse variation across registers. *Nordic Journal of English Studies* 14(1). 258–288.
- Labov, William. 1966. *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- Labov, William. 1969. *A study of non-standard English*. Washington, D.C.: Center for Applied Linguistics.

- Labov, William. 1971. The study of language in its social context. In Joshua Fishman (ed.), *Advances in the sociology of language*, vol. 1, 152–216. The Hague: de Gruyter Mouton.
- Labov, William. 1978. Where does the linguistic variable stop? A response to Beatriz Lavandera. *Working Papers in Sociolinguistics* 44. 1–22.
- Labov, William. 1982. Building on empirical foundations. In Winfred P. Lehman & Yakov Malkiel (eds.), *Perspectives on historical linguistics*, 79–92. Amsterdam: John Benjamins.
- Labov, William. 1984. Field methods of the project on linguistic change and variation. In John Baugh & Joel Sherzer (eds.), *Language in use: readings in sociolinguistics*, 28–66. Englewood Cliffs: Prentice Hall.
- Labov, William. 2001. *Principles of linguistic change, Volume 2: social factors*. Oxford: Wiley-Blackwell. 592 pp.
- Labov, William. 2004. Quantitative analysis of linguistic variation. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics. An international handbook of the science of language and society*, vol. 1 (HSK), 6–21. Berlin: de Gruyter.
- Labov, William. 2013. *The language of life and death*. Cambridge: Cambridge UP. 248 pp.
- Lambert, Wallace E., Richard C. Hodgson, Robert C. Gardner & Samuel Fillenbaum. 1960. Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology* 60. 44–51.
- Langacker, Ronald W. 1999. *Grammar and conceptualization*. New York: Mouton de Gruyter.
- Lassiter, Daniel & Noah D Goodman. 2017. Adjectival vagueness in a Bayesian model of interpretation. *Synthese* 194(10). 3801–3836.
- Lavandera, Beatriz R. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7. 171–182.
- Legare, Cristine H., Michael T. Dale, Sarah Y. Kim & Gedeon O. Deák. 2018. Cultural variation in cognitive flexibility reveals diversity in the development of executive functions. *Scientific Reports* 8(1). 1–14. DOI: [10.1038/s41598-018-34756-2](https://doi.org/10.1038/s41598-018-34756-2). <https://doi.org/10.1038/s41598-018-34756-2>.
- Lewis, David. 1969. *Convention: A philosophical study*. Cambridge, MA: Harvard UP.
- Lüdeling, Anke. 2017. Variationistische Korpusstudien. In Marek Konopka & Angelika Wöllstein (eds.), *Grammatische Variation: Empirische zugänge und theoretische modellierung* (Jahrbuch des Instituts für Deutsche Sprache), 129–144. Berlin: de Gruyter Mouton. DOI: [10.1515/9783110518214-009](https://doi.org/10.1515/9783110518214-009).

- Lukin, Annabelle, Alison Moore, Maria Herke, Rebekah Wegener & Canzhong Wu. 2008. Halliday's model of register revisited and explored. *Linguistics and the Human Sciences* 4. 187–213.
- Maas, Utz. 2010. Literat und orat. Grundbegriffe der Analyse geschriebener und gesprochener Sprache. *Grazer Linguistische Studien* 73. 21–150.
- Macha, Jürgen, Elvira Topalović, Iris Hille, Uta Nolting & Anja Wilke. 2005. *Deutsche Kanzleisprache in Hexenverhörprotokollen der Frühen Neuzeit*. Berlin: de Gruyter.
- Mazzarella, Diana. 2015. Politeness, relevance and scalar inferences. *Journal of Pragmatics* 79. 93–106. DOI: [10.1016/j.pragma.2015.01.016](https://doi.org/10.1016/j.pragma.2015.01.016).
- McCready, Eric. 2012. Emotive equilibria. *Linguistics and Philosophy* 35(3). 243–283.
- McRae, Ken, Michael J. Spivey-Knowlton & Michael K. Tanenhaus. 1998. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language* 38(3). 283–312. DOI: [10.1006/jmla.1997.2543](https://doi.org/10.1006/jmla.1997.2543).
- Mercer, Sarah, Stephen Ryan & Marion Williams (eds.). 2012. *Psychology for language learning: Insights from research, theory and practice*. Houndmills, Basingstoke: Palgrave Macmillan.
- Milroy, Lesley. 1987. *Observing and analysing natural language*. Hoboken, NJ: Wiley. 248 pp.
- Mishra, Ramesh K., Niharika Singh, Aparna Pandey & Falk Huettig. 2012. Spoken language-mediated anticipatory eye-movements are modulated by reading ability – Evidence from Indian low and high literates. *Journal of Eye Movement Research* 5(1). 1–10.
- Neumann, Stella. 2014. Cross-linguistic register studies: Theoretical and methodological considerations. *Languages in Contrast* 14(1). 35–57. DOI: [10.1075/lic.14.1.03neu](https://doi.org/10.1075/lic.14.1.03neu).
- Nguyen, Dong, A. Seza Dođruöz, Carolyn P. Rosé & Franciska de Jong. 2016. Computational sociolinguistics: A survey. *Computational Linguistics* 42(3). 537–593. DOI: [10.1162/coli_a_00258](https://doi.org/10.1162/coli_a_00258).
- Nowak, Martin A., Joshua B. Plotkin & Vincent A. A. Jansen. 2000. The evolution of syntactic communication. *Nature* 404(6777). 495–498. DOI: [10.1038/35006635](https://doi.org/10.1038/35006635).
- O'Neill, Daniela K. 1996. Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development* 67(2). 659–677. DOI: [10.2307/1131839](https://doi.org/10.2307/1131839).

- Özge, Duygu, Jaklin Kornfilt, Katja Münster, Pia Knoeferle, Aylin Küntay & Jesse Snedeker. 2016. Predictive use of case markers in German children. In *Proc. of the 40th annual Boston University Conference on Language Development*, 291–303.
- Paul, Hermann. 1920. *Prinzipien der Sprachgeschichte*. 5th edn. Halle: Niemeyer.
- Pickering, Martin J. & Simon Garrod. 2004. The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences* 27(02). DOI: [10.1017/s0140525x04450055](https://doi.org/10.1017/s0140525x04450055).
- Pierrehumbert, Janet B. 2008. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology 7*, vol. 4–1 (Phonology and Phonetics), 101–139. Berlin: de Gruyter.
- Pinker, Steven. 1984. *Language learnability and language development* (Cognitive Science Series). Cambridge, MA: Harvard UP.
- Plag, Ingo, Christiane Dalton-Puffer & R. Harald Baayen. 1999. Productivity and register. *English Language and Linguistics* 3. 209–228.
- Rampton, Ben. 1995. *Crossing: Language and ethnicity among adolescents*. London: Longman.
- Romaine, Suzanne. 1981. The status of variable rules in sociolinguistic theory. *Journal of Linguistics* 17(01). 93–119. DOI: [10.1017/s0022226700006794](https://doi.org/10.1017/s0022226700006794).
- Romaine, Suzanne. 1982. *Socio-historical linguistics: Its status and methodology*. Cambridge: Cambridge UP.
- Ruette, Tom, Dir Speelman & Dirk Geeraerts. 2014. Lexical variation in aggregate perspective. In Augusto Soares da Silva (ed.), *Pluricentricity: Language variation and sociocognitive dimensions*, 103–126. Berlin: de Gruyter.
- Ruette, Tom & Dirk Speelman. 2012. Applying individual differences scaling to the measurement of lexical convergence between Netherlandic and Belgian Dutch. In Anne Dister, Dominique Longrée & Gérald Purnelle (eds.), *Proc. of the 11th International Conference on Textual Data Statistical Analysis*, 883–895.
- Säily, Tanja, Arja Nurmi, Minna Palander-Collin & Anita Auer (eds.). 2017. *Exploring future paths for historical sociolinguistics*. Amsterdam: John Benjamins.
- Sampson, John. 1997. «Genre», «Style» and «Register». sources of confusion? *Revue belge de Philologie et d'Histoire* 75. 699–708.
- Santos, Ava, Sergio E. Chaigneau, W. Kyle Simmons & Lawrence W. Barsalou. 2011. Property generation reflects word association and situated simulation. *Language and Cognition* 3(1). 83–119.
- Schenker, Walter. 1977. *Die Sprache Huldrych Zwinglis im Kontrast zur Sprache Luthers*. Berlin: de Gruyter.
- Schirmer, Alfred. 1911. *Wörterbuch der deutschen Kaufmannssprache auf geschichtlichen Grundlagen. Mit einer systematischen Einleitung*. London: Trübner.

- Schnelle, Gohar. 2018. *Funktional bedingte variation in der evangelienharmonie oftrids von weißenburg: Eine methodische annäherung an eine variationistische korpusbasierte registerstudie des althochdeutschen*. Humboldt-Universität zu Berlin. (MA thesis). DOI: <http://dx.doi.org/10.18452/19250>.
- Schöller, Anthea & Michael Franke. 2017. Semantic values as latent parameters: Testing a fixed threshold hypothesis for cardinal readings of few & many. *Linguistics Vanguard* 3(1), 20160072. 1–15. DOI: [10.1515/lingvan-2016-0072](https://doi.org/10.1515/lingvan-2016-0072).
- Siegroth-Nellesen, Gabriele. 1979. *Versuch einer exakten Stiluntersuchung für Meister Eckhart, Johannes Tauler und Heinrich Seuse*. Wilhelm Fink.
- Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23(3–4). 193–229. DOI: [10.1016/s0271-5309\(03\)00013-2](https://doi.org/10.1016/s0271-5309(03)00013-2).
- Smith, E. Allyn, Kathleen Currie Hall & Benjamin Munson. 2010. Bringing semantics to sociophonetics: Social variables and secondary entailments. *Laboratory Phonology* 1(1). 121–155. DOI: [10.1515/labphon.2010.007](https://doi.org/10.1515/labphon.2010.007).
- Smith, K. Aaron. 2007. The development of the English progressive. *Journal of Germanic Linguistics* 19. 205–241.
- Snell, Julia. 2017. Enregisterment, indexicality and the social meaning of ‘Howay’: Dialect and identity in northeast England. In Emma Moore & Chris Montgomery (eds.), *Language and a sense of place*, 301–324. Cambridge: Cambridge UP.
- Soliz, Jordan & Howard Giles. 2014. Relational and identity processes in communication: A contextual and meta-analytical review of communication accommodation theory. *Annals of the International Communication Association* 38(1). 107–144. DOI: [10.1080/23808985.2014.11679160](https://doi.org/10.1080/23808985.2014.11679160).
- Spivey, M. J., M. K. Tanenhaus, K. M. Eberhard & J. C. Sedivy. 2002. Eye-movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology* 45. 447–481.
- Spivey, Michael J. & M. K. Tanenhaus. 1998. Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24. 1521–1543.
- Staum Casasanto, Laura. 2008. Does social information influence sentence processing? In B. C. Love, K. McRae & V. M. Sloutsky (eds.), *Proc. of the 30th Annual Meeting of the Cognitive Science Society*, 709–804. Washington DC: The Cognitive Science Society.
- Szczepaniak, Renata & Fabian Barteld. 2016. Hexenverhörprotokolle als sprachhistorisches Korpus. In Sarah Kwekkeboom & Sandra Waldenberger (eds.), *Per-*

- spektivwechsel oder: die wiederentdeckung der philologie. bd. 1 sprachdaten und grundlagenforschung in historischer linguistik.* 43–70. Berlin: Erich Schmidt.
- Szmrecsanyi, Benedikt. 2013. Aggregate data analysis in variationist linguistics. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 433–455. Cambridge: Cambridge UP.
- Szmrecsanyi, Benedikt. 2017. Variationist sociolinguistics and corpus-based variationist linguistics: Overlap and cross-pollination potential. *Canadian Journal of Linguistics/Revue canadienne de linguistique*. 1–17. DOI: [10.1017/cnj.2017.34](https://doi.org/10.1017/cnj.2017.34).
- Taavitsainen, Irma, Turo Hiltunen, Anu Lehto, Ville Marttila, Päivi Pahta, Maura Ratia, Carla Suhr & Jukka Tyrkkö. 2014. Late Modern English medical texts 1700–1800: A corpus for analysing eighteenth-century medical English. *ICAME J* 38. 137–153. DOI: [10.2478/icame-2014-0007](https://doi.org/10.2478/icame-2014-0007).
- Tagliamonte, Sali A. 2016. *Making waves: The story of variationist sociolinguistics*. Malden, Oxford: Wiley-Blackwell.
- Tanenhaus, M. K., M. J. Spivey-Knowlton, K. Eberhard & J. C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268. 632–634.
- Teich, Elke, Stefania Degaetano-Ortlieb, Peter Fankhauser, Hannah Kermes & Ekaterina Lapshinova-Koltunski. 2016. The linguistic construal of disciplinarity: A data-mining approach using register features. *Journal of the Association for Information Science and Technology* 67(7). 1668–1678.
- Tomasello, Michael. 1992. The social bases of language acquisition. *Social Development* 1(1). 67–87. DOI: [10.1111/j.1467-9507.1992.tb00135.x](https://doi.org/10.1111/j.1467-9507.1992.tb00135.x).
- Trudgill, Peter. 1974. *The social differentiation of English in Norwich*. Cambridge: Cambridge UP.
- Trudgill, Peter. 1983. *Sociolinguistics: An introduction to language and society*. London: Penguin Books.
- Trudgill, Peter. 2002. *Sociolinguistic variation and change*. Edinburgh: Edinburgh UP.
- Trueswell, John C. & Michael K. Tanenhaus. 1994. Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In Charles Clifton, Lyn Frazier & Keith Rayner (eds.), *Perspectives in sentence processing*, 155–179. Hillsdale, NJ: Lawrence Erlbaum.
- Van Berkum, Jos J. A., Danielle van den Brink, Cathelijne M. J. Y. Tesink, Miriam Kos & Peter Hagoort. 2008. The neural integration of speaker and message. *Journal of Cognitive Neuroscience* 20(4). 580–591. DOI: [10.1162/jocn.2008.20054](https://doi.org/10.1162/jocn.2008.20054).
- van Berkum, Jos J. A., Peter Hagoort & Colin M. Brown. 1999. Semantic integration in sentences and discourse: Evidence from the n400. *Journal of Cognitive Neuroscience* 11(6). 657–671. DOI: [10.1162/089892999563724](https://doi.org/10.1162/089892999563724).

- Wagner, Laura, Maia Geene-Havas & Rebecca Gillespie. 2010. Development in children's comprehension of linguistic register. *Child Development* 81(6). 1678–1686.
- Wagner, Laura, Mariana Vega-Mendoza & Suzanne van Horn. 2014. Social and linguistic cues facilitate children's register comprehension. *First Language* 34(4). 299–314. DOI: [10.1177/0142723714544413](https://doi.org/10.1177/0142723714544413).
- Wegera, Klaus-Peter. 2007. *Die Entstehung der neuhochdeutschen Schriftsprache*. Frankfurt/Main: Peter Lang.
- Wells, John C. 1982. *Accents of English: An introduction*. Cambridge: Cambridge UP. 300 pp.
- Wieling, Martijn & John Nerbonne. 2015. Advances in dialectometry. *Annual Review of Linguistics* 1(1). 243–264. DOI: [10.1146/annurev-linguist-030514-124930](https://doi.org/10.1146/annurev-linguist-030514-124930).
- Wiese, Heike. 2020. Language situations: A method for capturing variation within speakers' repertoires. In Yoshiyuki Asahi (ed.), *Methods in dialectology XVI*, 105–117. Frankfurt/Main: Peter Lang.
- Wolfram, Walt & Natalie Schilling-Estes. 2015. *American English: Dialects and variation*. New York, Oxford: Blackwell. 456 pp.
- Wu, L. L. & Lawrence W. Barsalou. 2009. Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychologica* 132(2). 173–189.
- Wulff, Stefanie. 2010. Marrying cognitive-linguistic theory and corpus-based methods: On the compositionality of English V NP-idioms. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches* (Cognitive Linguistics Research 46), 223–238. Berlin, New York: de Gruyter Mouton.
- Yang, Charles. 2000. Internal and external forces in language change. *Language Variation and Change* 12. 231–250.
- Yang, Charles. 2002. *Knowledge and learning in natural language*. Oxford: Oxford UP.
- Yang, Charles. 2010. Three factors in language variation. *Lingua* 120(5). 1160–1177.
- Yatsushiro, Kazuko, Uli Sauerland & Artemis Alexiadou. 2017. The unmarkedness of plural: Crosslinguistic data. In Maria LaMendola & Jennifer Scott (eds.), *BU-CLD 41: proc. of the 41st annual Boston University Conference on Language Development*, 753–765.
- Zeldes, Amir. 2012. *Productivity in argument selection. From morphology to syntax* (Trends in Linguistics). Berlin: de Gruyter Mouton.

Appendix

Project	Title	Research area	Project leader (institute)
Area A: Register and grammar			
A01	<i>Metaphor and metonymy as register phenomena</i>	Semantics, pragmatics	Markus Egg (IAA)
A02	<i>Speaker's choices in a creole context: Bislama and Morisien</i>	Creole linguistics, syntax, morphology	Manfred Krifka (ZAS)
A03	<i>"Expressive" dislocation and register in Czech vs. Russian</i>	Syntax, corpus linguistics	Roland Meyer, Luka Szucsich (IS)
A04	<i>Situated syntax: Exploring and modeling syntactic register variation in German</i>	Bayesian modeling, constraint-based syntax	Stefan Müller (IdSL)
A05	<i>Modeling meaning-driven register variation</i>	Formal & experimental semantics, game-theoretic pragmatics	Uli Sauerland, Stephanie Solt (ZAS)
A06	<i>Disentangling cross-linguistic and language-specific aspects of register</i>	Syntactic variation	Elisabeth Verhoeven (IdSL), Aria Adli (RS)
Area B: Register and change			
B01	<i>Register and the development of periphrasis in the history of English</i>	Historical morphosyntax	Artemis Alexiadou (IAA), Thomas McFadden (ZAS)
B02	<i>Emergence and change of registers: The case of Lithuanian and Latvian</i>	Baltic linguistics, syntax, lexicon	Anna Helene Feulner, Wolfgang Hock (IdSL)
B03	<i>Register knowledge in Ancient Egypt</i>	Ancient Egyptian language, diachronic grammar	Frank Kammerzell, Silvia Kutscher (IA)
B04	<i>Register emergence and register change in Germanic</i>	Historical morphosyntax, historical lexical semantics	Muriel Norde (NI), Karin Donhauser (IdSL), Lars Erik Zeige (IdSL)
B05	<i>Register choice as a key to early stages of language change</i>	Historical syntax, language variation and change	Richard Waltireit, Anne Wolfsgruber (IfR)
Area C: Register and cognition			
C01	<i>Register through the prism of mono- and multilingual acquisition</i>	Language acquisition	Natalia Gagarina (ZAS)

C02	<i>Variation in situated interaction</i>	Sociophonetics, sociolinguistics, multimodality, language attitudes	Stefanie Jannedy (ZAS)
C03	<i>Situation-register congruence meets morphosyntax and verb-argument violations: Real-time and post-sentence comprehension</i>	Psycholinguistics, neurolinguistics, morphosyntax	Pia Knoeferle, Katja Maquate (IdSL)
C04	<i>Register knowledge in advanced learner language</i>	Second language acquisition, corpus linguistics	Anke Lüdeling (IdSL)
C05	<i>Development of specialized knowledge in linguistics and register flexibility in early adulthood</i>	Language acquisition, languages for specific purposes	Beate Lütke (IdSL)
C06	<i>Non-native addressee register</i>	Phonetics, phonology, morphology, syntax, corpus linguistics	Christine Mooshammer, Anke Lüdeling (IdSL)
Central projects			
MGK	<i>Integrated Graduate School</i>		Uli Sauerland (ZAS), Richard WALTEREIT (IfR)
INF	<i>Data management and statistical analysis</i>	Corpus linguistics, statistics, modeling, service	Malte Dreyer (CMS), Thomas Krause (IdSL), Anke Lüdeling (IdSL)
Z	<i>Central services</i>		Anke Lüdeling (IdSL), Artemis Alexiadou (IAA)

Table 2: Overview project groups and projects at the time of the proposal.⁷

⁷The project list has changed since the proposal. For a current list of projects, see <https://sfb1412.hu-berlin.de/>