Form-meaning mismatches

DFG Research Training Group
Georg-August-Universität Göttingen

RESEARCH PROGRAM
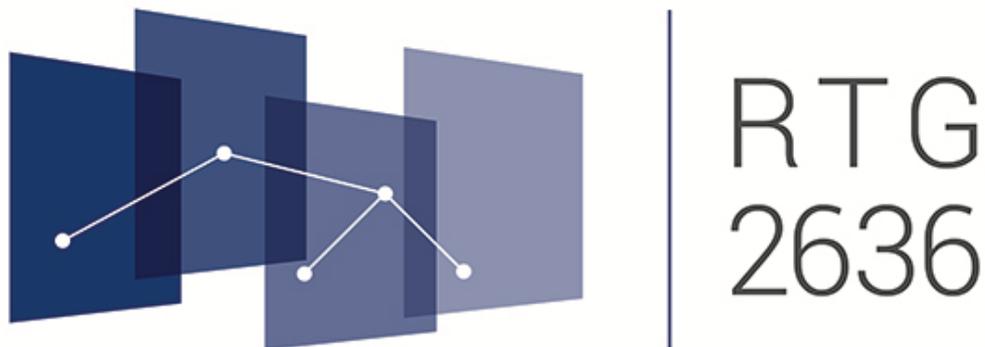
Duration: 01.09.2021 – 30.04.2026

1 Summary in German and English

German summary: Die Annahme, dass jeder Ausdruck in einem Satz zu dessen Bedeutung beiträgt und jeder Bedeutungsbeitrag einem Ausdruck im Satz entspricht, scheint intuitiv richtig zu sein. Bei genauerem Hinsehen sind Mismatches zwischen Form und Bedeutung in natürlichen Sprachen allerdings allgegenwärtig. Ihr genaueres Verständnis wird uns viel über das Wesen und die Funktionsweise menschlicher Sprache(n) verraten und die theoretische Sprachwissenschaft bereichern. Außerhalb der Linguistik im engeren Sinne sind Erkenntnisse über Mismatches zwischen Form und Bedeutung überall dort relevant, wo es um die exakte Entschlüsselung von Texten geht, in der maschinellen Sprachverarbeitung, im Fremdspracherwerb u.v.a.m.

Ein Mismatch zwischen Form und Bedeutung liegt zum Beispiel vor, wenn ein und derselbe Bedeutungsbestandteil mehrmals ausgedrückt wird. So wird etwa in dem Satz *Die schwarzen Hunde sitzen auf dem Dach* die Mehrzahl gleich viermal kodiert. Das Gegenteil ist der Fall, wenn ein Bedeutungsbestandteil keine formale Entsprechung hat. Beispielsweise bedeutet *Mit Maria habe ich gestern telefoniert* als Antwortauf die Frage *Mit wem hast Du gestern telefoniert?*, dass der*die Angesprochene *ausschließlich* mit Maria telefoniert hat. Darüber hinaus gibt es auch Fälle, in denen ein Ausdruck keinen Bedeutungsbeitrag leistet, und wiederum solche, wo ein Ausdruck mehrere Bedeutungskomponenten beisteuert, so dass sich in erster Annäherung vier Typen unterscheiden lassen.

Auch wenn derartige Mismatches zwischen Form und Bedeutung in einzelnen Bereichen immer wieder diskutiert wurden, erwarten wir durch eine umfassende Betrachtung der relevanten Phänomene in verschiedenen Sprachen und auf verschiedenen Ebenen (Morphosyntax, Semantik, Pragmatik) neue Erkenntnisse. Nur auf einer breiten empirisch und theoretisch motivierten Basis lassen sich Fragen beantworten wie: Lassen sich Generalisierungen treffen, die über Einzelphänomene hinweg angewendet werden können? Welche Typen von Mismatches zwischen Form und Bedeutung kommen in natürlichen Sprachen vor? Warum verfügen natürliche Sprachen überhaupt über derartige Mismatches? Wie entstehen sie in der zeitlichen Entwicklung von Sprachen? Welche Auswirkungen hat die Existenz von Mismatches zwischen Form und Bedeutung in der Sprachverarbeitung und im Spracherwerb? Die Beantwortung derartiger Fragen hat potenziell tiefgreifende Konsequenzen für unser Verständnis des Zusammenspiels von Grammatik und Bedeutung.

Im Rahmen des hier beantragten Graduiertenkollegs (GK) soll eine solche breite Basis an Einzeluntersuchungen entstehen, wobei Fragen wie die oben genannten im Dialog zwischen und mit den Doktorand*innen im Blick behalten werden, um sich in bis zu drei Kohorten an eine Theorie von Mismatches zwischen Form und Bedeutung anzunähern. In den entstehenden Dissertationsschriften werden verschiedene Arten von Mismatches zwischen Form und Bedeutung untersucht, wobei zunächst die o.g. vier Haupttypen zugrunde gelegt werden, die die umfangreiche Untersuchungsdomäne grob strukturieren, aber im Laufe der Arbeiten zu prüfen und ggf. zu modifizieren sind. Viele der zu untersuchenden Phänomene wurden bislang wenig bis gar nicht behandelt. Das GK beschäftigt sich dazu auch mit unterschiedlichen Sprachen/-familien. Die spezifischen Fragestellungen der Projekte decken alle relevanten Teilbereiche der Linguistik ab: Grammatiktheorie (Morphologie, Syntax, Semantik, Pragmatik), sprachliche Variation inkl. Gebärdensprachen, Sprachwandel, Sprachverarbeitung und Spracherwerb. Jede*r Doktorand*in beschäftigt sich mit einem empirischen Problem, hinter dem jeweils eine konkrete theoretische Frage steht. Dieser Ansatz gewährleistet eine breite Ausbildung sowohl in der empirischen Datenerhebung als auch in der theoretischen Analyse, die die Absolvent*innen angemessen auf die Anforderungen des akademischen und nicht-akademischen Arbeitsmarktes vorbereitet.

English summary: The assumption that every expression in a sentence contributes to its meaning and every contribution to meaning corresponds to an expression in the sentence seems intuitively correct. On closer inspection, however, form-meaning mismatches are abundantly present in natural languages. A better understanding of them will teach us a lot about the nature and the functioning of human language(s) and will enrich theoretical linguistics. Outside core linguistics such findings about form-meaning mismatches are relevant for the precise decoding of texts, for natural language processing, and for second language acquisition among many others.

On the one hand, a mismatch between form and meaning arises when one and the same meaning component is expressed multiply. In the sentence *Die schwarzen Hunde sitzen auf dem Dach* ('The black dogs are sitting on the roof'), for instance, plurality is expressed three times. On the other hand, the opposite occurs when a meaning component does not have a formal counterpart at all. For example, the sentence *Mit Maria habe ich gestern telefoniert* ('With Maria') as an answer to the question *Mit wem hast du gestern telefoniert?* ('Who did you call?') means that the addressee *only* called Maria. In addition, there are also cases where several expressions make a joint contribution to meaning, and again cases where one expression makes several meaning contributions. That is, four groups of form-meaning mismatches can be distinguished.

Even though such discrepancies between form and meaning have been discussed in particular areas more than once, we expect novel insights from a comprehensive investigation of relevant phenomena in various languages and on different linguistic levels (morphosyntax, semantics, pragmatics). Only on the basis of a broad, empirically and theoretically motivated foundation can the following questions be addressed: Can we make generalizations that are applicable across phenomena? Which types of form-meaning mismatches occur in natural languages? Why do natural languages exhibit such mismatches to begin with? How do they develop over time? Which implications does the existence of form-meaning mismatches have for language processing and acquisition? Answering such questions has potentially deep consequences for our understanding of the interplay of grammar and meaning.

In the course of the present Research Training Group (RTG), such a broad foundation of individual investigations will be developed. Questions such as the ones mentioned above will be considered in dialogue between and with the PhD students, allowing us to approach a theory of form-meaning mismatches in up to three cohorts. The dissertations to be developed will investigate various kinds of form-meaning mismatches. The four main types mentioned above will form the starting point for structuring this extensive domain of investigation. Over time these types will be evaluated and, where applicable, modified. Many of the phenomena have so far received little or no attention. The RTG also deals with different languages and language families in this sense. The specific questions of the projects cover all core areas of linguistics: grammatical theory (morphology, syntax, semantics, pragmatics), linguistic variation including sign languages, language change, language processing and language acquisition. Each PhD student deals with an empirical problem, which is based on a concrete theoretical question. This approach ensures a broad training in both empirical data collection and theoretical analysis, which will prepare the graduates adequately for the requirements of the academic and non-academic labor market.

2   Profile of the Research Training Group

The central aim of this RTG is to provide a large group of PhD students (36 in total, funded through the entire RTG) with an excellent training in linguistic research by jointly investigating one of the most pressing questions in linguistics, namely to what extent morphosyntactic structures are mapped onto semantic/pragmatic representations and vice versa. This research program will be carried out in an environment with a very strong, dynamic and well-organized linguistics community and will be guided by a qualification program that combines individual training by multiple supervisors with a wide range of courses, lectures, summer schools and international short- and long-term visitors. This enables PhD students to obtain all necessary linguistic and other academic skills for becoming high-end researchers, ready to pursue a career in empirical and theoretical linguistics or related fields.

It has been a standard assumption in linguistic theory, based on Frege's principle of compositionality (cf. Frege 1884, 1892), that the meaning of a sentence is composed on the basis of the meaning of its parts and the way these parts are structured. However, it is not really clear that there exists a transparent, bi-directional mapping between (morphosyntactic) form and (semantic/pragmatic) meaning. Many phenomena challenge the existence of such a mapping. Such challenges either exist in the form of morphosyntactic elements that do not seem to provide a semantic/pragmatic contribution, or in the form of meaning parts that lack a morphosyntactic realization. Hence, either the mapping between form and meaning (in its broadest sense) in natural language is less transparent than one might think, or

morphosyntactic and/or semantic/pragmatic structures may be richer than they appear. Addressing the question as to what this mapping exactly amounts to is at the heart of contemporary linguistic theory, and requires an in-depth investigation of a wide range of case studies involving potential challenges for this mapping as well as an overarching perspective that integrates the various findings of individual research projects. All PhD students in this RTG will carry out such a case study, and together with the participating researchers and other RTG-members, such as the current cohort of postdocs and other junior faculty members, they will implement the outcomes of their studies in an overarching theoretical perspective that this RTG aims to develop. PhD students in this RTG will thus receive excellent individual linguistic research experience working on their own projects while collectively developing a much-needed contribution to the core of linguistic theory.

In order to best instruct the PhD students and to maximally prepare them for future academic careers, they will all receive a training program consisting of a combination of broad supervision and an extended qualification program. As for the supervision program, every PhD student will have at least two main supervisors (both participating researchers) with different backgrounds (e.g., one supervisor working on syntax, one on semantics). Apart from this, other researchers, including existing postdocs and other junior faculty, are involved in the supervision process. The two main supervisors and one or more additional supervisors form the individual thesis committee of each PhD student that monitors the student's progress and meets every semester to discuss the progress and to signal potential problems. Next to the individual supervision, every doctoral candidate obtains a customized qualification program consisting of a selection of various courses (parts of four modules: research & methods, teaching skills, key qualifications & career planning, and organizational skills). In addition, PhD students will co-organize an RTG-colloquium series, host RTG-special guests and participate in an RTG-summer school. Finally, every PhD student is strongly encouraged to spend a period of three to four months at a foreign top-ranked institute for his/her specialization. This way the RTG will deliver a new generation of strong, independent and internationally oriented junior scientists. In addition, the existing postdocs and other junior faculty will gain experience in assembling a group of PhD students to jointly arrive at answers to overarching questions, enlarge their methodological and theoretical expertise, and further develop teaching and scientific leadership skills.

In recent years, Göttingen has established a strong linguistic community that pursues a cooperative teaching and research agenda. To join forces and to foster collaboration, linguistic researchers, spread over different philologies, the general linguistics department as well as other institutions like the University Center *Text Structures*, have established the platform *Linguistics in Göttingen* (*LinG*). The research focus of *LinG* lies on the relation between (grammatical) structure and meaning, both from a cross-linguistic perspective and a perspective of language change, as well as language processing and acquisition. In addition, *LinG* upholds a rich, long-standing tradition of investigating a large spectrum of spoken languages and hosts an ambitious center for the study of sign languages. All professors associated with *LinG* participate in the RTG, which means that PhD students can maximally profit from the present infrastructure in linguistics, interact with a large community of scientists, and make use of all the present facilities.

3   Research

3.1   Background

One of the major functions of language is its mapping between form and meaning. Ideally, such a mapping should be transparent in the sense that the meaning of a sentence can be determined on the basis of the form of the sentence and vice versa, where form is taken to be the full overt structure of the sentence, and meaning the full, final interpretation of that sentence. If that were not the case, it would be unclear how speakers would be able to assign meaning to new sentences, and thus use language productively.

Against this background, it is mysterious why languages, at the same time, exhibit various digressions from a complete transparency between form and meaning: several elements do not appear to bring in a meaning contribution of their own, as is the case in agreement and concord phenomena. And various meaning aspects do not always have a

reflection in the uttered form of the sentence, e.g. when the meaning of a sentence appears to be more informative than what is literally said. Of course, such elements can be analyzed in such a way that a full form-meaning mapping can be maintained, for instance by arguing that particular morphemes may lack either a phonological realization or semantic content. However, that does not address the question of why natural language would exploit such elements in the first place.

Whereas various individual form-meaning mismatches have been investigated in a fair amount of detail, the overarching question of why the form-meaning mapping in natural language prima facie is not 100% transparent has never been satisfactorily addressed. Moreover, it has hardly ever formed a guiding research question in linguistic theory. This RTG aims at filling this lacuna. In this doctoral training program, we will investigate a wide range of form-meaning mismatches – some of them novel, others having been studied before – with the objective of understanding to what extent apparent intransparencies between form and meaning may emerge. What determines the kinds of attested form-meaning mismatches? Do all form-meaning mismatches result from the same underlying principle, or are form-meaning mismatches more heterogeneous in nature? And why does natural language allow such mismatches in the first place?

Such questions have strong repercussions for linguistic theory, especially when it concerns the interface between morphosyntax and semantics/pragmatics. But addressing such questions may also help us better understand language variation, language change, language processing and language acquisition: To what extent do languages vary with respect to form-meaning mismatches? How do form-meaning mismatches emerge over time? Can different types of form-meaning mismatches be attested among spoken and signed languages? How are form-meaning mismatches processed in the brain? And how can a child acquire language if the mapping between form and meaning cannot be relied upon as being transparent?

At least four different types of form-meaning mismatches can be distinguished: (i) cases where "words" or particular instances of morphology do not seem to convey any additional meaning; (ii) cases where multiple apparently semantically active elements only make a single meaning contribution; (iii) cases where a meaning contribution is not reflected in the morphosyntax at all; and (iv) cases where a single morphosyntactic element gives rise to multiple semantic effects. Below, we give examples of each of these four categories.

An example of type (i) occurs, for instance, in agreement morphology. Consider the examples from Arabic in (1) (after Aoun et al. 1994). In Arabic, a subject preceding the verb triggers full co-varying agreement on this verb. The verb in (1a) shows 3rd person feminine plural morphology, given that the subject is 3rd person plural feminine. That this verbal agreement is semantically redundant is shown in (1b), where the subject follows the verb. Now, the verb only agrees with the subject in gender and person, but no longer in number. The verb is no longer marked for plural, despite the subject being a plural. This shows that presence of number morphology in (1) lacks a semantic interpretation.

(1)   a.   L-banaat-u   dˤarab-na/*- at   l-ʔawlaad-a
               the-girl.PL-NOM   hit.PAST-3F.PL/*-3F.SG the-boy.PL-ACC
               'The girls hit the boys.'
        b.   dˤarab-at/*-na   ʔal-banaat-u   l-ʔawlaad-a
               hit-PAST-3F.SG/*-3F.PL the-girl.PL-NOM   the-boy.PL-ACC
               'The girls hit the boys.'

Instances of type (ii), where multiple apparently semantically active elements only make a single meaning contribution, can be found in so-called Negative Concord constructions reported in a range of languages including Italian, Czech, Bavarian, West Flemish, Spanish, among others (cf. Ladusaw 1992; Haegeman & Zanuttini 1996; Giannakidou 2000; Zeijlstra 2004 and others). We illustrate with Bavarian, where negative concord has been argued to be attested in certain varieties (Bayer 1990). As shown in the example in (2), even though in Bavarian, the words *nicht* 'not' and *keine* 'no' can each introduce a semantic negation of their own, their joint meaning contribution in (2) is only one single negation.

(2)   Ich   bin froh, dass ich keine Rede   nicht   halden brauch

<pre>
I   am glad that   I no   talk   not   give   need
</pre>
'I am glad that I don't have to give a talk.'   (Bayer 1990:15)

Cases of type (iii), where a particular meaning contribution is not reflected in the morphosyntax, can be attested when the meaning of an utterance appears to be semantically enriched. For instance, the answer in (3) does not only suggest that the speaker talked to Mary at the party, but that s/he talked to *only* her, even though that does not follow from the literal meaning.

(3)   Q: Who did you talk to at the party?   A: To Mary (I talked).

Finally, it can be the case that a particular element brings in a meaning effect that only indirectly adds to the meaning of the sentence (type (iv)). For instance, McCready (2010) presents (politically incorrect) examples like (4) below, where it is entailed that Peter is German, but where it is also inferred that the speaker has a low opinion of Germans (see also Potts 2003, 2007; Gutzmann 2012, 2017 for discussion on expressive contents).

(4)   Peter is a kraut.

3.2   Scientific motivation of the RTG

We believe that now is the proper time to address in its full breadth the question why the mapping of sound to meaning and vice versa does not appear to happen in a fully transparent way, both for theoretical and empirical reasons. Let us explain why.

3.2.1 Theoretical motivation: It has been a long-standing assumption in linguistic theory that the meaning of a sentence follows from the meanings of its parts and the way in which these parts are structurally combined with each other, often (incorrectly) attributed to Frege (1892) (cf. Partee 1975, 1984; Janssen 1997). In the same vein, it has also often been assumed (though sometimes not as strongly as the reverse) that every morphosyntactically active element brings in a meaning contribution. If these assumptions were indeed correct, it would mean that there exists a fully transparent mapping between (morphosyntactic) form and (semantic/pragmatic) meaning. As the examples above already showed, however, this is not the case.

Hence, either the mapping between form and meaning (in its broadest sense) in natural language is less transparent than initially thought, or we must assume that morphosyntactic and/or semantic/pragmatic structures may be richer than they appear, and postulate that there are overt meaningless elements or covert elements that bear semantic content. In addition, the notion of sentence meaning would be in need of enrichment in comparison to original truth-conditional semantics. However, it is not really clear to what extent transparency between form and meaning can be maintained if it comes at the expense of such enrichment of form or meaning. Such a postulation requires independent evidence. Different linguistic theories take different stands, though, with regard to the number and range of such elements or operators that can be assumed to be present – and consensus is often lacking even within single frameworks.

Addressing the question as to what exactly form-meaning mappings amount to, and how (in)transparent they can be, lies at the heart of contemporary linguistic theory and requires an in-depth investigation of a large variety of phenomena involving potential challenges for this mapping, as well as an overarching perspective that integrates the various findings of the research projects. Concretely, this amounts to evaluating a wide range of form-meaning mismatches with respect to the question as to whether the mismatch is only apparent (and fully transparent on some more abstract level), or whether it is real and requires us to think more deeply about the interaction between morphology, syntax, semantics, and pragmatics.

— Form-meaning mismatches at the morphosyntax-semantics/pragmatics interface: Studies concerning transparency between linguistic form and meaning traditionally focus on the interface between morphosyntax and semantics. While in certain parts of grammar, the mapping between morphosyntactic form and meaning seems to be transparent, other parts of grammar pose a challenge to such a transparent mapping. For instance, grammatical gender and noun-class systems, or cases of apparently semantically redundant agreement morphology, as exemplified in (1), do not contribute to the meaning in a straightforward and perspicuous way; phenomena such as verbal agreement or grammatical gender seem rather orthogonal to matters of meaning. In addition, we find complex idioms, fixed locutions, and discourse particles, such as modal or presentational particles, whose meaning is very hard to

capture, but whose usage is abundant, as well as scope mismatches, and meaning enrichments, which all appear to refute decomposition and invite analyses in terms of non-transparent form-meaning mappings. To make things even harder, many parts of language(s) do not carry their semantic nature on their sleeve. Consider, for instance, the well-studied phenomenon of negation. Even though, semantically, negation is part of the core compositional system, mapping propositions to propositions with reverse truth conditions, data from natural languages challenge the idea that negation is a simple transparent operation: Negative Concord (where multiple negatively marked elements yield one semantic negation, as illustrated in (2)), scope-related mismatches between the syntax of negation and its semantics, and many other facts suggest that the mapping between negative form and negative meaning is much less transparent than seems to be the case at first sight. Yet, phenomena like negation have received theoretical analyses that are fully in line with a transparent mapping between form and meaning, such as the analysis of negation in terms of overt (negative) agreement markers triggering abstract (negative) operators (Ladusaw 1992; Haegeman 1995; Haegeman & Zanuttini 1996; Brown 1999; Weiß 2002; Zeijlstra 2004, 2008b; Haegeman & Lohndal 2010). Similar kinds of analyses have been proposed to account for apparent compositionality problems in various other domains, such as tense, modality or (*wh-*) questions.

Such examples show that the idea of transparency can often be maintained if the analysis makes use of elements that only signal the presence of a (potentially covert) controller, as assumed in the case of negation or in the case of verbal agreement. Apart from elements that lack a particular meaning, elements that have a clear meaning but lack any overt form have also been assumed, such as items indicating speech acts, covert operators like exhaustifiers, generic operators, and other abstract elements (such as null pronouns like PRO). Moreover, many semantic theories rest on operations such as type shift or quantifier movement, tacit quantification, and other morphosyntactically triggered operations, thus adding to the stock of elements whose sheer contribution is to let elements make their meaning contribution at a locus that is distinct from their surface position. All these elements serve to retain the idea of a fully transparent mapping between form and meaning. For instance, if (subject-) agreement markers are purely functional elements that lack any semantic content, no compositionality problems arise in combinations of subjects and finite verbs. Likewise, a clear mapping between the number of syntactic arguments and the number of thematic roles can always be maintained by postulating unpronounced arguments such as PRO.

However, if transparency between form and meaning can only be maintained if it comes at the expense of postulating overt meaningless elements or covert meaningful operators, the question arises whether such postulated theoretical ingredients can be independently motivated. Although most scholars agree that semantically vacuous elements and covert operators should only be postulated once their effects are visible, criteria for visibility differ and can only be formulated in particular theoretical frameworks. For instance, generative (minimalist) syntax often assumes a rich inventory of formal interpretable and uninterpretable features that trigger syntactic operations (Chomsky 1995, 2000, 2001; Harley & Ritter 2002; Adger 2003). Corresponding semantic theories assume that the meanings of utterances are compositionally derived from meanings of words and morphemes, possibly enriched by certain covert (type-shift and other) operators and by universal pragmatic processes (cf. Heim & Kratzer 1998). But other theories take different stands. For instance, construction grammar takes a holistic stance: It associates meanings *directly* with structures (see, among others, Goldberg 1995, 2006; Tomasello 2003; Fischer & Stefanowitsch 2007). Moreover, even within the same theoretical framework, there is often no consensus. As Meisel (2010: 975–976) remarks with respect to such postulated elements (functional categories, FCs, in his terms) in syntax, "[t]he particular importance of FCs for numerous areas of linguistic research stands in stark contrast to the fact that there exists no broadly accepted consensus on their theoretical status". Hudson (2000:8) even concludes that "[a]ny notion as important as Functional Category should be subjected to the most rigorous scrutiny, but this seems not to have happened […]. Instead it has been accepted more or less without question". Clearly, a systemic investigation of the extent to which form-meaning mismatches can be taken to represent underlying transparent mappings between form and meaning is strongly called for. Only if there is independent, (ideally) theory-neutral,

evidence for the postulation of (invisible) elements that can restore the transparency between form and meaning, can the mapping between form and meaning be said to be fully transparent.

— **Form-meaning mismatches at the semantics-pragmatics interface**: Many alleged form-meaning mismatches are generally understood as mismatches between morphosyntax and semantics. However, meaning, understood as the information conveyed by a particular form, is not always entirely determined by the morphosyntactic form. It may also incorporate aspects that are traditionally dealt within pragmatics. Traditionally, conversational implicatures were assumed to be computed post-grammatically, i.e., after the sentence form had been created (with the literal meaning), thus constituting a different kind of form-meaning mismatch. But the clear distinction between semantic and pragmatic aspects of meaning that this picture presupposes has been challenged frequently, and is currently under special scrutiny.

First, the information conveyed by sentences often depends on the linguistic and extra-linguistic context (as in the case of anaphora and demonstratives), but context dependency does not stop there: Tense, indexicals, aspect and mood, adjectives, adverbs, and quantifiers may encode aspects of meaning which may all be further enriched by context. Second, a large class of expressions directly incorporates meaning contributions that apparently only aim at marking peculiarities of a context or impose constraints on the context in terms of common-ground information (e.g., presupposition triggers), the question under discussion, the aims of conversation, or even possible further discourse developments (cf. Zeevat 2003; Potts 2003, 2007; Beaver & Clark 2008; Gutzmann 2012, 2017; Krifka 2014). Third, context has been argued to play a role in the disambiguation of meaning by way of selecting stronger or more likely readings by making different assumptions about the structure of sentences. This has even led scholars to assume that syntax and pragmatics are intertwined more deeply. Chierchia (2004, 2006, 2013) proposes that exhaustification operators and implicature computation are part of the grammar and must be located in syntax, just like Speas & Tenny (2003) presume speech act centers to be encoded syntactically. Moreover, Fox (2007) and others show that many other classical pragmatic problems, such as free choice effects, appear to be related to the syntactic integration of corresponding operators.

This shows that the traditional divide between syntax-semantics and pragmatics will have to be re-considered, as it has been many times over in the history of linguistics and philosophy of language. To date, it is an open issue where new principles should replace the old division of labour. Studying form-meaning mismatches and, in particular, the notion of 'meaning strengthening', i.e., the assignment of additional "invisible" meaning to an utterance, has thus become highly important in determining the boundaries between semantics and pragmatics, and will provide novel insights for the exploration of the relation between language system and language use.

3.2.2   Empirical motivation: The question of why the mapping between form and meaning is not always as transparent as one would prima facie expect can be approached from different empirical angles. Understanding form-meaning mismatches requires investigations in various linguistic disciplines, including linguistic typology and language variation (including sign languages), diachronic linguistics, language processing and language acquisition - disciplines that all belong to the core specializations of the Göttingen linguistics community. Full understanding of (non-)transparent form-meaning mapping is only gained by thoroughly investigating the structural encodings of form-meaning mismatches, the range of cross-linguistic variation they are subject to (within and across modalities), their historical emergence, the way children acquire them, and how they are processed in the brain. Exploring form-meaning mismatches empirically will, in turn, inform theories of language variation, change, acquisition and processing.

In order to carry out such a widespread empirical investigation, various data sources are required. Recently, a lot of new corpora and other methodological tools have been made available: For spoken language, various databases, such as CORAL-ROM (Crest & Moneglia 2005) for Romance languages or the Corpus of Contemporary American English (COCA, https://corpus.byu.edu/coca/) for English, as well as typological repositories that provide large collections of grammatical features or annotated text collections in diverse languages, such as The World Atlas of Linguistic Structures (WALS, http://wals.info/), Terraling

(http://test.terraling.com/), and TypeCraft: The Interlinear Glossed Text Repository (https://typecraft.org/tc2wiki/Main_Page); for language change, the series of the Penn-notation-parsed corpora of Old, Middle, and Early Modern English, as well as the Corpus of Early Modern English Correspondence, the Referenzkorpus Altdeutsch (Donhauser et al. 2018), and the Referenzkorpus Mittelhochdeutsch (1050–1350) (Klein et al. 2016); for language processing and acquisition, the Child Language Data Exchange System (CHILDES database, MacWhinney 2000), alongside sophisticated statistical methods to model them; and for sign language, a whole series of experimental techniques and corpora, such as the Hamburg Deutschen Gebärdensprache (DGS) Corpus. Apart from work with corpora, the PhD students will use a wide range of other methods as well (for instance, eye tracking, association tasks, sentence reproduction tasks, acceptability judgement studies, etc.).

Below, we spell out the way in which empirical investigations into language variation (incl. sign languages), language change, and language processing and language acquisition can inform us about the nature of form-meaning mismatches and vice versa.

— **Form-meaning mismatches in language variation (across modalities)**: The relation between form and meaning has been of key importance in understanding language diversity since the early years of linguistic thought. Adam Smith, in his *Dissertation on the Origin of Languages* (Smith 1811 [1767]), speculates that rigid word order compensates for the loss of case and agreement in modern European languages. The reasoning of a functional complementarity between formal devices is pervasive in accounts of language diversity, e.g., overt *wh*-movement and question particles in clause typing (Cheng 1991); null-subjects and agreement (Rizzi 1986); head-marking and dependent-marking (Nichols 1986). All these accounts are motivated by the idea that there is some higher-order function to be fulfilled and that languages vary in selecting from an array of formal devices for this purpose. The wealth of empirical data from the world's languages that has been made available in the last decades challenges these generalizations and calls for refinements of our hypotheses about the relation between form and meaning. What we finally learn from the typological data is that natural languages can afford a vast amount of vagueness and redundancy, departing from a one-to-one mapping between form and meaning (see, e.g., Comrie 1991 on disambiguation in a language without case and argument positions; Levinson 2010 on speech-act processing in a language without question particles, question intonation, and interrogative word order). The broad spectrum of languages that are being explored at Göttingen University serves as a powerful basis for the empirical domain of research. Beside numerous Indo-European languages, our range further extends to non-Indo-European varieties such as Bantu or Mayan languages, and also covers older language varieties. Form-meaning mismatches attested for spoken languages can also be found in sign languages (cf., e.g., Davidson 2013, 2014). However, the visual-gestural modality of sign languages may have an impact on form-meaning mismatches (Meir 2002; Aronoff et al. 2005). Three modality-specific properties seem to be particularly relevant to the study of form-meaning mismatches across the two modalities. First, sign languages make use of a three-dimensional signing space to express grammatical, semantic, and pragmatic features (cf. Pfau & Steinbach 2016). This means that the signing space is used to express quite different features, some of which seem to only have a grammatical function while others directly realize (parts of) propositional content or pragmatic meaning, such as presuppositions or implicatures. A systematic investigation of spatial form-meaning correlations is still pending. Second, unlike spoken languages, sign languages can use different articulators simultaneously to express various grammatical, semantic, and pragmatic features. Each articulator can either be used to express a different feature, or various articulators are used to express only one feature. In the latter case, multiple apparently semantically active markers together only make a single meaning contribution. An especially interesting case is formed by so-called nonmanual markers (Pfau & Quer 2010), which are typically underspecified multifunctional markers. Third, sign language and (non)manual gestures use the same modality. Therefore, gestures can be integrated at various levels into the grammatical system of sign languages. Consequently, the interface between the two systems is permeable. This has important consequences for the interaction of form and meaning. On the one hand, gestural elements can enter the grammatical system and develop into lexical, grammatical, or pragmatic markers. On the other hand, linguistic and gestural elements can interact in a modality-specific way to express

complex propositions and realize specific speech acts (Schlenker 2018; Davidson 2015). Studies within the RTG are expected to provide substantial contributions to linguistic typology by refining(cross-modal) typologies that consider the interaction between grammatical features in spoken and sign languages and employ state-of-the-art methods in empirical research (i.e., cross-linguistic experiments and statistical modelling).

— **Form-meaning mismatches in language change**: Form-meaning mismatches are often the result of diachronic change, and their evolution can be observed in language history. For example, agreement morphology is frequently eliminated in diachrony (e.g., by phonetic erosion) and can remain absent (as in Afrikaans or Japanese), but it is also often re-established through grammaticalization of pronouns into clitics and finally into inflectional elements. Intermediate stages of such processes can be closely studied, for instance, in Northern Italian, Occitan, and French subject clitic systems, which show different degrees of co-occurrence of older agreement systems with incipient newer systems; present-day dialectal variation reflects different diachronic stages (cf. Polo 2007). The diachronic decline and rise of agreement shows that agreement is part of the nature of language. At the same time, it is unclear why agreement systems may remain stable for longer periods of time in some languages, whereas they disappear more or less quickly in others. Cases in which a meaning contribution is not overtly reflected in speech can result from processes such as coalescence, semantic bleaching, and phonetic erosion. Here, it must be examined how the semantic system reacts to such changes, e.g., by means of covert operators (if they are to be assumed). Other cases can be argued to be triggered by changes in the morphosyntactic system. For instance, the rise of certain types of so-called null-subjects can often be regarded as being connected to the rise of strong agreement morphology. However, Chinese and other East Asian languages do not have verb agreement but still allow for null subjects (see Huang 1984; Neeleman & Szendrői 2007). This may bear upon the morphosyntax-pragmatics interface, as it has been argued that, in such languages, the omitted argument can be retrieved from the discourse (see D'Alessandro 2014 for literature and discussion). The inverse case also offers interesting research perspectives. For example, it has remained an open question what triggered the loss of null-subjects in Old French. These and other issues have partially been addressed in studies on grammaticalization, as well as on what has recently been dubbed pragmaticalization (cf. Diewald 2011), for instance, in the case of emergent pragmatic markers. Cognitive and functional frameworks have been focussing on mechanisms of semantic shift, explanations based on a tension between linguistic economy and the necessity of communication, as well as universal grammaticalization pathways. But a systematic study of the diachrony of form-meaning mismatches has never been undertaken. An important task will be to establish a typology of the diachronic mechanisms that lead to such mismatches. Diachrony serves as a testing ground for synchronic formalizations, as these must be compatible with the diachronic mechanisms identified.

— **Form-meaning mismatches in language processing and language acquisition**: Form-meaning mismatches have a massive influence on the way in which language is processed (both in production and in perception) and how it is acquired. Given that both processing and acquisition take place in real time and are sensitive to cognitive resources, this influence can take on two forms. On the one hand, mismatches of the type exemplified in (1) create redundancies on the level of morphosyntactic encoding without any visible semantic effect. It is reasonable to assume that language processing and acquisition exploit such redundancies: Processing becomes more robust or automatic by relying on such multiple encodings of structural features; this robustness, or automaticity, saves resources for other processes (cf. Levy 2008 and others). In a similar vein, acquisition processes may use such redundancies and the excess in cognitive resources to bootstrap into the complex system of form-meaning mappings and, ultimately, to straddle developmental hurdles (see Han et al. 2007 for an example). On the other hand, there are types of mismatches which can pose problems for processing and acquisition: in processing, mismatches like non-local dependencies tax the cognitive resources that both language production and comprehension tap into. And in acquisition, acquiring a form which mismatches the corresponding function may put severe strain on the learner's cognitive resources (e.g., by having to resort to context, or other cues). By systematically identifying the resource-relative benefits and charges of mismatch phenomena, and tracing their effects in processing and

acquisition by means of controlled experimentation, an empirical bedrock can be supplied for the overall theme of the RTG, and thereby add to our understanding of how the overall cognitive system reacts to the challenges that form-meaning mismatches impose. In addition, the RTG will investigate to what extent form-meaning correspondences (or the lack thereof) drive language acquisition: do children start out with 1:1 form-meaning correspondences and later on acquire potential deviations from that, or do children pursue different acquisitional pathways.

3.3   Research questions

Given the argumentation above, the following sets of theoretical and empirical research questions will be addressed in our RTG *Form-meaning mismatches*.

Theoretical questions
- Why does natural language employ elements that do not appear to contribute to the meaning of an utterance?
- Why can utterances convey more information than follows from the audible/visible elements they contain?
- What kind of different meaning contributions can be introduced by linguistic elements?
- What constrains the range of variation that can be attested with respect to form-meaning mismatches?

Empirical questions
- What is the range of variation that can be attested cross-linguistically with respect to the (in)transparency of the mapping between form and meaning?
- To what extent do spoken and sign languages differ with respect to the transparency of the mapping between form and meaning?
- How do form-meaning mismatches emerge and disappear over time?
- To what extent do form-meaning mismatches have reflexes in processing and computation? And what can studies on language acquisition tell us about the way in which (first-) language learners deal with form-meaning mismatches?

3.4   Methods and architecture of the RTG

The main objective of this RTG is to address these research questions through a number of case studies carried out by PhD students, and to collectively approach the theoretical foundations of the mapping between form and meaning and its consequences. This entails that this RTG will investigate a wide array of topics without spelling out an articulated framework first. The reason for this is that the RTG is ultimately intended for the education of PhD students. What we want to establish is a common ground between students of various research backgrounds and paradigms (varying from theoretical linguistics to psycholinguistics, and from diachronic linguistics to typology) who will develop new theoretical and empirical tools in a core issue of linguistic research. In order to achieve this goal, the participating researchers set up an agenda of challenging topics, but we explicitly do not restrict the type of approach in terms of a linguistic theory first. Such a framework is to be developed in the course of the RTG by the participating researchers in close collaboration with their associates and the students.

This research aim of the RTG will therefore require case studies, each of which considers a particular kind of form-meaning mismatch on the basis of a particular set of data from a particular empirical domain, and which will be carried out by our PhD students. Each of these case studies should have a theoretical component (involving the interfaces between morphosyntax, semantics, and pragmatics), and a focus on one of the four distinguished types of form-meaning mismatches as well as an empirical basis, along the lines discussed above: involving language variation (across modalities), language change, language processing and language acquisition. These case studies will be complemented by surveys in different frameworks. Theoretically, the participating researchers in the RTG and their associates will then connect the outcomes of the various subprojects. This enables PhD students to embed their research results in a bigger theoretical and empirical perspective. The development of a theoretical perspective should be the guiding force behind the selection of future follow-up projects.

The proposed RTG consists, in total, of three cohorts of PhD students in two funding periods of 4.5 years each. In each cohort, a balanced distribution of projects on the interfaces

between morphosyntax, semantics, and pragmatics will be realized. Moreover, in order to address the questions listed above, each cohort will reflect a balanced distribution of case studies in the empirical disciplines that this RTG covers. All projects within each cohort are well-connected, so that results from one project will naturally feed other projects. This guarantees a maximal integration between members of the RTG while retaining balance and broadness. The first cohort provides the empirical and theoretical building blocks that will form the basis of the second cohort, and the same should hold for the connection between the second and the third cohorts. The major advantage of this architecture is that the students and the participating researchers together build a project at large, which yields both an empirical basis and a theoretical framework, able to account for the landscape of form-meaning mismatches and their different causes. In Tables 2 and 3 below, we list several case studies that could be covered in PhD theses in the first cohort. They span a broad range of form-meaning mismatches, known from the literature and previous research. These case studies involve both investigations into the interface between morphosyntax, and semantic and pragmatics (block I), as well as the relation between the latter two (block II). They are also spread over the RTG's three empirical research domains.

Block I: Form-meaning mismatches in the morphosyntax-semantics/pragmatics interface

	Empirical domain		
Type of form-meaning mismatch	Language variation (across modalities)	Language change	Language processing and lang. acquisition
Type (i): 1:0 form-meaning mismatches	1. Verbal agreement in and across sign languages	2. From discourse to syntax	3. Word-form and word-meaning mismatches in language acquisition
Type (ii): many:1 form-meaning mismatches	4. Modal concord cross-linguistically	5. Diachrony of split NP/DP constructions	6. Multiple *wh*-questions: Processing of ex-situ/in-situ variation

Table 2: Envisaged projects in the first cohort (Block I)

Block II: Form-meaning mismatches in the semantics-pragmatics interface

	Empirical domain		
Type of form-meaning mismatch	Language variation (across modalities)	Language change	Language processing and lang. acquisition
Type (iii): 0:1 form-meaning mismatches	7. Imperatives and imperative speech acts across modalities	8. Null objects	9. Processing exhaustive inferences
Type (iv): 1:many form-meaning mismatches	10. Particles with question and disjunction uses	11. Presentational particles	12. Identifying context-dependent meaning components of figurative expressions

Table 3: Envisaged projects in the first cohort (Block II)

The set of projects outlined above reflects all dimensions that constitute the RTG. Within the domain of language variation (across modalities), the range of variation that verbal agreement exhibits across sign languages is investigated, as well as how that compares with agreement in spoken languages. Also Modal Concord is studied, the phenomenon where multiple modal elements yield a single modal reading, and how languages may differ with respect to that. A third case study looks at how imperative meanings follow (in)directly from the imperative form: How and to what extent is the illocutionary meaning of imperative expressions encoded syntactically? Finally, in this empirical domain the RTG aims at investigating the exact meaning contribution of question particles in relation to their position in the sentence.

Within the domain of language change, the focus is on those phenomena where diachronic analyses may shed more light on the nature of form-meaning mismatches. For instance, particular fixed word order systems appear to have evolved from more flexible ones, where different word orders coincided with different pragmatic / discourse effects. Similarly, the question of why certain languages have determiners/DPs and not only NPs can be investigated

by looking at what caused DPs to emerge in the first place. The same holds for the diachronic developments with respect to the usage and distribution of covert elements, such as null objects. Finally, the RTG will look at the emergence of a particular subclass of particles, namely presentational particles, to see what original syntactic and semantic properties of such elements (and their subcomponents) were, and how such properties may be affected over time.

Within the domain of language processing and language acquisition, the RTG will investigate how the cognitive system deals with cases where sentence form and sentence meaning do not align. Starting with infants, it is investigated to what extent they think that words that overlap in form overlap in meaning and what triggers children to semantically differentiate elements that formally look very similar. Another project tries to see what the effects of prosody, morphological marking, and *wh*-movement are in processing *wh*-questions. Given that languages differ with respect to how and when exhaustification effects are triggered, we will also look into how exhaustification is processed in the first place, and whether processing exhaustification works similarly across languages. Finally, the RTG will investigate global processing of figurative expressions, which primarily function to reflect speakers' attitudes.

All projects are well-connected with each other. For e.g., in order to investigate the processing of exhaustification, cross-linguistic differences pertaining to pragmatic strengthening are also investigated, as is the case in the project on imperatives. Similarly, understanding the grammaticalization of certain discourse-structural processes in diachronic terms also has important consequences for the understanding of verbal agreement. The borders between language variation (across modalities), language change, and language processing and acquisition are not always rigid, so that, even though various projects are characterized under one domain, together they form a natural continuum of case-studies that jointly address the research questions of this RTG. While they do not cover the investigated field of research exhaustively, together they capture a well-adjusted spectrum of problems and phenomena concerning form-meaning mismatches. Figure 1 below shows how exactly all these projects interact.
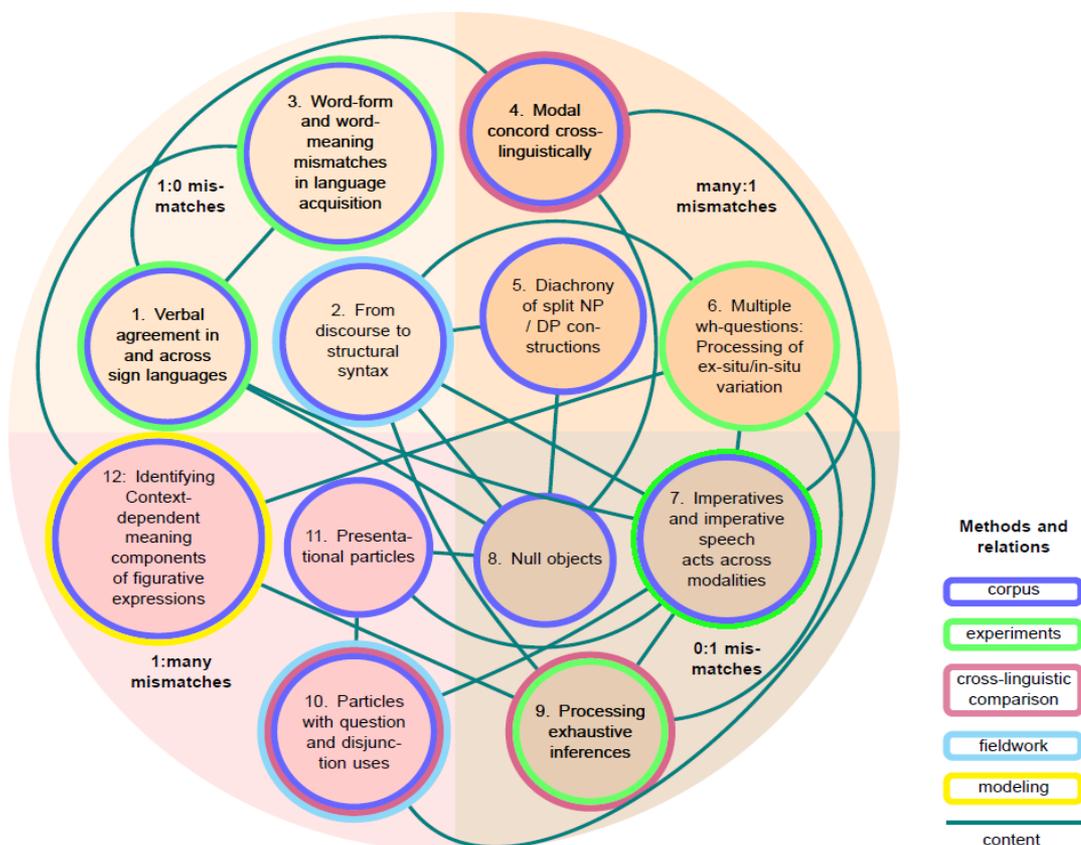

Figure 1: Connections across projects

At the same time, this does not preclude that, as a result of actual applications or of

developments in the course of the RTG, adjustments can be made regarding the envisaged case studies. In fact, we emphasize that we need to remain flexible both in altering the order of the projects within each project area and in introducing new research topics that also fit the central theme of the RTG, if there are good reasons to do so. Modifications concerning the proposed projects are possible both in the context of the recruitment process for the first cohort of PhD students (for instance, if highly qualified PhD students apply with a research project of their own) as well as in the course of the RTG (where, for instance, obtained results or other developments call for other follow-up studies). At the same time, we will make sure that the case studies are properly balanced in every cohort along the dimensions sketched.

Below, these envisaged projects are introduced in more detail. We first introduce the relevant form-meaning mismatches and motivate why they are in need of investigation. After this, a draft for a PhD project within this area is presented. In this draft, the research questions as well as first hypotheses and possible research methods are addressed. Also, the intended supervisors and their collaborators are mentioned, as well as how the project connects with other projects. Finally, potential follow-up projects are listed.

Project 1: Verbal agreement in and across sign languages

Supervisors: Markus Steinbach, Hedde Zeijlstra

The form-meaning mismatch: Agreement is a typical example of a 1:0 form-meaning mismatch. Like many spoken languages, most sign languages have a complex system of verbal agreement inflection (Lillo-Martin & Meier 2011; Mathur & Rathmann 2012; Pfau et al. 2018). However, unlike spoken languages, sign languages do not realize agreement sequentially, by adding agreement affixes, but rather simultaneously, by means of a transparent spatial modulation of the verb's phonological features 'movement' and/or 'orientation'. In sign languages, verbs can be modulated such that their beginning/end points (and/or the orientation of the hands) coincides with referential loci in signing space to which the corresponding arguments of the verb are (anaphorically) linked. An interesting modality-specific feature of sign language agreement is that the grammatical system has a gestural origin, which is still (at least partly) transparent and semantically active (Meir 2002; Aronoff et al. 2005; Steinbach 2011). Moreover, only a subclass of verbs, i.e., so-called agreement verbs, can express subject and object agreement. By contrast, so-called plain verbs cannot express agreement overtly, as their phonological movement features are lexically specified. Over time, some plain verbs have lost this lexical specification and become agreement verbs. But even more importantly, many sign languages have developed specific agreement auxiliaries that are used with plain verbs to express agreement (Steinbach & Pfau 2007). The grammaticalization of agreement auxiliaries provides evidence that sign-language agreement develops into a morphosyntactic system that becomes more grammatical in the course of time (Pfau & Steinbach 2013; Pfau et al. 2018). At the same time, recent studies have shown that agreement in sign languages is subject to similar restrictions as differential object marking in spoken languages (Bross 2018; Börstell 2019), i.e. agreement is not completely arbitrary but constrained by (discourse) semantic properties of the corresponding argument(s). In this context, it has also been observed that in sign languages, overt agreement inflection is optional at least to a certain extent (Fenlon et al. 2018; Oomen 2019). This specific combination of morphosyntactic and semantic properties raises the question to what extent verbal agreement in sign languages constitutes a 1:0 form-meaning mismatch.

Motivation: In order to understand the interaction between morphology, syntax and semantics in verbal inflection and the development of inflectional systems, agreement in sign languages is an important test case. As mentioned, agreement in sign languages has a gestural basis and some modality-specific properties (e.g. simultaneous realization of agreement and modality-specific lexical restrictions). At the same time, there is evidence that sign languages develop more grammaticalized agreement systems that share many properties with fully grammaticalized agreement systems found in spoken languages. An in-depth empirical investigation of the properties of agreement in different sign languages will thus tell us how inflectional systems emerge in the visual-gestural modality at the interface between gesture and

sign and whether theories designed for the analysis of spoken language agreement can also (or at least partly) account for the specific properties of these systems.

Research questions: The question immediately arises as to how the specific properties of agreement in sign languages can be explained and whether they can be derived from more general properties of the visual-gestural modality. One important aspect in this context is whether the analysis of sign language agreement can be implemented in recent morphosyntactic theories developed for spoken languages or whether a completely different account built on the modality-specific properties of sign languages has to be pursued. Especially the gestural origin of agreement in sign languages poses a challenge for classical morphosyntactic theories of agreement. In addition, the investigation of agreement in sign languages will shed new light on the question to what extent agreement in sign language is an arbitrary grammatical system and to what extent it still has a transparent semantic basis. It is thus directly relevant to the question of how syntactic categories have emerged from underlying semantic concepts and how specific morphosyntactic restrictions on these categories have developed.

Hypotheses and method: The baseline of this project is the hypothesis that the agreement system in DGS is hybrid in combining transparent semantic with arbitrary morphosyntactic properties. Focusing on DGS, the PhD student will first carry out a corpus study, using data from the new Hamburg DGS corpus, which is one of the largest sign language corpora available for linguistic research in the world. In the first step, the PhD student will build on two pilot corpus studies on specific aspects of agreement marking in DGS (Macht 2016; Oomen 2019) and focus on (i) semantic and contextual constraints on the inflection of agreement verbs, and (ii) the combination of agreement auxiliaries with plain verbs. Both aspects directly address three questions: (a) To what extent is agreement in DGS optional? (b) Is agreement in sign languages subject to semantic constraints, e.g. can it be compared to differential object marking in spoken languages? (c) Do younger signers show different patterns than older signers, i.e. can we observe a process of grammaticalization? As a second step, the PhD student will conduct experimental studies on the gestural origin of sign language agreement and the grammatical status of agreement. The gestural origin will be investigated with a semantic association task with hearing non-signers, hearing L2 learners of DGS and deaf signers (Strickland et al. 2015). For the grammatical status of agreement, the PhD student will conduct a sentence reproduction task with deaf signers manipulating overt agreement inflection and the distribution of the agreement auxiliary. In the final step, the results of the empirical studies will be compared to competing analysis of verbal agreement in sign languages (Holler & Steinbach 2018; Pfau et al. 2018). The project is thus innovative in two respects: the empirical studies close an important gap in the analysis of agreement in sign languages and the theoretical implementation at the interface between gesture and sign take up recent theoretical developments in sign language linguistics (Goldin-Meadow & Brentari 2017).

Connection to other research projects: As projects 2 and 3, this project investigates type (i) form-meaning mismatches, and as projects 4, 7 and 10, it focuses on language-variation. Contentwise, this project is connected to projects 3, 7, and 8: Project 3 also investigates potential mismatches between word form and word meaning. Whereas, the focus of project 3 is on language acquisition, project 1, by contrast, investigates various aspects of language change and contextual restrictions on language use. Project 7, like project 1, deals with the impact of modality and the relevance of gestures for linguistic structures. Finally, the findings of project 8 will be important for project 1 since sign language agreement licenses null objects. Methodologically, it connects with virtually every other project, as it involves both corpus studies and experiments.

Possible follow-up studies:
1. Classifier agreement in sign languages
2. The emergence of agreement systems across modalities
3. Gestural agreement in spoken languages

Project 2: From discourse to syntax

Supervisors: Marco Coniglio, Stavros Skopeteas

The form-meaning mismatch: How do 1:0 form-meaning mismatches evolve in language change? The syntacticization of processes that were once used to express particular discourse functions covers an array of phenomena for investigating this question. For e.g., Old Germanic languages (as most Ancient Indo-European/IE languages) had an OV clause structure and employed V-fronting under particular conditions, related either to information structure or to the illocutionary force of the utterance. Examples are verb-first in thetic sentences and imperatives, or verb placement in front of a narrow focus domain, see (5) for Old High German (for German, cf. Lenerz 1984; Önnerfors 1997; Axel 2007, 2009; Schallert 2007; Petrova 2009, 2011; Hinterhölzl & Petrova 2009, 2011; Coniglio 2012; Schlachter 2012; for other Germanic and IE languages, cf. Hock 1986; Van Kemenade 1987; Kroch 1989; Matras & Sasse 1995; Kiparsky 1996; Pintzuk 1999; Fuß & Trips 2002; Taylor & Pintzuk 2012; Walkden 2014):

(5)   thaz   in   mir   habet   sibba
       that   in   me   you.have peace
       'so that you have peace in me'
       Lat. ut in me pacem habeatis (T 290,8, adapted from Petrova 2009: 266)

In Germanic languages, including German, verb placement eventually became rigid, sometimes conditioned by clause type, as evidenced by V2 placement in main declarative clauses (by the end of the Old High German period) and V-final placement in subordinate clauses. Hence, the discourse-related functions of the OV/VO alternation in a previous diachronic stage were lost after reanalysis (approximately in the Early New High German period).

Motivation: Within an environment that develops the theoretical and methodological tools for the investigation of form-function mappings, we expect that this PhD project will provide a precise account of this type of language change and its functional consequences. Given the collaborative supervision, we expect this PhD project to combine the knowledge from diachronic syntactic studies with the knowledge about currently spoken languages that display similar cases of OV/VO flexibility. For instance, several Southern Caucasian languages (e.g., Georgian and Eastern Armenian) display an alternation between OV and VO that is at least superficially similar to the flexible linearization of VPs in Old Germanic (Comrie 1985; Harris 2000; Skopeteas & Fanselow 2010). One of the best studied instances of word order flexibility in these languages is the role of focus, as shown in (6) for Georgian: (6a) illustrates the canonical V-final order, while (6b) illustrates the fact that a preverbal focus (obligatorily) attracts the finite verb.

(6)   a.   p'it'er-ma   es   p'roblema   čkara   gada-č'r-a
            Peter-ERG this   problem(NOM)   quickly   PR-(IO.3)solve-AOR.S.3.SG
            'Peter solved this problem quickly.'
       b.   p'it'er-ma   gada-č'r-a   es   p'roblema   čkara
            Peter-ERG PR-(IO.3)solve-AOR.S.3.SG this   problem(NOM)   quickly
            'Peter solved this problem quickly.' (Skopeteas & Fanselow 2010: 1378)

In modern languages, word order generalizations have been supported by types of evidence that are not available for languages only accessible through corpora, most importantly by speakers' intuitions, including, among others, prosodic judgments (see e.g. Skopeteas et al. 2009, 2018). What can the knowledge established for modern languages with flexible order of the V within VPs contribute to the understanding of word order variability in Old Germanic (and vice versa)?

Research questions: The first research question concerns the *motivation of change*: What are the grammatical conditions that are behind the reanalysis from a discourse-related alternation as a syntactic rule? The second one concerns the *consequences of change*: What are the consequences of the loss of particular pragmatic functions? Is there a residual of discourse functions that remains after reanalysis (Gärtner 2002; Gärtner & Michaelis 2010)? Does the loss of such functions mean that particular functions of word order (e.g., creating an optimal linearization of discourse-related domains) are compensated by other properties of expressions (such as prosodic means, scrambling, etc.) in a later stage?

Hypotheses and method: The PhD student is expected to offer a cross-linguistic study of the functional properties of OV/VO alternation in Old Germanic and a currently spoken language

having flexible order of the V, such as Georgian. The major aim of this study is to figure out whether the word order flexibility in Old Germanic can be understood by general principles of the linearization of flexible V-projections that hold cross-linguistically. Can the prosodic facts that have been reported for Georgian (e.g., the prosodic integration of preverbal foci to the verb or the presence of a p(rosodic)-phrase edge at the right side of the V in Skopeteas et al. 2009, 2018) also hold for Old Germanic, maybe manifested through phonological processes (e.g., apocope for avoiding hiatus in the absence of a p-boundary)? Do the effects of V-fronting in anchoring (or not anchoring) the assertion to the speaker, or in marking the illocutionary force of a sentence (cf. Meinunger 2004, 2006; Antomo & Steinbach 2010 etc.) apply to Georgian? In order to tackle these questions, the PhD student will use corpora of Old Germanic (for example, Referenzkorpus Altdeutsch (ReA) for Old High German, Donhauser et al. 2018) and the spoken Georgian corpus of the General Linguistics Department (Asatiani et al. 2019) and will examine qualitative hypotheses through linguistic fieldwork. Based on this background, he/she will develop and test hypotheses about the sources and consequences of OV-to-VO change in diachronic syntax. More generally, the question will be answered how the syntacticization of discourse- or information-structural processes affect compositionality and how such 1:0 form-meaning mismatches can be interpreted.

Connection to other research projects: As projects 1 and 3, this project investigates type (i) form-meaning mismatches. The empirical focus is language change, shared with projects 5, 8 and 11. Contentwise, this project is connected to project 7 – both projects investigating syntacticization phenomena beyond the propositional level – and will collaborate with project 9 in the investigation of information structure. The project shares methods and ideas of corpus research with virtually all other projects. It also shares methods of fieldwork with project 10.

Possible follow-up studies:
1. Syntacticization of left/right dislocations
2. From topic markers to case assigners
3. Clause typing: movement vs. grammaticalization of sentence type markers

Project 3: Word-form and word-meaning mismatches in language acquisition

Supervisors: Nivedita Mani, Thomas Weskott

The form-meaning mismatch: Why is a chair called a *chair* in English, *Stuhl* in German and /kʊr.siː/ (kursi) in Hindi? It is typically assumed that the mapping between the form of a word, i.e., its component phonemes, and its meaning is arbitrary and conventionalised (Saussure 1916; Hockett 1960), such that the meaning of a word is not inferable from its form (a 1:0 form-meaning mismatch). While exceptions exist, e.g., onomatopoeia, sound symbolism or iconicity in form-meaning mappings (see Dingemanse et al. 2015 for a review), these have largely been considered rare exceptions. However, this viewpoint is being revaluated, given recent studies showing that the probability that two words that are related in meaning also overlap in form is greater than that expected by chance (e.g., Monaghan et al. 2014; Dautriche et al. 2016). Such findings highlight a potential systematicity of form-meaning mappings, which may impact the acquisition and processing of words.

Motivation: There is renewed interest in the arbitrariness of form-meaning mappings following suggestions that there may be a greater correlation between the semantic and phonological distance between words than previously assumed (e.g., Dautriche et al. 2016). Regularities in form-meaning mappings not only question the pervasiveness of lexical form-meaning mismatches, but have also been shown to boost word learning, category formation and lexical retrieval (e.g., Monaghan et al. 2011; Imai & Kita 2014). Against the background that such regularities may particularly help young learners establish links between phonological and semantic levels of representations (Dautriche et al. 2016), there is a need for further evaluation of the extent to which such correlations may be found in developing lexicons. At the same time, there is considerable disparity in the findings of studies examining the extent to which overlap at the level of form and meaning may boost early word learning and word processing. On the one hand, word-form familiarity or word-meaning familiarity boosts acquisition of similar-sounding or similar-meaning words (Newman et al. 2008; Altvater-Mackensen & Mani 2013; Borovsky et al.

2016). On the other hand, children find it difficult to simultaneously acquire and process words that overlap on multiple dimensions (Dautriche et al. 2015). In other words, the influence of form-meaning *mismatches* on *in-the-moment* language learning and processing remains unclear. The disparity in these findings highlights the need for further research into the extent to which the presence or absence of form-meaning overlap between words may influence language acquisition and processing.

Research questions: Against this background, the proposed project will examine two related research questions. First, the project will address the extent to which the correlation between semantic and phonological distance in adult lexicons is similarly found in developing lexicons (c.f. Monaghan et al. 2014). Second, the project will examine the consequences of potential form-meaning overlap on lexical acquisition and processing, i.e., the extent to which overlap at the level of both form and meaning boosts or hinders word learning and word processing.

Hypotheses and method: The first research question will examine the correlation between semantic and phonological distance in developing lexicons. The PhD student will examine this using data from Wordbank (www.wordbank.stanford.edu), which contains data pertaining to the words known to children learning 29 different languages at different ages across development. At the same time, he/she will also examine this question with regards to the input to children using corpora of parent-child interactions (childes.talkbank.org). Thus, he/she will examine the correlation between semantic and phonological distance in early language use with regards to both the input that children receive and their vocabulary knowledge. As in Dautriche et al. (2016), the raw Levenshtein-distance between two words will provide a measure of phonological distance, while data from German associative norms and corpus analyses of the contexts in which words occur could be used to infer semantic distance. The Pearson correlation between semantic and phonological distance in words across a set of chosen languages would provide a measure of the extent of form-meaning mismatches in the early lexicon. We hypothesize that, in keeping with adult lexicons, there is a correlation between the semantic and phonological distance in words, highlighting early systematicity in form-meaning mappings.

The second research question will examine the extent to which the presence of form-meaning overlap between two words influences children's acquisition and processing of these words. Here, using classic paradigms to examine word learning in young children (Eiteljörge et al. 2019; Taxitari et al. 2020), the PhD project will train children on novel word-object associations that either overlap at the form-meaning level or do not, and examine their learning and recall of overlapping and non-overlapping associations. By modulating the extent to which children are familiar with the overlapping word-object associations, this project will examine whether potential learnability constraints incurred by form-meaning match are reduced by prior familiarity with the information to be learned. Thus, potential studies could examine the extent to which form-meaning overlap differentially impacts acquisition and processing of novel and familiar word-object associations.

Connection to other research projects: This is the third project next to projects 1 and 2, which investigates type (i) form-meaning mismatches. The empirical focus of the project, language acquisition and processing, is shared with projects 6, 9 and 12, which means that the project will benefit from cross-talk with these projects examining the processing of form-meaning mismatches. In addition, there will be considerable exchanges with Project 1 with overlapping interests in the opacity of agreement expressions. Methodologically, this project involves corpus studies and experimentation – virtually all other projects involve corpus studies, and the methodology of experimentation is also employed by projects 1, 6, 7 and 9.

Possible follow-up studies:
1. The development of arbitrariness and systematicity in form-meaning mapping
2. The influence of arbitrariness and systematicity in retention of newly learned information
3. Differences in the learning of arbitrary and systematic form-meaning mappings

Project 4: Modal concord cross-linguistically

Supervisors: Caroline Sporleder, Hedde Zeijlstra

The form-meaning mismatch: Expressions consisting of multiple modal elements normally yield a cumulative reading. Sentences like *Maybe Mary has to leave* and *John should be allowed to read this file* contain two modal elements each (*maybe*/*has to* and *should*/*allowed*, respectively). However, if two modal elements are of the same modal type (epistemic/deontic) and have similar quantificational force (universal/existential), the most salient reading is not a cumulative one but a concord reading, where the semantics seems to contain only one modal operator, as shown below:

(7)   You may possibly have read my book.
         'The speaker considers it possible that the hearer read his/her book.'
         [?]'The speaker considers it possible that it is possible that the hearer read his/her book.'
(8)   Power carts must mandatorily be used on cart paths.
         'It is obligatory that power carts are used on cart paths.'
         [?]'It is obligatory that it is obligatory that power carts are used on cart paths.'

                                                   (Geurts & Huitink 2006: 15)

This phenomenon has first been observed by Halliday (1970) and Lyons (1977) and has been first analyzed by Geurts & Huitink (2006), who have argued that the readings in (7) and (8) are not entailed by the iterative readings and thus exist in their own right. They refer to this phenomenon as *modal concord (MC)*. Further analyses have been formulated by Huitink (2006), Zeijlstra (2008a), Anand & Brasoveanu (2010) and Grosz (2010).

Motivation: Given the principle of compositionality, two elements that in isolation can introduce some particular meaning should give rise to a cumulative reading when the two elements appear together in one sentence. However, often such cumulative readings do not arise. For instance, in cases of negative concord, two negative expressions yield only one negation, and in cases of sequence-of-tense, the presence of two tense morphemes often indicates the presence of only one semantic tense operator. Phenomena like negative concord and sequence-of-tense have been fairly widely investigated and various compositional analyses have been provided (cf. Zeijlstra 2004, 2008b; Haegeman & Lohndahl 2010 and De Swart 2010 for recent approaches to negative concord). Other concord phenomena, however, have been less well investigated. By investigating MC, a phenomenon that has hitherto never been satisfactorily explained in detail, it can be explored whether these readings only appear to be a violation of the principle of compositionality, or whether it shows that the modes of interpretation of morphosyntactic structures are more complex than the principle of compositionality suggests.

Research questions: The question immediately arises as to why the two modal elements do not both give rise to a transparent semantic modal interpretation. For Geurts & Huitink (2006), this is due to some semantic absorption process, where the two modals melt together. For Zeijlstra (2008a) and Anand & Brasoveanu (2012), by contrast, the phenomenon arises due to the fact that multiple ingredients of modality may be introduced by different elements. Concretely, Anand & Brasoveanu (2012) suggest that modal force is introduced by modal auxiliaries and modal flavour by modal adverbs. For auxiliaries this is quite clear: *must* always has a universal modal reading, but is ambiguous between an epistemic and deontic interpretation, even though other modals, such as *might*, do not reflect this ambiguity. For modal adverbs, this is harder to show: is *obligatorily* a quantificationally underspecified deontic adverb, or does it also have universal force? Only if the former is the case, do MC readings not violate compositionality: the universal deontic modal reading in (8) then transparently comes from the deontic adverb and the universal auxiliary. If modal adverbs induce modal force as well, compositionality appears to be violated and other explanations for the phenomenon are called for.

Hypotheses and method: Focusing on English, the PhD student will first carry out a corpus study, using data from corpora such as the Corpus of Contemporary American English (COCA), where the semantic contexts of a large series (approx. 25) of modal adverbs are analyzed. Based on these data, he/she establishes with what other modal elements such adverbs may yield a modal concord reading, focusing on the modal type and/or force. The crucial question is whether the distribution of MC is restricted to auxiliaries of one particular modal force or whether the distribution is more flexible. It will thus provide us with empirical clues on whether modal adverbs do introduce modal force of their own or not. This opens up a way to see whether MC

is the result of modal auxiliaries introducing modal force and modal adverbs introducing modal flavour, or whether the phenomenon calls for a different type of explanation. In the second part of this project, the PhD student will conduct a cross-linguistic comparison of the availability and restrictions on MC in a sample of approximately 15 languages from different families and regions. From the literature, it is known that such differences exist (cf. Matthewson et al. 2007 for examples from modals in St'at'imcets), but it is not investigated or understood yet how this range of variation is constrained and how different types of MC can be predicted from an underlying syntactic-semantic theory.

Connection to other research projects: This project connects with projects 5 and 6 in investigating type (ii) form-meaning mismatches, and with projects 1, 7 and 10 as regards its empirical focus, i.e. language variation. Contentwise, this project is connected to projects 1, given the potential correspondences between modal concord and agreement: with project 7, given that imperatives may have a modal component (cf. Kaufmann 2012), with project 8, given that modal concord may involve covert operators, and with project 9, given that modality can be subject to strengthening effects as well, and may cross-linguistically vary with respect to that. Regarding methodology, the project connects with virtually all other projects since it involves corpus studies. It also overlaps with projects 9 and 10 for cross-linguistic survey methods.

Possible follow-up studies:
1. Clause-internal vs. clause-external Modal Concord
2. Modal Concord diachronically
3. Comparing Modal Concord, aspectual concord and temporal concord (Sequence of Tense)

Project 5: Diachrony of split NP/DP constructions

Supervisors: Marco Coniglio, Götz Keydana

The form-meaning mismatch: Hyperbaton, especially discontinuity in the realm of complex noun phrases, is a major challenge in the general context of this RTG as "[t]he phrase is presumably reconstructed into a single continuous noun phrase at some level of semantic interpretation" (Devine & Stephens 2000: 247). Consider the following examples:

(9) pollḕn   gàr   pánu katélipen   ho patḕr   autõi   ousían
     much.ACC PTCL very leave.AOR.3SG   the father.NOM he.DAT property. ACC
     'For his father left him a very large property.' (Ancient Greek: Aeschin.1.42, Devine & Stephens 2000: 261)

(10) táṃ   vo   vājānām   pátim   áhūmahi
      this.ACC you.DAT   prize.GEN.PL lord.ACC   call-upon.AOR.3PL
      'We have called upon the lord of prizes on your behalf.' (Vedic Sanskrit: Rigveda 8.24.18)

In (9), sentence-initial *pollḕn* and sentence-final *ousían* form a single DP. In (10), the DP *táṃ vājānāmpátim* is broken up by a Wackernagel clitic. In both cases, a listener processing these utterances has to analyse two or more constituents compositionally despite their discontinuity. Hyperbaton is thus a showcase example of a many:1 relation between (surface) form and meaning. While, in early Indo-European languages, this kind of construction is very frequent (cf. Lühr 2016), in later stages it typically becomes much more restricted (e.g. to a subset of quantified constructions). In this project, we will concentrate on hyperbata *stricto sensu* like (9).

Motivation: Given compositionality, a split between a head and the "dependent" DP/NP is unexpected. The computational load in processing hyperbata is necessarily higher than that of continuous constituents. We expect not only additional information- or discourse-structural functions connected to such constructions, but also more complex syntactic structures with properties licensing them (especially under the consideration that both the head and the dependent DP/NP may be displaced). From the diachronic point of view, the unidirectionality of the change in hyperbaton distribution is remarkable: In most Indo-European languages, hyperbaton is in no longer available (with the notable exception of Greek), but never totally ousted (cf. for example Giusti & Iovino 2010, 2014, 2016 on Latin and development in Italo-Romance). Re-emergence of hyperbaton is extremely rare (but see Mensching in press). The

development of restrictions raises the question of the mechanisms behind them: do such constructions really violate the principle of compositionality or do we have to assume information- and discourse-structural factors ruling discontinuity? How do these factors get lost and how are they related to other syntactic changes? Why do languages with massively restricted hyperbata end up with the patterns currently attested?

Research questions: In the research into discontinuous syntax, various analyses have been proposed (cf. Fanselow 1988; Van Riemsdijk 1989; Ott 2011). For instance, Bošković (2005) ties hyperbata to a global parameter that inter alia also determines the possibility of pro-drop. Thus, shifts in the distribution of subject pro are expected to go hand in hand with the emergence of restrictions on hyperbata. Furthermore, Roehrs (2007) interprets the two parts of the split construction as being base-generated separately, with one part containing "a null semantic anaphor that needs to be semantically identified on the basis of the split-off [...]" (Roehrs 2007: 335). Hence, for example, changes in the internal make-up of quantified expressions (e.g. the change from partitive to cardinal expressions described in Roehrs & Sapp 2016), may block the discontinuous use of the dependent NP/DPs. Focussing on Greek, Devine & Stephens (2000) argue that hyperbaton is licensed if the NP/DPs are part of complex predicates (or an information-structural unit) – i.e. if their heads are weak – or if they occupy a low position in the argument hierarchy.

Hypotheses and methods: Concentrating on quantified expressions in one or more Indo-European languages (for example Greek, Lithuanian, Latin, German, etc. – the number of languages depends on time depth and attestation), the PhD student will investigate the factors leading to the loss and re-emergence of hyperbata over time based on diachronic text corpora. On a broad empirical ground, he/she will test, among others, the following hypotheses: (i) Hyperbaton and subject pro-drop develop in parallel fashion; (ii) The development of Q(uantifier)-heads into modifiers leads to a loss of hyperbata; (iii) The weaker the head of a dependent NP/DP with respect to information structure and/or argument hierarchy, the more prone it is to hyperbaton (and the more resilient it is against continuity constraints); (iv) Similarly, the lower the dependent NP/DP is in the argument hierarchy, the higher the availability of hyperbata is at later stages. In addition to these specific hypotheses, the PhD student will explore the more general hypothesis that hyperbata originate in structures in which the head and the dependent are to a certain degree syntactically and semantically independent from each other and do not give rise to form-meaning mismatches (see Luraghi 2010). Their (later) reinterpretation as parts of one constituent that is split due to the information-structural properties either of the dependent or the head brings in a violation of compositionality. Thus, the ultimate loss of hyperbata in most languages could be seen as an attempt to recreate a perfect match between form and meaning. In addition (or alternatively), the PhD student will investigate whether the types of hyperbata still attested in some languages have been grammaticalized over time in a process that led to the loss of their information-structural load and to their syntacticization.

Connection to other research projects: Project 5, like projects 4 and 6 discusses type (ii) form-meaning mismatches. Empirically, it focuses on language change with projects 2, 8 and 11. With regard to content, this project is connected to project 2 because both projects investigate the syntacticization of information-structural processes, to project 8 on null-objects (since pro-drop and hyperbata are possibly driven by a single parameter), and to project 9 since focus-fronting could be connected to the hyperbaton phenomena investigated in this project. From a methodological perspective, the project involves corpus studies, which connects it with virtually all other projects.

Possible follow-up studies:
1. Split-NP vs. floating quantifiers in diachronic perspective
2. Diachrony of right dislocation/heavy NP-shift
3. Pied piping vs. stranding in relative constructions.

Project 6: Multiple *wh*-questions: Processing of ex-situ/in-situ variation

Supervisors: Uwe Junghanns, Thomas Weskott

The form-meaning mismatch: The strategies languages exploit to form constituent questions vary along many dimensions. One such dimension is *in-situ* vs. *ex-situ,* i.e. whether the *wh-* expression can remain in the argument position, or if it has to be located in a peripheral position. For example, Romanian allows only for ex-situ (see Comorovski 1996), and Mandarin Chinese only for in-situ *wh*-expressions (see Huang 1982). There is a fair number of languages which exhibit a considerable degree of freedom in the placement of *wh*-words. Take, for example, multiple *wh*-questions in a language like Czech, which allows *wh*-pronouns like *kdo* ('who') and *co* ('what') each to remain in situ, as well as to be moved to the left periphery, respectively (see Toman 1981; Rudin 1988; Błaszczak & Fischer 2001 and Meyer 2004 for the pertinent facts in Slavic languages). This will give us four forms of a constituent question: (a) *kdo.in-situ/co.in-situ*; (b) *kdo.ex-situ/co.in-situ*; (c) *kdo.in-situ/co.ex-situ*; and (d) *kdo.ex-situ/co.ex.situ*. At first blush, all four forms correspond to one meaning: for a sentence like *Kdočetl co?* ('Who read what?'), the meaning apparently common to all four forms is: for which *x* and which *y* is it the case that *x* read *y*?

Motivation: While from a theoretical point, one might ask whether languages should allow a ratio as uneconomic as four forms to just one meaning, a more fruitful question is wherein the exact differences in the semantics and/or pragmatics of the variants consist: the null hypothesis claiming that there are no meaning differences is hard to maintain; but what exactly are the factors along which the meaning of the forms vary—e.g., presupposition, or information structure more generally? Implicature? Register? Sociolinguistic factors? Under a processing perspective, the question why different forms of *wh*-marking are allowed becomes even more pressing: ex-situ *wh*-marking creates filler-gap dependencies (see Fodor 1978; Clifton & Frazier 1987), and thus taxes processing resources (attention, working memory), while in-situ *wh*-marking does not, or, at least, to a lesser degree. If the meaning of a constituent question is the same for both the in- and the ex-situ variant, why should speakers bother to produce a more complex form, and force it upon their hearers to deal with the comprehension problem this form creates? The project will investigate the variation of (morpho-)syntactic *wh*-marking in a number of relevant languages, taking the null hypothesis — that there is no co-variation between different forms and the meanings they express — as the starting point. While admittedly this null hypothesis runs danger of being a straw man, the alternative hypotheses to replace it will have to accommodate the complexity of both the syntactic and the semantic/pragmatic factors involved in question formation and interpretation.

Research questions: The overarching question is what the factors driving the variation in the placement of *wh*-phrases are. As a first step, these factors have to be identified both by recourse to existing literature on interpretational differences (from Bolinger 1978, to more recent work by Cable 2007, and Kotek 2015), as well as by controlled acceptability judgment experiments. For each language to be investigated (e.g., Czech or Polish), the following factors and possible confounds known from the literature will have to be checked and possibly controlled for: (i) superiority effects; (ii) availability of in-situ variant, plus its reading (echo/ contrastive/ rhetoric/ indefinite/pair-list); (iii) (weak) cross-over effects; (iv) sensitivity to D-linking (Pesetsky 1987); and possibly a number of other factors like animacy that have been identified in the literature as pertinent to multiple *wh*-question formation.

Hypotheses and method: The PhD student will conduct an assessment of the acceptability profiles of different variants of multiple *wh*-question forms; a possible starting point could be the four variants of the Czech sentence *Kdočetl co?* ('Who read what?'). By operationalizing the factors identified as governing the variation, these factors can then be tested in small-scale acceptability judgment studies (systematically testing the acceptability of a question, variant given a context/scenario). Once the influence of the factors is established in off-line measures and can be quantified, their contribution to processing difficulty, or its amelioration, can be investigated further by employing self-paced reading, and, where necessary, eye-tracking measures in order to learn more about the influence of these factors on on-line comprehension. The working hypothesis behind the experimental part of the project is that in multiple *wh*-questions, even forms that employ ex-situ marking and are thus dispreferred from a processing perspective will correspond to an identifiable pragmatic meaning in a restricted set of contexts. This correspondence between forms on the one hand, and context-meaning pairs on the other

hand, can be read off from the acceptability and processing cost profiles these forms exhibit. To circumvent possibly strong effects of superiority violations, object-*wh*-phrases in ditransitives can be employed. The availability of readings will be systematically controlled by post-tests (testing, e.g., for pair-list readings, etc.). By relating the empirical observations back to the theoretical question of why multiple *wh*-questions in certain languages exhibit a form-meaning mismatch, the project will make an important contribution to a better understanding of variation in the expression of interrogative meaning.

Connection to other research projects: Like projects 4 and 5, this project investigates type (ii) form-meaning mismatches, and like projects 3, 9 and 12, the empirical focus of the project is language processing and acquisition. With regard to content, this project is related to projects 2, 10 and 12. Project 2 is concerned with syntactization processes of what used to be functionally determined variation of word order (OV/VO). Semantic and/or discourse factors restricting placement is one issue linking the two projects. Another question is whether variability is bound to be given up in favor of a more rigid syntax. To what extent and how can loss of syntactic freedom suggesting loss of functions be compensated for? Project 10 addresses question and disjunction uses of particles. Particles in polar questions display an affinity for particular positions high up in clause structure. An overlap with placement of *wh*-items is evident. What properties then allow in-situ realization of *wh*-phrases? Among the issues connecting the two projects are the semantic and pragmatic determinants of question formation. Project 12 deals with figurative expressions that have mixed interpretations. As an instance of 1:many form-meaning mismatches it is, in a way, a mirror case of the phenomena investigated in the present project. Issues of context dependency and semantics constitute a tie between the two projects. Finally, the project involves experimentation, which connects it to projects 1, 3, 7 and 9.

Possible follow-up studies:
1. Widening the empirical domain: ex-situ and in-situ occurrences of non-thematic *wh*-phrases (e.g., adjuncts)
2. Testing more complex *wh*-phrases and their Placement Behaviour (e.g., *who apart from x*, *was für*)
3. Fine-tuning the interaction of factors governing *wh*-phrase placement

Project 7: Imperatives and imperative speech acts across modalities

Supervisors: Nivedita Mani, Markus Steinbach

The form-meaning mismatch: In many different languages, imperative sentence types can be used to express many different but probably related speech acts such as commands, warnings, requests, advice, and pleas, among others, as can be seen in the examples in (11), cf. Condoravdi & Lauer (2012). These different usages lack a morpho-syntactic reflection, rendering imperatives 0:1 form-meaning mismatches.

(11) a. Stand at attention!   (command)
      b. Don't touch the hot plate!   (warning)
      c. Hand me the salt, please!   (request)
      d. Take these pills for a week! (advice)
      e. Please, lend me the money!   (plea)

One of the crucial questions in this context is whether there is one semantic core (sentence mood) and different pragmatic uses (speech acts) of that core or whether there are different semantic cores and therefore different pragmatic uses. A related question concerns the kind of mismatch we are dealing with, i.e. whether each speech act/semantic core has a corresponding grammatical marker or whether the calculation of the pragmatic use of an imperative does not depend on grammatical features of the clause but on contextual information and probably additional speech act indicating devices such as intonation, particles, and gestures. In the last years, various competing formal proposals of a unified semantic representation of imperatives have been developed such as Portner (2007), Condoravdi & Lauer (2012), Kaufmann (2012), von Fintel & Iatridou (2017). In this context, two things are of peculiar interest for new studies on imperatives in sign and spoken languages: (i) the distribution of additional linguistic (including lexical and prosodic) and gestural markers that can be used to narrow down the functions of

imperatives in spoken languages, and (ii) the forms und functions of imperatives in the visual-gestural modality since sign languages are known to directly grammaticalize or pragmaticalize gestural elements in a systematic way (Pfau & Steinbach 2011; van Loon et al. 2014).

Motivation: So far, imperative sentence types in sign languages have not been investigated in detail (Donati et al. 2017). Likewise, not much is known about the use of manual and nonmanual gestural elements to indicate specific speech acts in spoken and sign languages. The importance of facial expression for speech act detection has been investigated in a first experimental study on Italian by Domaneschi et al. (2017). A similar experimental study on the importance of prosody has been conducted by Hellbernd & Sammler (2016). Both studies show that prosodic and nonmanual gestural markers help identifying speech acts. Brentari et al. (2018) focus on the interpretation of imperative sentence types in two unrelated sign languages, i.e. American Sign Languages (ASL) and German Sign Languages (DGS), and show that prosodic cues such as sign duration, head nod, head tilt and wide eyes help the addressee to distinguish different imperative speech acts. Interestingly, these prosodic cues seem to be only partially grammaticalized since their function is also accessible to hearing non-signers. By investigating imperatives across modalities, this project directly builds on recent formal and experimental studies on the form and function of imperatives on the one hand and on the importance of the visual-gestural modality (including co-speech gestures) in the foundations of semantics on the other (Schlenker 2018).

Research questions: Three questions immediately arise: (i) What is the linguistic status of manual and nonmanual gestural markers of imperative speech acts, (ii) do these markers provide evidence for a specific theory of imperative sentence mood, and (iii) how can these markers be integrated into a multimodal (compositional) theory of meaning (an imperative sentence type corresponds to an imperative sentence mood, which can be used to realize different but related speech acts) or do we have a more fine-grained set of sentence types and corresponding sentence moods. A fourth question concerns the status of additional manual and nonmanual markers: are these markers gestural pragmatic speech act markers (pragmaticization) or grammaticalized syntactic markers of sentence types?

Hypotheses and method: As a starter, it is hypothesized that DGS uses various speech act indicating manual and nonmanual markers that have not yet been developed into more general sentence type markers to express different imperative speech acts (i.e. imperatives in DGS are an example of a 0:1 form-meaning mismatch involving further pragmatic specification). Focusing on DGS and spoken German, the PhD student will first start with a grammatical description of the form and function of imperatives in DGS. For that purpose, various controlled elicitation tasks will be conducted to elicit imperative sentence types and imperative speech acts in different contexts. In the second step, the PhD student will check the findings against the new Hamburg DGS corpus. Although the corpus does not yet contain syntactic annotation on the sentence type level, critical contexts and various manual markers that have been identified in the first step can be used to find a considerable number of imperatives in the corpus. As a third step, the PhD student will design an experimental study on the impact of different manual and nonmanual markers such as different facial expressions, pointing or the palm-up gesture on the interpretation of speech acts. This study will be conducted with hearing and deaf subjects. While deaf subjects will only see videos with DGS sentences, hearing subjects will see videos with sentences in German and DGS. The latter is important to decide whether these markers are transparent gestural elements available in both modalities, or whether they are used in sign language as specific markers indicating speech acts or sentence types. In the final step, the PhD student will develop an analysis that integrates recent theoretical developments in the semantics of imperatives and in the cross-modal semantics of 'visual' meaning.

Connection to other research projects: The project investigates type (iii) form-meaning mismatches along with projects 8 and 9. Empirically, it focuses on language variation like projects 1, 4 and 10. This exemplified project is linked to projects 1, 2, 6, 10, and 11 contentwise. Project 1 also deals with the relevance of gestures for linguistic structures and linguistic meaning. Project 2 investigates the relation between discourse and syntactic structures, a topic that is also relevant for project 7. The topic of project 6 is *wh*-questions,

which, like imperatives, is another major sentence type. And finally, projects 10 and 11 deal with particles in different contexts. Since particles are also important in the context of imperative speech acts, the current project will also collaborate with these two projects. Methodologically, the project involves corpus and experimental work, which links it to virtually all other projects for corpus studies, and to projects 1, 3, 6 and 9 for experimental work.

Possible follow-up studies:
1. Manual and nonmanual markers of different kinds of questions in DGS
2. Speech acts and co-speech gestures in spoken languages
3. The processing of speech act indicating gestures across modalities

Project 8: Null objects

Supervisors: Götz Keydana, Guido Mensching

The form-meaning mismatch: Non-overt objects are an emblematic case of the subject of our RTG, see already Rizzi (1986:501): "does an 'understood' [.i.e. 'implicit'] thematic role [...] always correspond to a structural slot in syntactic representations? If not, under what conditions can the mismatch between meaning and form be tolerated?" Both possibilities raise the important additional question of how the unexpressed piece of information is retrieved. Examples of null objects which can be argued to be represented in syntax are given in the following examples, (12) from Classical Latin and (13) from Early Latin:

(12) convocatis   suis   clientibus$_i$ facile incendit   ∅$_i$
      call.together-PPP-INS.PL own-INS.PL dependent-INS.PL easily excite-PERF-3SG
      '[Vercingetorix] summoned together his dependents, and easily excited them.'
                              (Caesar, *de bello Gallico 7.4.1)*, cf. Luraghi (2004: 242)
(13) Si   ∅$_i$ in   ius   ∅$_k$ vocat,   ∅$_k$ ito.
      if   in   law-ACC   call-PRS-3SG   GO-IPV.3SG
      'If someone summons someone else to court, the latter shall come.' (Ius XII tab. I,1)

These cases must be distinguished from other types of empty categories (such as traces) or cases which contain a null element only by appearance (such as intransitive counterparts of transitive verbs). In contrast, (12) can hardly be explained by anything else than a null pronoun. Likewise (13), where we have to reckon with indefinite null subject and object pronouns in the first clause and an anaphoric null subject in the second. Syntactic tests (applicable to historical data) for identifying null object pronouns can be found, among others, in Troberg (2004) and Keydana (2009). The identity of a null object "must be readily ascertainable and understandable to the interlocutor" (Donaldson 2012:63). In addition, it is usually assumed that some type of formal (syntactic) licensing can or even must apply.

Motivation: Null objects of different types have been shown to exist in ancient Indo-European languages (such as Hittite, Sanskrit, Greek, and Latin; see, among many others, Luraghi 2004; Keydana & Luraghi 2012; Inglese et al. 2019). They are attested with anaphoric, cataphoric and arbitrary readings and they occur in any argument position (Keydana 2009). As is evident from (13), in Early Latin they may also have an indefinite reading. In the course of language history, their use has become more and more restricted. However, the diachrony of null objects has rarely been examined in depth. It is expected that the change started with some of the functions and/or contexts mentioned above before spreading to others. Also, it is probably related to other developments in the language system. For e.g., whereas arbitrary and anaphoric null objects can still be found in the modern Romance languages (see Raposo 1986; Rizzi 1986; Authier 1989, among many others), there seems to be no trace of cataphoric usages. Indefinite readings of the type in (13) were probably already obsolete in Classical Latin and became extinct later. Null objects compete with overt (clitic and non-clitic) object pronouns, probably in a non-arbitrary way including complementary distribution. Thus, we expect changes in the pronoun system (e.g. the rise of clitics in Romance) to have an influence on the distribution of null objects. Since at least some null objects are syntactically licensed, we expect their distribution to change if the syntactic system changes. Such a change may even give rise to new configurations in which null objects are licensed. The step from Latin to Romance is a good example. Troberg (2004) shows that Old French had what she calls "topic-comment resumptive

pronouns", the distribution of which depends, among others, on the specific shape of the left periphery in (Old) Romance. The latter being a Romance innovation (cf. Ledgeway 2012:158-168), this type of null objects cannot have been inherited directly from Latin. Another case is the so-called *écrasement* in Old French, i.e. the suppression of an accusative clitic/weak pronoun in dative-accusative clusters (Donaldson 2012:69): Latin had no comparable pronouns. Contrarily, a case of preservation concerns null objects in (mostly final) infinitival Old French adjunct clauses (Donaldson 2012:73-74), which resemble similar cases already found in Early Sanskrit (Keydana 2009:128).

Research questions: If the constraints and rules that regulate the recoverability of null objects guarantee a transparent mapping between form and meaning, a crucial question (within the context of our RTG) that this gives rise to is the following: how is the usage of null objects affected by language change? Specifically, what are the structural and discourse conditions at a certain diachronic stage, and how do they interact with the grammatical system of that stage? Does the loss of a certain type/function of null object correlate with other grammatical changes? Are there shifts within the types of covert objects, i.e. do we find cases in which what is an ellipsis in one stage is reinterpreted as a null pronoun in another stage? Other scenarios are possible, e.g. the emergence of intransitive readings of transitive verbs as a consequence of the loss of a rule that regulates the recoverability of an object. Further perspectives can be found in the literature; e.g. Luraghi (2004:247) hypothesizes (albeit on the evidence of very few data) that during the development from Latin to Early Romance "null objects must have disappeared first in contexts where they were discourse conditioned, while they proved more resistant in syntactically conditioned contexts." Eventually, the answers to such questions can provide insights into the question of why phonetically empty elements exist in the first place. For now, it seems – as far as null objects are concerned – that they follow some kind of economy strategy, but such an idea must be thoroughly assessed on an empirical basis.

Hypotheses and method: The PhD student will perform in depth-research on the development from Late Latin to Early Romance (with a focus on Old French) along the lines of Johnson (1991), which studied this subject, but while focusing on the loss of null objects, lacks a consistent distinction between null objects and intransitive uses of transitive verbs (Luraghi 2004: 236). The research will be corpus-based. For Late Latin, the POS-tagged corpus Le passage du Latin au Français: Corpus Latin (PaLaFraLat), a subcorpus of Base de Français Médiéval (BMF) (see below), is available, for Old French, the Base de Français Médiéval (BFM) and the Nouveau Corpus d'Amsterdam (NCA). On retrieval strategies see, for e.g., Keydana (2009). What is of interest here is, among others, the frequency of null objects versus overt objects over time and the distribution of null objects versus clitic and other resumptive pronouns and the syntactic and discourse-structural conditions governing this distribution. Apart from the general hypotheses that can be drawn from the research questions above, additional hypotheses worth exploring can be found in the literature. E.g. Vincent (2000) claims that the emergence of clitics in Romance led to the end of object omissions inherited from Latin (for discussion, see Donaldson 2012:82, but there is no quantitative research up to now). Schøsler (1999) expects some interference with the null subject property; this is relevant since Old French had a much more restricted null subject system than Latin (cf. Zimmermann 2009, a.o.). The relation between recoverability and the change of null-object systems might be crucial here. The hypothesis is that competition between different accessibility hierarchies (see Holler & Suckow 2016 for an overview) and changes in their role influence the use of null objects over time. Apart from a quantitative study on several types of Latin and/or Romance null object types/configurations, the PhD student is expected to reconstruct the pragmatic and syntactic licensing mechanisms determining the licensing of null objects over time.

Connection to other research projects: With projects 7 and 9, this project examines type (iii) form-meaning mismatches, and with projects 2, 5 and 11, the empirical focus of the project is language change. Contentwise, this project is connected to project 2 for the part that bears on discourse conditions. It is also connected to project 5, which is also concerned with non-canonical forms of arguments. Methodologically, it involves corpus studies and hence connects virtually with all other projects - more specifically, it shares part of the corpora with project 11 as that project, too, bears on Romance languages.

Possible follow-up studies:
1. Null-objects from Old French to Middle French
2. Null-objects in languages with similar time-depth as, for example, Greek or Indo-Aryan
3. Null objects in modern Romance languages: A comparative approach

Project 9: Processing exhaustive inferences

Supervisors: Stavros Skopeteas, Clemens Steiner-Mayr

The form-meaning mismatch: In Hungarian, focus-fronting yields a so-called exhaustive inference. The sentence in (14) with a focus-fronted object regardless of the context in which it is uttered, suggests that Macario fished only a trout.

(14)   [Pisztrángot]$_F$ fogott Macario
         trout   fished Macario
         'M. fished a TROUT and nothing else.' (modified from Skopeteas & Fanselow 2011)

Since there is no explicit expression present in (14) contributing a meaning akin to *only*, this is an instance of a 0:1 form-meaning mismatch.

Motivation: A major challenge for understanding such 0:1 form-meaning mismatches concerns cases where surface-parallel constructions give rise to different interpretations across languages. The cross-linguistic variation in the interpretation of focus-fronting constructions is a case in point (see É.Kiss 1998; Skopeteas & Fanselow 2011). Unlike the Hungarian case in (14), the parallel Greek case in (15) does not always give rise to an exhaustive inference. This is despite the fact that object-fronting has the same properties in both languages according to syntactic diagnostics (the focus must be left-adjacent to the verb; the operation is not clause-bound, and is sensitive to islands for extraction, giving rise to weak crossover effects).

(15)   [Mia   péstrofa]$_F$ psárepse   o   Makarios
         a   trout   fished   the   Makarios
         'M. fished a TROUT.'   (modified from Skopeteas & Fanselow 2011)

Speakers of Hungarian judge (14) always exclude further alternatives (i.e., that M. fished no other fishes) independent of the context it is uttered in. For native speakers of Greek, the exclusion of alternatives, by contrast, is context-dependent (see Skopeteas & Fanselow 2011). In particular, only narrow-focus contexts like (16a) lead to exhaustive inferences.

(16)   a.   in the context: What did M. fish?
                 Hungarian (14) & Greek (15): 'M. fished a trout and nothing else.'
         b.   in the context: What happened?
                 Hungarian (14): 'M. fished a trout and nothing else.'
                 Greek (15): 'M. fished a trout.'

Research questions: At least three kinds of questions may be asked with regards to the exhaustivity effect related to focus fronting. First, could the underlying representations of (14) and (15) in context (16b) differ despite surface appearance? That is, focus fronting in Hungarian may involve an obligatory abstract exhaustivity operator excluding alternatives (Groenendijk & Stokhof 1984; Fox 2007; Spector 2007; Chierchia et al. 2012, a.o.), unlike Greek. Second, is it conceivable that both (14) and (15) always involve exhaustification but that in the Greek case in context (16a), the process is vacuous? Vacuity could be due to the absence of the relevant alternatives for the exhaustification process to negate (e.g. Fox & Katzir 2011; Meyer 2016; Singh et al. 2016, a.o.). Finally, could the exhaustivity effects in Greek be derived in a completely different way than in Hungarian? In particular, this would suggest that the Greek way of doing so is less obligatory.

Hypotheses and method: The hypothesis is that the underlying structural representations for focus-fronting constructions may differ between languages. On the one hand, focus-fronting involves obligatory exhaustification due to a covert operator in Hungarian. For Greek, on the other hand, the effects are brought about by Neo-Gricean pragmatic reasoning (Gazdar 1979; Sauerland 2004 and others). Crucially, such pragmatic inferences are non-obligatory. The aim of this PhD project is to examine the impact of the semantic vs. pragmatic access from form to

meaning on sentence processing. The PhD student will examine this hypothesis by developing methods to examine the reflexes of these different types of exhaustification inferences in processing. Building on previous eye-tracking studies on the effects of focus (Braun et al. 2005; Kim et al. 2015, see overview of psycholinguistic studies in Gotzner & Spalek 2019), the PhD student will examine the impact of the cross-linguistic differences in the sources of exhaustivity (in Hungarian vs. Greek) on the visual search for competing alternatives. Regarding the differences in (16b), participants will be presented with alternatives to the focused referent in (14)/(15) (along with distractors): in a language with inherently exhaustive focus expressions (Hungarian) we expect a larger number of fixations to the alternatives than in a language in which focus evokes alternatives only under particular pragmatic conditions (Greek).

Connection to other research projects: Like projects 7 and 8, this project examines type (iii) form-meaning mismatches, and like projects 3, 6 and 12, this project focuses on language processing and acquisition. It is strongly connected to project 6 (multiple *wh*-questions) and project 2 (from discourse to syntax), since these projects aim at understanding the functional properties of similar syntactic operations in different languages. Furthermore, it is related to project 7 (imperatives), which investigates speech acts. Methodologically, it shares a comparative perspective with projects 4 and 10, and an experimental perspective with projects 1, 3, 6 and 7.

Possible follow-up studies:
Studies on the mapping between syntax and information structure, e.g.,
1. Discourse conditions of cleft constructions in cross-linguistic perspective
2. Semantics of topicalization: scrambling vs. left dislocation
3. Passivization and information structure: Germanic vs. Slavic languages

Project 10: Particles with question and disjunction uses

Supervisors: Uwe Junghanns, Clemens Steiner-Mayr

The form-meaning mismatch: Particles used to form polar questions often can also be used to give rise to other meanings. Slavic languages are a case in point. In some Slavic languages polar questions are formed with a sentence-initial polar particle, e.g. *čy* in the Ukrainian (17a). In some others, the enclitic *li* is used as in the Russian example in (17b).

(17) a. Čy vin pryjde? (Ukrainian) b. Pridët li on? (Russian)
          part he come   come part he
          'Will he come?'

In both languages, however, the particle is also used in non-question environments, in particular for declarative disjunctive sentences as in (18). (While in Old Church Slavonic, *li* occurs on its own in both question and disjunctive uses, in modern Slavic languages *li* in the disjunctive use always co-occurs with another item.)

(18) a. Vin buv veselyj čy   napidpytku.   (Ukrainian)
          he   was happy part   tipsy
          'He was happy or tipsy.'
      b. On   byl vesel ili   pod xmel'kom.   (Russian)
          he   was happy and-part tipsy
          'He was happy or tipsy.'

Motivation: The question-disjunction affinity exhibited by (17) and (18) is observed in a typologically diverse set of languages (e.g. Japanese *ka*, Hindi *ya*, Hausa *ko:*). This suggests that we are not dealing with accidental homophony but rather a more systematic phenomenon. This is to say, one form is associated with more than one meaning. And the two meanings it is associated with are not arbitrary. The fact that the question use is associated with a particular sentence type and thereby also a corresponding speech act, unlike the disjunctive use, makes the phenomenon particularly interesting and difficult to account for. While there is a substantial amount of work on this (e.g. Haspelmath 2007; Haida 2012; Szabolcsi 2015; Mayr accepted), no fully satisfactory proposal exists, whereby consensus is lacking.

Research questions: Assuming one is not facing accidental homophony, the most important question becomes how the two seemingly very different meaning contributions of *čy/li* and other such particles should be related to each other. That is, does a semantic property exist for these particles that makes them usable in both questions and disjunctions? And if so, one would like to know why that property has the effect observed. In order to answer these questions, one also must ask whether the two different uses correlate with distinct morphological, syntactic and/or prosodic properties.

Hypotheses and method: There are at least three hypotheses based on the data pattern one can think of and that will be investigated. Building on observations about English *whether* (i.e., *wh+either* e.g. Karttunen 1977; Guerzoni 2004), the first one is the morphological complexity hypothesis, whereby the uses in (17) are morphologically and therefore semantically more complex than those in (18) despite appearance. This reflects the common assumption that question meanings are taken to be more complex than those of declaratives. The second hypothesis is the operator hypothesis, according to which the meanings of (17) and (18) are indeed equivalent at an embedded level. The examples would, however, differ in that these equivalent constituents are embedded under different syntactic operators giving rise to different meanings (Kratzer & Shimoyama 2002; Fox & Johnson 2016; Ciardelli et al. 2019). Finally, there is the pragmatic hypothesis. Here even the semantic type/representation of (17) and (18) would be parallel, the difference being due to the pragmatics (e.g. through speech act theory, Lauer 2013). It is possible that further variants of these hypotheses must be taken into account (e.g. sentence mode could play a role for hypothesis three). Finally, different hypotheses might well be correct for different languages.

It is a goal of this project to enlarge the set of available empirical data speaking to this phenomenon. This can be achieved by the PhD student through e.g. field work, in-depth work with informants (also via Skype), questionnaires, the mining of text corpora, or large scale online gathering of linguistic judgments using Amazon Mechanical Turk (MTurk). Which of these methods is appropriate depends to a large degree on the languages to be chosen for investigation. Since question speech acts often differ from assertive ones in the prosody, the PhD student will also carry out work analyzing the prosodic patterns using the PRAAT software (Boersma & Weenink 2007). The PhD student is expected to either do a detailed cross-linguistic comparison of languages within one branch of a language family (e.g. Slavic, Chadic, …) or investigate two typologically disparate languages (e.g. Russian vs. Japanese). Based on the generalizations established by this work, he/she will then give a formal analysis of the data, taking into account different levels of linguistic analyses. In particular, the morpho-syntactic analysis should be complemented by an adequate semantic-pragmatic account. Drawing on the collected data, the PhD student will choose one of the hypotheses outlined above (or an alternative one) to develop his/her account.

Connection to other research projects: With projects 11 and 12, this project investigates type (iv) form-meaning mismatches, and with projects 1, 4 and 7, this project focuses on language variation empirically. Contentwise, this is connected to projects 6 and 11. Project 6 dealing with *wh*-questions shows a relation to the present project because the question-disjunction affinity has been argued to be a subcase of the broader question-existential affinity covering also *wh*-questions. The work of project 11 is important to the present project given that it also deals with a 1:many form-meaning mismatch involving particles. The common interests are the structure of particles, the reconstruction of meanings/functions, and the development of particles in different languages. Methodologically, the project shares the cross-linguistic perspective with projects 4 and 9, and fieldwork methods with project 2. Since the project also involves corpus study, it is connected to virtually all other projects.

Possible follow-up studies:
1. The wider typology of particles with question and disjunction uses
2. The diachronic development of particles with question and disjunction uses
3. Modal particles incorporating question particles

Project 11: Presentational particles

Supervisors: Anke Holler, Guido Mensching

The form-meaning mismatch: This case-study concerns so-called presentational particles, like Standard Arabic *hā*, Modern Greek *na*, French *voilà*, *voici*. All of these mean something like 'here/there is/are', but they also contribute additional meaning components beyond stating the presence of something/somebody, thus constituting a case of 1:many form-meaning mismatches. Syntactically, these elements partially behave like verbs, in that they can take a theme complement which can also appear as a clitic (e.g., Greek *náto*, French *le voilà*, Italian *eccolo*, Spanish *helo aquí*, all meaning 'here he/it is'). While French *voilà*/*voici* still shows some degree of transparency (as it contains the morpheme *voi* 'to see' and the locative deictic pronouns *ci* and *là*), Italian *ecco* and Spanish *he* are not transparent and differ in the expression of locative deixis (which is included in Italian *ecco*, but not in Spanish *he*). This PhD project, therefore, mostly concerns itself with mismatches where meaning contributions are not or only partly overtly reflected, and partially also with cases where morphosyntactic elements appear in a position that does not correspond with their locus of interpretation.

Motivation: This area is particularly attractive for the study of form-meaning mismatches, because presentational particles encode multiple non-expressed meaning components, which are partially pragmatic in nature. For e.g., Julia (2016) describes their meaning as to invite the hearer to take note of a sudden occurrence or the arrival of a person, or to highlight an object during the conversation. Sometimes, such pragmatic components still show a certain degree of transparency. For e.g., the fact that presentational particles realize commissive speech acts can still be seen in some languages, in which these particles behave syntactically like imperatives, e.g. with respect to the position of clitics (enclisis) in Greek, but also in Spanish and Italian (unlike in French, which shows proclisis as in indicatives), see the following contrast between Italian and French:

(19) | Lo | guardi. | Guardalo! | Eccolo. | (Italian) |
| --- | --- | --- | --- | --- |
| him= | watch.PRES.IND-2SG. | watch.IMP.2SG =him | PTCL=him | |
| 'You watch him.' | | 'Watch him!' | 'Here he is.' | |

(20) | Tu | le | regarde. | Regarde-le! | Le | voilà. | (French) |
| --- | --- | --- | --- | --- | --- | --- |
| you= | him= | watch.PRES.IND-2SG | watch.IMP.2SG =him | him= | PTCL | |
| 'You watch him.' | | | 'Watch him!' | 'Here he is.' | | |

Diachronic analyses can provide us with insights into the emergence of such meaning components that today are only partially or even implicitly represented. The verb-like behaviour of presentational particles can sometimes be traced back to their verbal origin, but this is not always the case (e.g. in Greek, if it is true that *na* derives from the conjunction *hína*, see, among others, Roussou & Tsangalidis 2010). These and other observations (e.g. concerning the behaviour of locative elements) trigger the impression that languages tend to build up presentational particles and their structure according to some universal scheme. Against this background, it is important to consider how presentational particles can be analysed in a concrete framework. While they might lend support for an analysis within construction grammar (cf., e.g., Bergen & Plauché 2005), within the RTG we would like to address the question if and how these and other meanings/functions can be brought about in a compositional analysis as well. For example, within a generative syntactic model, part of the properties should be encoded in the lexical entry of the item itself, whereas others ought to follow from feature-checking or movement to certain positions in the left periphery (cf. Zanuttini's 2014 analysis of Italian *ecco* carrying a covert locative element in FocP and being the head of a higher SpeakerP). Since these particles also encode directive force (diachronically derived from the imperative in French), other categories, such as Rizzi's ForceP, must also be involved. In an HPSG framework, by contrast, presentational particles can be modelled as sign-based constructions, without contradicting the compositional approach as a whole.

Research questions: The leading question is how presentational particles develop over time in different languages. When and in which order do the different components that make up a presentational particle and their structure appear? Does a particular pragmatic function appear together with certain elements or structures? Language contact is of particular interest, because presentational particles are often loans from other languages. Does a particular language adopt a presentational particle together with the structure of the source language? Finally, it should be

asked which framework of analysis (see above) is the most appropriate one to model the detected changes.

Hypotheses and method: The main objective of this project is to follow the diachronic paths of these particles in at least two Romance languages, mostly using corpus analyses in order to find out how the complex meaning, the syntactic structures, and the pragmatic functions of these items arose and developed. Examples for corpora that can be used for this purpose are El Corpus Diacrónico del Español (CORDE) and Corpus de Documentos Españoles Anteriores a 1800 (CODEA) (for the history of Spanish), FRANTEXT for French (plus specific corpora for the medieval stages, such as NCA and BMF, introduced earlier, for Old French), Opera del Vocabolario Italiano (OVI) for Old Italian, etc. The elements at stake are, among others, French *viola*/*voici*, Italian *ecco*, Spanish *he*, and similar Romance items. Diachronically, Italian *ecco*, Spanish *he*, and French *voilà*/*voici* have totally different origins: the former probably stems from the Latin presentational particle *ecce* plus the masculine singular accusative pronoun *eum*; Spanish *he* is a loan from the Arabic presentational particle mentioned above, whereas *voilà/voici* is a Romance formation that was already visible in Old French (von Wartburg 1961: 426–427). One basic hypothesis is that the history of presentational particles is an interplay between semantic bleaching and a tendency to reconstitute the parts that are not interpretable anymore. However, it seems that overt transparency is not built up again immediately. For e.g., the locative deixis seems to have been implicitly present in the Spanish particle *he* (corresponding to the Arabic source item), but started to become obligatorily accompanied by locative adverbs in the 15th century (Corominas & Pascual 1989: 334–335) – here we expect that this was preceded by a period in which the adverb was optional. As far as the formal representation of presentational particles are concerned, a basic working hypothesis in this project is that a mere description of items such as defective verbs (see Bouchard 1988) does not suffice, given the numerous other properties of those particles, many of which lie outside the verbal domain in the strict sense. If these particles can be analyzed in compositional terms, the question arises as to how this can be realized in a concrete model.

Connection to other research projects: The project is connected to projects 10 and 12 in that it examines type (iv) form-meaning mismatches, and to projects 2, 5 and 8 in that it focuses on language change empirically. Contentwise, this project is connected to project 7, with which it shares the interest for directive speech acts and imperative sentence mood, as well as to project 10, which focuses on another type of particles, also with respect to a type (iv) mismatch. Finally, it connects with practically all other projects since it involves corpus studies. Specifically, it shares part of the corpora with project 8, which also bears on Romance languages.

Possible follow-up studies:
1. Presentational particles in Arabic dialects
2. Presentational relative clauses
3. Diachronic sources of presentational particles typologically

Project 12: Identifying context-dependent meaning components of figurative expressions

Supervisors: Anke Holler, Caroline Sporleder

The form-meaning mismatch: Many figurative expressions can also have a literal interpretation, depending on the context. This is particularly prevalent for verb phrase idioms as in examples (21a) and (21b). Additionally, in some cases, both interpretations can even hold simultaneously (22-23). Such "mixed usages" do not form a homogeneous class: In example (22) both the literal and the figurative interpretation would be felicitous given the context, though one — typically the idiomatic one — might be much more likely. In other cases, an idiom is embedded in a metaphor and takes on a literal meaning within the figurative meaning of the surrounding metaphorical context (example (23)). In (24), this figurative reading is dominant but the idiom is chosen in such a way that its constituents exhibit lexical cohesion (Halliday & Hasan 1976) with the literal context. Such mixed usages are often chosen for humorous effects. They constitute a form-meaning mismatch of type (iv) in that one form can take on literal and figurative meanings, components of which may even be accessible simultaneously.

(21)   a.   Dad had to <u>break the ice</u> on the chicken troughs so that they could get water.

> [Gigaword Corpus NYT200008]
>
> b.   If you've just moved to a new area a good way to <u>break the ice</u> for you and your child is a parent and toddler group. [BNC AAY]

(22)   <u>Left holding the baby</u>, single mothers find it hard to fend for themselves. [BNC CRA]

(23)   You're like a restless bird in a cage. When you get out of the cage, you'll <u>fly</u> very <u>high</u>. [BNC FR6]

(24)   Wenn das Kind gerade einen fieberhaften Infekt hat, ist den meisten Eltern klar, dass das Planschen buchstäblich erst einmal <u>ins Wasser fallen</u> muss.
[Web, www.praxisvita.de]

Motivation: Figurative expressions generally pose a challenge for natural language processing (Sag et al. 2002), and a robust system for natural language understanding needs to be able to detect potentially figurative expressions and to determine their interpretation in the given context. This is typically done by comparing the constituents of the target expression to the context and computing for example the semantic 'fit' (e.g. in terms of lexical cohesion) between the context and the expression (Li & Sporleder 2009; Feldman & Peng 2013; Peng et al. 2014; Salton et al. 2016; King & Cook 2018). Under the assumption that the semantics of figurative expressions is not (or not entirely) compositional a good 'fit' with the context indicate a literal interpretation. However, this approach is bound to fail for mixed usages, which exhibit cohesion under both interpretations. While mixed usages are rare, their frequency is not negligible (Sporleder et al. 2010).

Research Questions: This project aims at exploring linguistically-informed algorithmic approaches to identifying and interpreting occurrences of figurative expressions which have mixed interpretations in a given context. Given that multi-word figurative expressions are generally assumed to be morpho-syntactically 'frozen' (Fraser 1970) and semantically non-compositional to varying degrees (Nunberg et al. 1994; Fellbaum 2014), the question arises under which circumstances mixed usages are possible. For instance, one might hypothesize that expressions which are semantically more transparent and/or syntactically less fixed are more likely to allow mixed interpretations. In addition, it is also possible that certain properties of the context license such interpretations. Furthermore, current computational models of figurative expression interpretation tacitly assume that figurativeness is a boolean, all-or-nothing category, whereas examples such as those above suggest that one needs to account for (i) different interpretations holding simultaneously and (ii) the possibility of nested figurative and literal interpretations (as in example 23).

Hypotheses and method: As a first step, the PhD student will carry out a corpus study in order to categorise mixed usages and identify subclasses. A good starting point is the Idioms in Context (IDIX) corpus (Sporleder et al. 2010) which contains annotations of literal, figurative and mixed interpretations for 5.836 occurrences of 78 different idioms in the British National Corpus (BNC). The next task will be to hypothesize about linguistic markers or licensors for mixed usage, both with respect to the target expression itself and the context. Finally, and building on recent advances in distributional semantics and neural language processing (Salton et al. 2016; King & Cook 2018), the PhD student will develop a computational model of mixed usage interpretations.

Connection to other research projects: With projects 10 and 11, this project investigates type (iv) form-meaning mismatches, and with projects 3, 6 and 9, this project focuses on language processing and acquisition. Contentwise, this project is connected to several projects, very closely to projects 3, and 9 since it will benefit from cross-talks on word-form and word-meaning mismatches in language acquisition and on the impact of the semantic vs. pragmatic access from form to meaning on sentence processing.

Possible follow-up studies:
1. Mixed usage figurative expressions across domains; a corpus study
2. Sensitivity of processing of figurative language to internal factors (degree of predictability) and external factors (context sensitivity) using eye-tracking in reading

3. Modeling of mixed usages at the grammar-pragmatics interface

Summing up: The proposed project areas cover a whole range of phenomena crucial to answer the central research question of this RTG. The overall architecture of the RTG and the particular project areas are designed in such a way that together they provide a balanced combination of phenomena touching on different grammatical levels and different linguistic disciplines. Even though it should be emphasized that the suggested research program may be subject to modifications with respect to the final choice of the PhD projects as well as the exact range of project areas, this balance is essential for the overall design of the RTG and the success of its research program.

4 Qualification Programme

PhD students attending the envisaged RTG will be qualified for various careers within and also outside academia. The aim of the RTG is to qualify young researchers for a professional academic career in various fields of modern linguistics at universities, research centers, and similar institutions in Germany and abroad. Other professional avenues include science management and academic consultancy, teaching, and (scientific) writing and publishing. The ongoing organizational changes in science and research drive a strong demand for highly skilled personnel able to serve bridging-functions between science and administration; such 'science managers' require a track record in research, and we expect increasing job opportunities for people with a scientific background and management skills in the near future. Likewise, the demand for external consultants and trainers for science- and communication-related skills is growing. Management and teaching skills will also open employment opportunities in (public) organizations and NGOs. Although the publishing market is currently subject to major organizational and structural changes, and career paths in journalism become increasingly insecure, there is an increasing investment in communication departments in universities, non-university research institutions, and media consulting companies. Linguistic expertise and methodological training – as provided within the qualification program of the RTG – are requested in all of these professional sectors.

Our training program, which consists of four modules, is designed in such a way that students can aim for individual career paths. It addresses specific theoretical and methodological knowledge as well as foundational academic and general skills. It will be implemented in close collaboration with the Graduate School of Humanities Göttingen (Graduiertenschule für Geisteswissenschaften Göttingen, GSGG), which offers a detailed qualification program for PhD students in the humanities. In addition, experimentally oriented PhD students will be enrolled in the Göttingen PhD program Behavior and Cognition (BeCog).

As the scientific environment forms a key factor for the successful training of PhD students, we ensure that students profit in the maximal way from the present faculty members, postdocs, and fellow-students, and from the present infrastructure and facilities. Moreover, international experiences as well as extensive contacts with scholars outside of Göttingen are of the upmost importance for a successful qualification program. To this end, a number of both well-established and young and promising scholars will be invited to come to Göttingen, some for shorter visits, others for longer ones. Also, every candidate is strongly encouraged to frequently attend international and national conferences, and to spend one academic term at a relevant top university or research institute outside Germany, for instance, at MIT, UMass, UConn, UCLA, USC, NYU, the Hebrew University in Jerusalem, Oslo University, Cambridge University or the Radboud University Nijmegen, all universities which run excellent linguistic programs and with which ongoing collaborations do already exist.

Finally, we emphasize that a successful completion of a PhD program is highly dependent on a proper concept of supervision. Therefore, apart from regular supervision by two supervisors, we establish a thesis committee for every PhD student. While the two supervisors and the PhD student meet on a regular basis to discuss all aspects of the ongoing research, the PhD student will have a meeting with the entire thesis committee twice a year to discuss the overall progress and to identify potential problems or obstacles for the student. Postdocs that are affiliated with Göttingen during the RTG will actively be encouraged to participate in thesis committees as well, which helps them in their training to become future supervisors.

4.1 Qualification programme

The qualification activities of the RTG will be organized around four modules illustrated in table 5 and explained below. All PhD students choose courses with an extent of 16 ECTS points (see table 4 for courses offered) and participate in some regular as well as some one-time RTG activities. Table 5 gives a model timetable for the activities in all four modules. These include mandatory and optional activities, the latter naturally in accordance with the individual needs of the PhD students, as it needs to be ensured as well that the students finish their theses on time. Depending on the individual planning, some activities such as organization of the RTG colloquium, co-teaching, presentations at conferences and workshop organization also can be shifted. All activities will be explained in more detail below and in sections 4.2 and 4.3.

The qualification program is designed to meet the different requirements in the course of completing a PhD. In the first semester, the emphasis will be on theoretical background, research methods and consultation with the thesis committee. The second and third semester are dedicated to advanced research and methodological courses and a teaching workshop (given the strong value of teaching experience when entering the job market). In the second and third year, students present their results at national and international conferences. Towards the end of the three-year period, they should attend courses preparing for the post-doctorate career phase. Finally, each PhD student is strongly encouraged to make use of the possibility to spend up to four months as visiting researcher abroad (and to obtain credits there if necessary).

The four modules, around which the activities of the RTG are organized, are set up as follows.

Module 1: Research & methods: In consultation with the thesis committee, the PhD students select three RTG-courses which they attend in the first three semesters. Within the RTG, a program of dedicated theoretical courses (including typology and theoretical models) and a methodological course program (including statistics, experimental methods, data elicitation and corpus linguistics), necessary for a successful accomplishment of the PhD thesis, will be offered. To make sure that all PhD students receive a profound training in the research areas and methods relevant for the RTG, 12 of a total of 16 ECTS points have to be obtained in the courses in module 1 (the other 4 ECTS points have to be obtained in modules 2 and 3). These will be provided by members of the RTG and by experienced guest researchers. In addition, this module consists of a weekly RTG colloquium with presentations of the participating researchers, PhD students, existing postdocs, as well as outstanding international guest researchers (cf. section 4.2), two mandatory annual meetings with the thesis committee (see below) and a yearly two-day retreat of all members of the RTG (cf. section 4.3). The PhD students are expected to hold one presentation in the RTG colloquium each year to present the progress they made. This also serves as an exercise for presentations at international conferences. Finally, in every cohort one thematic summer school will be organized.

Module 2: Teaching skills: Gaining teaching experience within the course of the RTG is strongly encouraged. In the first year, students can attend a basic workshop on academic teaching. In the second year, they can participate in the Hetairos Program of the GSGG (https://www.uni-goettingen.de/en/127348.html), or its future successor (https://www.uni-goettingen.de/en/409598.html). These programs consist of different components, among which: (i) specific didactic workshops that give an intensive introduction to the theory and practice of academic teaching, and (ii) co-teaching of a joint seminar together with an experienced university teacher. After successfully completing such a program, the PhD students are able to provide a tutorial for undergraduates on their area of expertise.

Module 3: Key qualifications (incl. conference presentations) and career planning: In this module, a number of workshops are organized ranging from science and project management, presentation skills, scientific writing (including abstract writing) to grant writing and academic career planning. While some of these workshops are realized by the GSGG, other workshops are arranged by members of the RTG. In addition, we will organize workshops with professional coaches specialized on topics such as career planning, work-life balance etc. Because of the growing importance of public communication of scientific activities, we will also offer a workshop on the topic of dissemination in which students will be trained in the writing of press releases, the handling of interviews, social media marketing and related issues by a professional coach. These

activities will be particularly useful for students who consider a career in media or communication departments. But based on our own experiences, training on PR and media matters gains increasing importance also within academia.

Concerning conference presentations, each PhD student is expected to give two (poster) presentations at peer reviewed international conferences. These presentations can be in collaboration with the supervisor(s). Additional participation in international conferences is highly encouraged right from the beginning of the doctorate for reasons of networking and international experience. However, PhD students are not expected to present their own work on all conferences they participate in as this would mean too high a work load besides their actual research.

Module 4: Organizational skills: Every semester a group of generally four PhD students organizes the weekly RTG colloquium in consultation with the speaker of the RTG (i.e., each PhD student organizes the colloquium twice in the three-year period). In addition, they meet in reading groups on topics relevant for their doctoral projects. Finally, all PhD students can get involved in the organization of an international workshop at the end of the second year.

The overall 4-module qualification program is designed to meet the individual requirements of each PhD student. Therefore, it must be adaptable to individual needs and remain flexible, including only a small number of mandatory events (each type of event will be discussed in detail below):

- Two meetings per year with the thesis committee
- Kick-off meeting (two-day retreat) and annual retreat (two-day retreat)
- Weekly RTG colloquium
- Special events with RTG-special guests (e.g. a lecture series, a workshop or round table discussions, approx. one per semester)
- One summer school per cohort
- One international workshop per cohort
- Reading groups

Furthermore, each PhD student can choose from a number of courses that will be offered within the RTG or in collaboration with the GSGG (see table 4 for details). As stated above, a total of 16 ECTS points have to be obtained in these courses, meaning that every PhD student has to participate in five courses in total.

Mod.	Content	Responsible	Sem.	Duration	ECTS	
1	Morphology and syntax	Mensching	1.	1 Sem.	4	12 ECTS points in total
1	Semantics and pragmatics	Steiner-Mayr	1.	1 Sem.	4	
1	Morphosyntax/semantics interface	Zeijlstra	2.	1 Sem.	4	
1	Morphosyntax/pragmatics interface	Skopeteas	2.	1 Sem.	4	
1	Theories of grammar	Holler	2.	1 Sem.	4	
1	Experimental linguistics	Mani	3.	1 Sem.	4	
1	Statistical analysis	Weskott	3.	1 Sem.	4	
1	Corpus linguistics	Coniglio	3.	1 Sem.	4	
1	Language documentation	Keydana	3.	1 Sem.	4	
1	Data editing and data management*	Sporleder	3.	1 Sem.	4	
2	Teaching skills*	Steinbach	2.	2 days	2	4 ECTS points in total
3	Scientific writing*	Weskott	3.	2 days	2	
3	Presentation skills*	Mani	3.	2 days	2	
3	Conflict management*	Junghanns	4.	2 days	2	
3	Science and project management*	Holler	4.	2 days	2	
3	Dissemination/transfer of knowledge*	Steinbach	5.	2 days	2	
3	Grant writing, career planning*	Zeijlstra	5.	2 days	2	
3	Work-life balance*	Skopeteas	5.	2 days	2	

Table 4: List of RTG courses with ECTS points[1]

---

[1] 'Courses marked with * are to be provided by external consultants and/or professional coaches under auspices of the listed participating researchers and the speaker.

In the table below (table 5) we show what a schedule of activities and courses of a RTG member may look like. Note that this combines mandatory as well as optional activities. More details on the different activities will be given subsequently.

			Semester					
			1.	2.	3.	4.	5.	6.
Module 1	Research & methods	(At least) 3 selected RTG courses (see table 4) (M)	X	X	X			
		Thesis committee meeting (M)	X	X	X	X	X	X
		Kick-off meeting and retreat (M)	X	X		X		X
		Weekly RTG colloquium (participation every semester, presentation once per year) (M)	X	X	X	X	X	X
Module 2	Teaching skills	Teaching workshop		X				
		Co-teaching (Hetairos Program or follow-up)			X			
		Tutorial				X		
Module 3	Key qualifications	(At least) 2 selected RTG courses (see table 4) (M)				X	X	
		Active conference presentation (M)			X		X	
Module 4	Organizational skills	Organization of the RTG colloquium (M)	X				X	
		Reading group (M)		X	X	X	X	
		Workshop organization				X		
Special events		Events with RTG-special guests (M)	X	X	X	X	X	X
		Summer school (M)				X		
		International workshop (M)				X		

Table 5: Sample schedule of the qualification activities (M = mandatory)

Thesis committees: Apart from regular supervision, each PhD student will be provided with a thesis committee which consists of at least the two main supervisors (both participating researchers) and one consultant advisor, chosen from the list of participating researchers, or – in special cases – from outside the RTG, e.g. RTG-special guests (cf. section 4.2) or other external academic experts. At the beginning, the PhD student and the thesis committee members sign an agreement of supervision as provided by the examination regulations for PhD students of the Faculty of Humanities. Regular meetings of the thesis committee and the PhD student take place twice a year. At the first meeting, the candidate discusses his/her career goals with the thesis committee and identifies targets for an individual development plan. Based on this discussion, the thesis committee and the PhD student develop an individual research and study plan including a selection of courses chosen from table 4. This plan will be evaluated and adjusted at the next meeting according to the individual development of each PhD student. Since all students will be enrolled in the GSGG, they also benefit from the qualification program and all other activities, possibilities and facilities offered by the graduate school.

Kick-off meeting: The kick-off meeting, which will be held in the first or second month of each cohort, pursues multiple targets. In order to provide sufficient time for discussion and socializing, this first meeting will be organized as a two-day retreat outside of Göttingen. First, we want to discuss the students' expectations for the RTG, as these shall be accommodated in the RTG's curriculum. Second, the students briefly present and discuss their PhD proposals. Third, we establish a first network of topics investigated in the first cohort to promote intellectual exchange between the PhD students. And finally, the students are introduced to various topics such as the goals, structure and requirements of the RTG, the rights and obligations of students and supervisors, research ethics and rules of good scientific practice (including regulations for studies on human subjects), data handling and storage procedures. The university's Ombudsperson and the Göttingen eResearch Alliance will be asked to contribute to the topics of good scientific practice and research data management. The scientific discussion will be continued at the annual retreat, which will be used to discuss first results and future research directions and to exchange ideas and experiences.

At the end of the third year, we will organize a feedback workshop with the PhD candidates of the first cohort. In addition, for the second cohort, PhD students from the first

cohort will be invited to a one-day workshop facilitating cohort-bridging networking. In the first part of this workshop, the former PhD students present the main results of their PhD theses. In the second part, they report and reflect their experiences in the RTG, thereby providing helpful insights and knowledge transfer. Selected alumni may also be invited for presentations in the RTG colloquium and for career workshops.

4.2 Visiting researchers and Mercator fellows

In order to maximally enhance contributions of visiting scholars, the RTG sets up two series of research visits: short visits of international and national scholars (both well-established senior researchers and promising junior researchers) and long term visits by internationally respected scholars (RTG-special guests).

The weekly RTG colloquium: As for the former, every semester the RTG will invite six scholars to contribute to the weekly RTG colloquium. In this colloquium, scholars will present their latest discoveries and developments in their research areas. An important aspect of this colloquium series is that it is organized in a bottom-up fashion: projects determine the lists of guests and may also enjoy other activities around the invited guests, for instance, an additional guest lecture, a round table session, or a best practice meeting. Every project will invite one colloquium speaker per year, yielding twelve speakers in total. To facilitate networking, the PhD students will be asked to propose invitees and invite them, serve as hosts, chair sessions and promote discussions. Apart from the usual dinner after the colloquium talk, which will be open for everybody, each visit shall also include lunch with the guests with the PhD students only. This further improves the exchange with the guests in an informal setting. Finally, the hosting PhD student also discusses the research of the guest before and after the visit in the reading groups. This weekly colloquium will be another cornerstone of the program's schedule guaranteeing regular exchange.

RTG-special guests: Apart from these twelve annual colloquium visitors, it is important that world-renowned experts on the various areas of investigation this RTG covers will stay for a longer time to have longer and more in-depth discussion and counseling meetings with students, existing postdocs and participating researchers. Those RTG-special guests should be able to provide a series of lectures or workshops, possibly also teach one of the classes offered in the qualification program (cf. section 4.1) and be available for discussions and joint research projects. In order to ensure that often very busy scholars will still be able to make it to Göttingen for a longer time, these visits will last four to max. six weeks. During the visit, the RTG-special guests will also participate in all regular RTG activities. Moreover, it is expected that RTG-special guests will be involved in two of the projects and support PhD students that invited them, for instance, as external advisor for the project, thesis committee member, host for research visits, etc. In order to ensure that every project will profit equally well from this special guest series, two projects jointly invite one special guest so that the number of six guests per cohort is reached. Potential guests include Angelika Kratzer (UMass Amherst), Philippe Schlenker (ENS Paris), Greville Corbett (University of Surrey), Maria Polinsky (University of Maryland), Gennaro Chierchia (Harvard) and Rafaella Zanuttini (Yale).

4.3 Additional qualification measures

Annual retreat: An important building block to ensure quality management will be the annual two-day retreat mentioned in section 4.1. This meeting outside Göttingen will not only provide intensive in-depth discussions about the development and progress of the PhD students, but stimulates social interactions beyond the scope of the weekly RTG colloquium and serves an important team-building function. As an initiative to foster gender equality, we will also invite a number of role model female and male scientists to the retreat to speak about how to combine a scientific career with family life. This shall provide our students with the opportunity to learn from best practice examples. We emphasize that a successful combination of work and family involves both genders. Furthermore, the annual retreat will contain a session on career development, in which we will invite previous PhD students of *Linguistics in Göttingen* to share their experience from the job market and the work opportunities (within or outside academia) with the PhD students of the RTG.

Summer school: The regular qualification program will be complemented by a summer school

held once for each cohort. This summer school will last six days (five days of teaching and a one-day workshop) and will be organized by the current postdocs and the involved participating researchers. The summer school aims to address one of the topics investigated in the RTG from various perspectives and to combine empirical and theoretical approaches to stimulate the development of concepts and future plans by the participating PhD students. To this end, three to five international scholars are invited to teach a course on the selected topic, which will also foster the international network of the PhD students. The summer school will also be open for up to 15 students from other universities. Several applicants (Coniglio, Holler, Steinbach, Steiner-Mayr, Skopeteas, Zeijlstra) have already organized linguistic summer schools of this kind. These previous experiences have shown that the PhD students benefit from the inclusion of peers from other universities and that summer schools can serve as advertisement for the RTG and the University of Göttingen.

International workshop: An international workshop will be organized once for each cohort. PhD students will be involved in the selection of speakers and can gain experiences in project management by supporting the speaker and other participating and associated researchers in the organization of the workshop. For this, an optional key skill course on workshop organization will be offered. Like in the RTG colloquium, PhD students will be encouraged to serve as hosts, chair talks and promote discussions.

Reading groups: The reading groups are an event organized by the PhD students, assisted by the participating and associated researchers, that should take place on a regular basis and cover literature that is important for a number of PhD students. Current postdocs will play an active part in guidance and discussion within the reading groups. These reading groups (consisting of approximately four PhD students per group), will also serve to help PhD students support each other during the time of their PhD with regard to diverse issues, such as handling the work load, communicating with the PhD advisor, etc. Such groups have the advantage that they can deal with daily concerns immediately and in a flexible way, so that PhD students profit from each other's different experiences and competences. We believe that the creation of such groups can foster a climate of cooperation as well as reflection of one's limits and that it can help in the search of solutions to common as well as individual problems. It is up to the PhD students whether and in which way they want to organize themselves. The supervisors and especially the postdocs will offer support and advice based on their own experiences with mentoring programs and peer-to-peer teaching.

International outreach: In line with our goal of reaching high international visibility, we expect our students to give presentations at international peer-reviewed conferences and – if possible – to publish their results in international peer-reviewed journals. Correspondingly, the new examination regulations for PhD students of the Faculty of Humanities allow the submission of cumulative, publication-based theses in place of monographs. These new regulations facilitate more flexibility for the PhD students. Especially candidates working in experimental linguistics are thus enabled to include already published experimental results in the PhD thesis submitted at the end of the three years.

 To further enhance international experience, to provide additional topic-specific academic training and to have students participate actively in an international community, all PhD students are strongly encouraged to spend up to four months as visiting researchers abroad. To foster further exchange, funding will be made available to visit other labs for short- and mid-term stays, i.e., to learn novel techniques or to work on joint projects. Based on our experiences, we are convinced that the PhD students will have excellent opportunities to establish international contacts, which, in turn, will give them a decisive advantage when pursuing their careers after PhD. The participating researchers have proficient international contacts that ensure that every PhD student can find an appropriate international department or research institute to spend a semester; previous linguistic PhD students from Göttingen have, for e.g., spent a semester at Harvard, MIT or the Universidade Federal de Santa Catarina (Florianópolis, Brazil).

Assistance: The coordination office of the RTG will support incoming students in all academic and non-academic matters. All students will be integrated in the research groups of their supervisors and events and courses organized by the RTG will in principle be held in English.