Predictable ways of being unpredictable: Unconventional uses of verbal constructions
Project description – BOF DOCPRO4

Supervisors: Astrid De Wit (astrid.dewit@uantwerpen.be) and Peter Petré (peter.petre@uantwerpen.be)

The main goal of this project is to unveil cross-linguistic systematicity in the way language users exploit certain syntactic structures to encode **unconventionality**, i.e. to indicate that the situation they are reporting is somehow non-canonical or that the circumstances that conventionally surround a speech event do not pertain. In what follows, we will provide a more precise discussion of the theoretical and methodological objectives of this project, as well as a detailed overview of the steps taken to attain these goals.

1. State of the art

Language users can opt for certain linguistic constructions for the sole purpose of conveying **unconventionality**, i.e. as a way to stand out communicatively. This phenomenon has also been referred to by means of related notions, such as surprisal (Levy & Jaeger 2007), salience (Schmid & Günther 2016), or extravagance (Haspelmath 1999; Petré 2017). These alternate concepts, however, are either not fully interchangeable or have accumulated too many different applications. As an information-theoretical notion, surprisal is more concerned with a correlation between syntactic patterns and processing difficulty. Salience has been used in a variety of ways, some speaker-oriented, others hearer-oriented, which makes the concept at risk of misinterpretation. Extravagance is a term used to avoid confusion with expressivity (which pertains to artistic language use), and probably comes closest to what we are after, yet it has only found limited use outside grammaticalization studies (but see Ungerer & Hartmann 2020). We have therefore opted for 'unconventionality', broadly defined as 'linguistically signaling that there are aspects of the speech event or its immediate circumstances that deviate from what is considered conventional by the speaker and/or that the combination of the construction and its linguistic context is somehow non-canonical'. Exploitation of certain linguistic elements for reasons of unconventionality has been attested across a wide variety of domains, in different languages – examples are (i) the pervasive spread of the affix *-ish* (as in *everybody-ish*) (Traugott & Trousdale 2013: 233-7); (ii) expressive diminutives in Dutch, e.g. *kletsjenat* very-dim-wet 'very wet' (Norde & Morris 2018: 47); or (iii) cross-linguistically attested reduplication strategies, as in English *I like it quite a lot but not A-LOT-a-lot* (Eitelmann & Haumann 2019). In spite of the scholarly attention devoted to unconventional language use, two crucial aspects remain understudied: (i) unconventionality is not only/primarily expressed by lexical and morphological elements (the "classical loci for exceptions" (Simon & Wiese 2011)), but also by syntactic constructions, and (ii) some types of syntactic constructions appear notably amenable to signal unconventionality, persistently so across different stages of their relative development and irrespective of the language in which they are used.

(i) Unconventionality in syntax. Functionalist accounts of morphosyntactic constructions have amply shown that these syntactic constructions do not just create coherence between words and clauses, but also carry meaning. Yet in many functionalist semantic analyses, such constructions are still regarded as automated (Givón 2001: 15), i.e. unconsciously and routinely used and therefore unavailable for deliberate communicative exploitation. For instance, most studies on serial-verb GO- and COME-constructions in different languages primarily focus on the prototypical, automated aspectotemporal functions that have evolved out of their original motion meaning, including, most notably, the expression of futurity for GO and recent past-time reference for COME (e.g. Bybee et al. 1994: 56, 266-70). Similarly, analyses of progressive constructions, such as English *be* V-*ing* or the Dutch posture-verb progressive (*zitten/liggen/staan* + to-inf), typically discuss uses in which these constructions express an aspectotemporal meaning of ongoingness (e.g. Leech 2004; Lemmens 2005). However, semantic analyses of GO and COME-auxiliaries and progressive constructions in individual languages are often faced with natural language examples that fall outside the purview of aspectotemporal meanings, thereby challenging the received wisdom that such meanings form the functional core of these constructions. For instance, accounts in terms of inchoativity or futurity do not

capture the difference between (1a) and (1b); (2a) and (2b) do not differ from each other in terms of pastness; and the difference between (3a) and (3b), featuring a Dutch posture-verb progressive, cannot be accounted for in terms of ongoingness:

(1) a. *T'* *es* *malade* – *où* ***as***-*tu* ***attrapé*** *ça ?*
2sg be.prs.2sg ill where have.prs.2sg-2sg catch.pst.ptcp that
b. *T'* *es* *malade* – *où* *t'* ***es*** ***allé*** ***attraper*** *ça ?*
2sg be.prs.2sg ill where2sg be.prs.2sg go.pst.ptcp catch that
'You're ill – where did you catch that / where did you go catching that?'
(Celle & Lansari 2015: 300)

(2) a. *So don't* ***whine*** *about it now*.
b. *So do not* ***come whining*** *now*. (Corpus of Contemporary American English)

(3) a. *En* *toen* ***keek*** *ik* *zo* *twee, drie* *seconden* *naar* *die* *fles.*
and then watch.pst.1sg 1sg so two three seconds at that bottle
'And then I looked at the bottle for two, three seconds or so.' (SoNaR)
b. *Bij* *elke* *straatje* *op* *de* *hoek* ***zat*** *ik* *echt* *zo* *te*
on every little.street at the corner sit.pst.1sg 1sg really so to
kijken *van* *achtervolgt* *mij* *niemand.* (Corpus of Spoken Dutch)
watch of follows me no one
'At every street corner I was really like looking like is nobody following me.'

Rather, a closer look at the contexts in which French *aller* + V-inf, English *come* + V-inf, and the Dutch progressive are being used reveals that they frequently feature expressions of emotional involvement on the part of the speaker, which reflects that the reported situations are perceived as somewhat out of the ordinary (e.g. Celle & Lansari 2015; Anthonissen et al. 2019). Some authors have paid attention to this non-aspectotemporal import that serial-verb GO and COME and the progressive appear to have in different languages (for GO and COME: e.g. Stefanowitsch 1999; Detges 2000; Bourdin 2003; Bres & Labeau 2013; Devos & van der Wal 2014; Celle & Lansari 2015; for the progressive: e.g. Goldsmith & Woisetschlaeger 1982; Franckel 1989; De Wit et al. 2020). Yet these existing accounts, revealing though they may be, focus on a single construction type only (serially used motion verbs or progressives) and have thus not looked for generalizations across these construction types. As a result, they fail to recognize how **similar** these different constructions are in their propensity to be used for reasons of highlighting – a similarity they moreover seem to share with other types of verbal constructions such as the present perfect and independently used infinitives (De Wit & Vandekerckhove 2021). This functional overlap suggests the presence of shared semantic features which have not been unveiled yet.

(ii) Stable unconventionality. What is also generally neglected is how **stable** this propensity for unconventional usage appears to be across languages. A common assumption underlying analyses of unconventional uses of linguistic constructions is that as soon as they become conventionalized, they automatically lose their capacity to be used for expressive purposes (cf. e.g. Haiman 2014: 78-80). Yet unconventional uses of the progressive, GO and COME are crosslinguistically attested (even) at advanced degrees of grammaticalization, i.e., apparently, these constructions continue to find new contexts in which they stand out and signal a sense of atypicality, even when they have become conventionalized in their aspectotemporal use in a given language. For instance, it has been noted that the progressive can be used in unexpected contexts to express unconventionality in a range of languages irrespective of its degree of grammaticalization (see e.g., Güldemann (2003) on Bantu languages; Donabédian (2012) on Western Armenian; De Wit et al. (2020) on English, Dutch and French). Similarly, pilot studies of French and English COME- and GO-auxiliaries reveal that these constructions have specifically been recruited to signal non-canonicity at all stages of their development (see e.g. Visser 1969: 1400; Bres & Labeau 2013). In other words, the current association of unconventional language use with transient properties such as novelty and low degree of grammaticalization (e.g., Haspelmath 1999; Kempf & Hartmann 2019) causes us to overlook the steadiness with which specific verbal constructions continue to be used for highlighting.

2. Describe the objectives of the research project.

This project sets out to demonstrate that, because of their intrinsic semantic properties, progressives and serial-verb GO/COME-constructions are recruited significantly more often for unconventionality purposes than their simple tense alternatives in the verbal paradigm. More specifically, it reveals that the use of the progressive and GO/COME to signal unconventionality is attested across a diverse array of languages and that, even as these constructions get more conventionalized in the inventory of grammatical constructions of individual languages, they continue to be exploited to attract attention in new linguistic environments in these languages. We hypothesize that this inherent and stable association of these verbal constructions with unconventionality is a result of their semantics: each of these constructions indicates that the situation they report is not seen as fully and regularly integrated in or somehow distant from what the speaker conceives of as the 'ground' (i.e. the speech event, its participants and its immediate circumstances (Langacker 2008)). This general research objective breaks down into five more specific goals:

1) *The concept of linguistic unconventionality*. In Section 1, we have proposed a working definition of linguistic unconventionality, based on insights from our previous work (Petré 2017; De Wit et al. 2020). Yet there are various other definitions and associated operationalizations of linguistic unconventionality, and related concepts such as extravagance and expressivity, leading to some vagueness and a proliferation of interpretations (see Ungerer & Hartmann 2020). A first objective of this project is therefore to chart the various existing definitions and operationalizations in order to create some clarity in this domain of linguistics, and, if necessary, take from these definitions what we need to supplement our own conceptualization for a cross-linguistic analysis.
2) *Operationalization*. The proposed project is also innovative in its combination of qualitative analysis (cf. goal (5)) with an elaborate quantitative methodology, allowing to establish if unconventionality is robustly rather than haphazardly attested. This first entails, as set out in detail in the methodology section, customizing the operationalization of the notion of unconventionality as developed in previous studies (Petré 2017; Ungerer & Hartmann 2020) so that it can be uniformly applied to a wider range of contemporary data in multiple languages.
3) *Cross-linguistic study*. Second, by making innovative use of translation corpora, we will test whether, across languages, the relative proportion of uses of progressives and GO/COME constructions for the expression of unconventionality is systematically higher than for that of their (non-progressive and non-GO/COME) counterparts.
4) *Grammaticalization from a comparative perspective*. In order to demonstrate that a higher degree of grammaticalization does not necessarily lead to the complete erosion of a construction's 'unconventionality potential', we need to compare grammaticalization degrees across case studies and across languages in a methodologically sound fashion. We aim to achieve this by adapting state-of-the-art quantitative methods (Petré & Van de Velde 2018; Saavedra 2019; Sun & Saavedra 2020) that were developed specifically for this purpose.
5) *Semantic analysis*. With the aid of this extensive data analysis, we will develop a cognitive-semantic analysis in terms of non-integration with respect to the ground. This will not only allow us to account for the inherent unconventional nature of progressive/GO/COME-constructions, it will also enable us to pinpoint formal and functional features that condition the inherent predisposition for unconventional exploitation of grammatical constructions more generally.

3. Describe the methodology of the research project.

The project is to be realized as a collaborative effort between the supervisors and a qualified PhD student. It involves six Work Packages (WPs) – each associated with one or more of the research objectives, and each comprising different Milestones (MSs).

WP1. Literature study: The theory behind unconventionality – Associated Research Objective 1

In a first step, the PhD student **charts the existing analyses** of the various approaches to the concept of linguistic unconventionality. At the same time, (s)he provides an overview of studies of progressives and GO-/COME-constructions that refer to non-temporal notions in their semantic accounts, compares them and verifies to what extent they can be aligned under the umbrella of 'unconventionality'. On the basis of this study, the list of unconventionality indices developed in previous work (Petré 2017; De Wit et al. 2020) is further elaborated for usage in the subsequent WPs. This first step should lead to a review article (MS1.1),

to be sent to a peer-reviewed journal such as *Functions of Language*.

WP2: The minimal pair method in three source languages – Associated Research Objectives 2 and 3

The goal of this second WP is threefold: (i) developing the **minimal-pair method** to measure unconventionality, (ii) analyzing expressions of unconventionality in **three source languages**, i.e. English, French and Dutch, and (iii) delineating the contexts that will serve as **input** for further crosslinguistic analysis in WPs 3 and 4.

The research conducted within this WP will be based on data from the English, French and Dutch components in the Europarl corpus (Koehn 2005) and the Harry Potter corpus (van der Klis et al. 2020). The reason for working with data from translation corpora (or so-called 'parallel corpora') is that we need these corpora for our subsequent crosslinguistic analysis (see WP3 for more details on the corpora and their selection). Our choice of source languages is motivated by our advanced knowledge of these languages, as well as the fact that, in these languages, the progressive and motion verb constructions exhibit different degrees of grammaticalization (see, e.g., De Wit et al. 2020), which allows us to encompass an extended functional range of the constructions under consideration.

For each language and each construction a sample of 200 instances will be retrieved from the corpora combined. In the case of the Harry Potter corpus, this will be done in collaboration with its developers at Utrecht University. For the Dutch progressive the sample will be divided into two equal subsamples, one containing data of the *aan het* construction, the other of the posture verbs progressive. Building on Petré (2017) and suggestions in Ungerer & Hartmann (2020), each of these occurrences will be matched with an example featuring the same verb in a contrasting verbal construction (i.e. non-progressive, non-GO or non-COME), extracted from the same corpus, so as to form contrastive corpus pairs that are minimally different (as far as the corpus allows this) in terms of clause type, subject and verb choices. This way an experimental setting is approximated that minimizes confounding factors and is not distorted by instances in the corpus of non-competing use (i.e., where the alternative expression is not an option). An example is (4a) versus (4b) (extracted from the Europarl corpus):

(4) a. *If we do nothing then the courts **will go deciding**. The Watts case will be next. But I believe that Parliaments, not lawyers, should decide policy*. (EP-05-06-07)
b. *If you would like to present your oral amendment, Parliament **will decide***. (EP-00-04-13)

Since each selected instance of GO, COME or the progressive will be matched with, respectively, a non-GO, non-COME or non-progressive counterpart in each of the three languages, we will arrive at a total of 1800 minimal pairs (3600 data points) (MS2.1). In tracking down near-minimal pairs, a time-consuming but routine undertaking, the PhD will be assisted by a student for two months. Each of these pairs will subsequently be analyzed by verifying whether the progressive, GO and COME occur significantly more often with contextual features ("unconventionality indices") suggesting a higher degree of unconventionality than their equivalent expressions. These features will be split up into two types. First a set of objectively measurable features will be annotated for, including co-occurrence with focus strategies (e.g., embeddedness in cleft constructions), and formal indications of surprise at an unconventional situation such as co-occurring emphatic markers (exclamation marks, intensifiers) or interjections (Petré 2017; De Wit et al. 2020). The second set requires more qualitative analysis, and will include indications of disapproval (as in (4)), intensity, flippancy or other elements suggesting unconventionality. To minimize the risk of annotator bias, (i) the minimal pairs will be mixed and the target string replaced by the capitalized infinitive of the main verb; (ii) a substantial subsample will be annotated not only by the PhD student but also by the supervisors, and inter-annotator agreement rates such as Cohen's kappa will be used to assess the need for a revision of the annotation system. The assumption that the presence of these contextual elements is indicative of the sense of unconventionality of a potentially neutral grammatical expression rests on the principle of redundancy as a strategy of highlighting one's message (cf. Detges & Waltereit 2002). In the course of this corpus study, the list of unconventionality indices can, if necessary, be further fine-tuned and elaborated, so as to develop a comprehensive description of the characteristics of unconventional grammar use (MS2.2).

This analysis – which we will report on in a journal article to be submitted to the peer-reviewed journal *Languages in Contrast* (MS2.3) – will allow us to determine whether it is indeed correct that, in English, Dutch and French, progressives, COME and GO are significantly more frequently used for the expression of unconventionality than their counterparts. At the same time, we will end up with an extensive data set for

which we know (i) whether or not they can be categorized as unconventional and (ii) that the progressive, GO or COME can in principle occur in these contexts. After pruning out instances from different languages that are co-referential, this data set will serve as input data for WPs 3 and 4.

WP3: Analyzing unconventionality in translation corpora – Associated Research Objectives 2 & 3

If there is something inherent in the semantics of progressives and GO/COME-constructions that makes them liable to exploitation for reasons of unconventionality, we can predict that such unconventional uses are attested for these constructions **across languages**. For the crosslinguistic analysis, we will rely on multilingual translation corpora, rather than using labor-intensive and methodologically more challenging elicitation tools used in semantic typology. Multilingual translation corpora, while of course pivotal data sources in translation studies (e.g. Zanettin 2012), are also increasingly exploited in linguistic typology (cf. e.g. the dedicated special issue of *Language Typology and Universals* 60 (2007); Wälchli & Cysouw 2012; van der Klis et al. 2020). There are, admittedly, potential pitfalls to their use in this context. Translations might for instance affect degree of idiomaticity (as is notably the case for the Open Subtitles Corpus), and some corpora are built on a limited number of source/pivot languages (e.g. the Dutch Parallel Corpus), or they contain data that cannot be considered representative for contemporary natural language use (e.g. the Multilingual Bible Parallel Corpus). If selected with care, however, the advantages outweigh these weaknesses: translation corpora constitute unique and large databases "of crosslinguistically comparable expressions in a large number of diverse languages" (Wälchli & Cysouw 2012: 673-674), and they grant analysts access to natural language data that they would not be able to interpret otherwise. The first most useful corpus for the present purposes is the Europarl corpus, which contains the proceedings of parliamentary speeches in 21 languages, with as of 2019 on average 40 million tokens per language (Ustaszewski 2019). For reasons of feasibility, we will only be working with languages using Latin script, which still leaves us with 19 languages, covering the Germanic, Romance, Slavic, Baltic and Finno-Ugric families. We realize that a 'Standard Average European' bias may affect the outcome of our study (cf. Drinka 2019 on the perfect), yet this will be taken into consideration in our statistical analysis (see below). A second potential issue is that Europarl contains data from a very specific genre, which is why we are also consulting the Harry Potter corpus. Just as for WP2, the data collection and analysis will proceed identically, yet for a smaller number of languages, i.e., Spanish, Italian, and German (in addition to the three source languages).

On the basis of the source data (merged into a single data set), we know which contexts are unconventional and which are not. We can presume this holds even if we might not master all the languages represented in the sample equally well, as we assume the degree of unconventionality will not be affected by the translation process, though we also provide additional safety nets in this regard (cf. below). From among the unique contexts in our data set retrieved in WP2, we will select, for each target construction (GO/COME/progressive) and its minimal pair counterpart, the top 200 contexts meeting most criteria for being 'unconventional', and the 200 showing least of these criteria (with random sampling in case of too many ties). This way we get (for each construction type) two more or less 'prototypical' sets of conventional and unconventional contexts, a method that should facilitate testing to what extent our target constructions are more strongly associated with the unconventional ones, as it reduces noise caused by fuzzy cases. Through the platform Sketch Engine, the selected contexts can easily be aligned with their translations in the other languages and exported for further annotation. For each of the 3*400 selected contexts, we will chart their translational equivalents in the languages available in the corpus. Based on grammatical descriptions of these well-documented languages, we will first verify whether the languages under consideration have these constructions in their grammatical inventory and how these constructions fit into the larger system of the language. Next, for those languages that have a progressive and/or serial-verb GO/COME in their systems, we will check whether the translations feature a progressive/GO/COME or not. This can be done semi-automatically by means of a (e.g. regex-based) pattern recognition algorithm assigning a value to each instance, which is then manually corrected. If native speaker assistance is occasionally required (e.g., to verify whether there are no other lexical or grammatical items systematically added to the translations to reinforce the sense of unconventionality that may therefore be less prevalent for the construction under scrutiny), we will use of our academic network or post a query on a linguistics platform like the LinguistList. Thus, every [± unconventional] context in which either a progressive/GO/COME or their non-progressive/non-GO/non-COME counterparts is being used in English, French or Dutch can now be matched to its translational equivalent in the way we visualize in Table 1 for English and a limited number of contexts/target languages.

The resulting data set (MS3.1) will then be analyzed by means of a number of language-specific and aggregate statistical diagnostics to establish whether there is a significant and robust correlation between [+ unconventional] contexts and the use of the progressive, GO or COME in each of the languages under investigation (MS3.2). Statistical analyses will be performed in R, with which the PhD student will familiarize themselves through targeted PhD courses (widely available for R).

	English	**Target language 1**	**Target language 2**	**Target language 3**	**…**
Sample 1 Context 1 [+unconventional]	[+PROG]	[+PROG]	[-PROG]	[+PROG]	…
Sample 1, Context 2 [-unconventional]	[-PROG]	[-PROG]	[-PROG]	[+PROG]	…
…	….	…	…	…	…
Sample 2, Context 1 [+unconventional]	[+COME]	[+COME]	n.a.	[+COME]	…
Sample 2, Context 2 [- unconventional]	[-COME]	[+PROG]	n.a.	[-COME]	
…	….	…	…	…	…

Table 1: Translational equivalents of [± unconventional] contexts

(i) Language-specific analysis: For each target construction (i.e. GO/COME/progressive in each language if used in it) a nonparametric correlation coefficient (such as the Kendall tau-b rank coefficient (Agresti 2010: 196)) will be calculated to test how strongly unconventionality correlates with each of these constructions rather than the contrasting verbal constructions paired with them. To get clean data sets for each case, some pruning is required. We illustrate this with the GO-construction. To test its use as a marker of unconventionality in, e.g., Italian, a data set is compiled containing all instances of the Italian GO-construction within the GO-sample, plus any instances of the GO-construction found in the two other (COME & progressive) samples. To this is added all instances of contrasting verbal constructions found in the GO-sample. In each case a power analysis will determine if the sample is sufficiently robust to infer any conclusions from it (Bonett & Wright 2000). Samples not reaching this threshold will still be retained in the analysis, but dealt with separately, to ensure maximum statistical validity. The correlation coefficient of the remaining samples will be transformed into an effect size (e.g., Cohen's d) to enable comparison of association strengths across constructions and languages, while also tracking internal consistency by measuring standard deviations. Comparison between the samples from Europarl and Harry Potter will also allow us to see if the results are consistent across genres.

(ii) Aggregate analysis: To verify to what extent the semantic core in the target constructions responsible for 'unconventionality use' is shared between languages, we pool together constructions of the same type but from different languages. We also pool together data from the two corpora. The sampling method is essentially the same as that for the language-specific analysis. On this data set we perform a linear mixed-effects model (Baayen et al. 2008), with the binary split [±unconventional] as dependent variable. The main independent variable is the construction type (target construction, contrasting construction, or other construction). Other independent variables are genre, separating out genre influence as represented by Europarl vs. Harry Potter texts, and language family, tracing genetic influences. Pivot language (indicating the original language from which the other versions are translations, cf. Ustaszewski 2019) is added as a random variable, in order to chart the impact of unwanted translation effects in the case of overrepresented pivot languages. The actual language of each instance is also added as a random variable, thus acknowledging the language-specific characteristics of each target construction, be it in terms of frequency or qualitative features. Rather than treating this information as noise, by calculating Intraclass Correlation Coefficients (which assess language-internal consistency; cf. Nakagawa & Schielzeth 2013) it is also possible to more precisely assess the relative weight of language-specific behavior and shared semantic features.
The results of these quantitative analyses will submitted to an international journal in corpus linguistics such as *Corpus linguistics and linguistic theory* (MS3.3).

WP4: Stable unconventionality: Comparing degree of grammaticalization and relative unconventionality – Associated Research Objective 4.
In order to demonstrate that a higher degree of grammaticalization does not automatically lead to the

complete erosion of a construction's potential to create a sense of unconventionality, we need to **compare the degree of grammaticalization** of the progressive/GO/COME in different languages in a methodologically sound fashion. As this is a fairly labor-intensive method, requiring a combination of manual and automatic analysis of each instance, this method can only be implemented for the original English, French and Dutch samples, yet this will still yield unique insight in the role played by degree of grammaticalization (MS4.1). Petré & Van de Velde (2018) measure the degree of grammaticalization of English *be going to* based on the calculation of grammaticalization features per instance in a data sample. Some of these features, such as loss of original lexical meaning, signals of bonding and subsequent phonetic erosion, or proportion of inanimate subjects, are applicable to a wide array of constructions (beyond *be going to*), and may serve as a basis for comparison. In addition, some broader measures have been proposed, which can be automated more easily, and have been proven to be crosslinguistically valid (Saavedra 2019). Dispersion, while not uncontested, has been proposed as one such measurement. It measures to what extent the construction at issue is evenly distributed, with a more even distribution pointing towards a higher degree of grammaticalization, while more lexical constructions are expected to be more 'bursty' in their distribution (cf. Gries 2020). In addition, Sun & Saavedra (2020) showed that in particular simple token frequency and collocate diversity were good cross-linguistic indicators of degree of grammaticalization. Each of these variables can be calculated in a semi-automatic fashion on the basis of a fixed-size subsample of each language corpus, with only some light manual noise-filtering required. On the basis of each of these measurements we can calculate the degree of grammaticalization of language-specific constructions, and establish (or disprove) the lack of a systematic correlation between degree of grammaticalization and decline of 'unconventionality potential' (as measured in WP3). As the results of this innovative study will be highly relevant for the crosslinguistic study of grammaticalization, they will be submitted to a journal with a crosslinguistic orientation such as *Studies in Language* or *Linguistic Typology* (MS4.2).

WP5: Semantic analysis - Associated Research Objective 5
In this overarching **theoretical WP**, which cuts through the other WPs, our hypothesis that the progressive and GO/COME-constructions encode a lack of integration of a situation with respect to the ground will be developed in more detail for each of these construction types. We will thereby take into account the semantic specificities of each construction type (e.g. the progressive can be argued to refer to non-structural situations (Goldsmith & Woisetschlaeger 1982), while motion verbs involve a move away from what is considered standard (Stefanowitsch 1999)), without losing sight of their shared semantics. In addition, this WP also serves to (i) find explanations for potential differences between construction types or language-specific instantiations of these types in terms of unconventionality disposition (if any), (ii) integrate other (pragmatic, sociolinguistic, formal) explanations that may be complementary to our semantic proposal, and (iii) chart other constructions which have the same semantic profile as the progressive and COME/GO to be able to make larger-scale predictions about the exploitation of syntax for unconventionality. A theoretical paper on this center-piece WP will be submitted to a leading journal within the domain of cognitive and crosslinguistic semantics, such as *Cognitive Linguistics* (MS5.1).

WP6. Writing the PhD – Associated Research Objectives 1-5
Each of the high-quality journal publications written in the course of the previous WPs will form an integral part of the PhD dissertation. This final WP consists in bringing together the outcomes of WPs 1-5 and presenting them in **a well-structured and unified book-length study**. The only MS associated with this WP is the PhD defense, following the completion of the dissertation (MS6.1).

4. Bibliographical references

Agresti, A. 2010. *Analysis of Ordinal Categorical Data*, 2nd edition. John Wiley & Sons. —**Anthonissen**, L., A. **De Wit** & T. **Mortelmans**. 2019. (Inter)subjective uses of the Dutch progressive constructions. *Linguistics*. 1111–1159.—**Baayen, R., D. Davidson & D. Bates**. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59. 390–412.—**Bonett**, D. & Th. **Wright**. 2000. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika* 65. 23–28.—**Bourdin**, P. 2003. On two distinct uses of GO as a conjoined marker of evaluative modality. In R. Facchinetti et al. (eds), *Modality in contemporary English*, 103-127. De Gruyter.—**Bres**, J. & E. **Labeau**. 2013. Allez donc sortir des sentiers battus! La production de l'effet de sens extraordinaire par *aller* et *venir*. *Journal*

of French Language Studies 23. 151-177.—**Bybee,** J., R. **Perkins** & W. **Pagliuca**. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. University of Chicago Press.—**Celle**, A. & L. **Lansari**. 2015. On the mirative meaning of *aller* + infinitive compared with its equivalents in English. *Cahier Chronos* 27. 289–305.— **De Wit**, A. & R. **Vandekerckhove**. 2021. *Lache*, *giere*, *boeie.* The use of insubordinate infinitives in informal computer-mediated communication. In G. Kristiansen, *Cognitive Sociolinguistics Revisited*, 359-370. De Gruyter.—**De Wit**, A., **Petré** P. & F. **Brisard**. 2020. Standing out with the progressive. *Journal of Linguistics* 56. 479-514.—**Detges**, U. 2000. Two types of restructuring in French Creoles: A cognitive approach to the genesis of tense markers. In I. Neumann-Holzschuh & Schneider, E.W. (eds), *Degrees of restructuring in Creole languages*. Benjamins.—**Detges**, U. & R. **Waltereit**. 2002. Grammaticalization vs reanalysis: A semantic-pragmatic account of functional change in grammar. *Zeitschrift für Sprachwissenschaft* 21. 151–95.—**Devos**, M. & J. **van der Wal** (eds). 2014. *COME and GO off the beaten grammaticalization path*. De Gruyter.—**Donabédian**, A. 2012. Evidentiel et progressif: Quel statut grammatical pour la saillance prédicative? *Faits de Langues*. 62–82.—**Drinka**, B. The perfects of Eastern "Standard Average European". In A. Danylenko & M. Nomachi (eds), *Slavic on the Language Map of Europe.*— **Eitelmann**, M & D. **Haumann**. 2019. Extravagant Morphology. SLE workshop.—**Franckel**, J.-J. 1989. *Etude de quelques marqueurs aspectuels du français*. Droz.—**Givón**, T. 2001. *Syntax*. Benjamins.—**Goldsmith**, J. & E.F. **Woisetschlaeger.** 1982. The logic of the English progressive. *Linguistic Inquiry* 13. 79-89.—**Gries**, S. 2020. Analyzing dispersion. In M. Paquot & S. Gries (eds.), *Practical handbook of corpus linguistics*, 99–118. Springer.—**Güldemann**, T. 2003. Present progressive vis-à-vis predication focus in Bantu. *Studies in Language* 27. 323-360.—**Haiman**, J. 2014. Decorative morphology in Khmer. In J.P. Williams (ed.), *The aesthetics of grammar,* 61–82. CUP.—**Haspelmath**, M. 1999. Why is grammaticalization irreversible? *Linguistics* 37. 1043–1068.—**Kempf, L. & S. Hartmann**. 2018. Schema unification and morphological productivity: A diachronic perspective. In G. Booij (ed.), *The Construction of Words*, 441–474. Springer.—**Koehn**, P. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation, MT Summit.—**Langacker**, R.W. 2008. *Cognitive Grammar: A Basic Introduction*. OUP.—**Leech**, G. 2004. *Meaning and the English Verb*. Routledge.—**Lemmens**, M. 2005. Aspectual posture verb constructions in Dutch. *Journal of Germanic Linguistics* 17. 183–217.—**Levy**, R. & T.F. **Jaeger**. 2007. Speakers optimize information density through syntactic reduction. In B. Schlökopf et al. (eds), *Advances in neural information processing systems* 19, 849–56. MIT Press.—**Nakagawa, S. & H. Schielzeth**. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4. 133–142.—**Norde**, M. & C. **Morris**. 2018. Derivation without category change: a network-based analysis of diminutive prefixoids in Dutch. In: K. Van Goethem et al. (eds), *Category change from a constructional perspective*, 47-90. Benjamins.—**Petré**, P. 2017. The extravagant and the progressive: An experimental corpus study on the grammaticalization history of [*be* V*ing*]. *English Language and Linguistics* 21. 227–250.— **Petré**, P. & F. **Van de Velde**. 2018. The real-time dynamics of the individual and the community in grammaticalization. *Language* 94. 867–901.—**Saavedra**, D.C. 2019. Measurements of Grammaticalization: Developing a quantitative index for the study of grammatical change. Antwerp/Neuchâtel: Unpublished PhD thesis—**Schmid**, H-J & F. **Günther**. 2016. Toward a unified socio-cognitive framework for salience in language. *Front. Psychol.* 7, 1110—**Simon**, H.J. & H. **Wiese** (eds). 2011. *Expecting the unexpected: Exceptions in grammar*. De Gruyter—**Stefanowitsch**, A. 1999. The go-and-verb construction in a cross-linguistic perspective: Image-schema blending and the construal of events. In D. Nordquist and C. Berkenfield (eds): *Proceedings of the Second Annual High Desert Linguistics Society Conference*, 123–134.—**Sun**, L. & D.C. **Saavedra**. 2020. Measuring grammatical status in Chinese through quantitative corpus analysis. *Corpora* 15. 317–342.—**Traugott**, E. & G. **Trousdale**. 2013. *Constructionalization and Constructional Changes*. OUP—**Ungerer**, T. & S. **Hartmann**. 2020. Delineating extravagance. *Belgian Journal of Linguistics* 34. 352–364.—**Ustaszewski, M**. 2019. Optimising the Europarl corpus for translation studies with the EuroparlExtract toolkit. *Perspectives* 27. 107–123.—**van der Klis**, M., B. **Le Bruyn** & H. **de Swart.** 2020. De la sémantique des temps verbaux à la traductologie: une comparaison multilingue de L'Étranger de Camus. In E. Corre et al. (eds.) *Linguistic approaches to Tense, Aspect, Modality, Evidentiality, based on the Novel L'Etranger ("The Stranger") by Albert Camus, and its Translations*. Benjamins—**Visser**, F. Th. 1969. *An Historical Syntax of the English Language. Part three, first half: Syntactical units with two verbs*. Brill—**Wälchli**, B. & M. **Cysouw**. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50. 671–710—**Zanettin,** F. 2012. *Translation-driven corpora: Corpus resources for descriptive and applied translation studies*. St. Jerome.