Time in translation: a semantics of the perfect

Project members: Prof. dr. Henriëtte de Swart (PI), dr. Bert Le Bruyn (PI), Martijn van der Klis (PhD1), NN (PhD2), NN (post-doc).

Advisors: Prof. dr. Nicholas Asher (Toulouse), Prof. dr. Bernhard Wälchli (Stockholm), dr. Joost Zwarts (Utrecht).

Funding: NWO Free Competition Programme.

Period: June 1, 2017 until September 30, 2021.

Institutional setting: UiL OTS, Department of Languages, Literature & Communication, Utrecht University, the Netherlands.

Key Words: tense, aspect, semantics, pragmatics, translation, corpus

Summary
The aim of the project is to analyze the meaning of the perfect. This sounds easy, but it is not. Verb forms like English ‘have left’ and their counterparts in other languages combine reference to the past and the present, in ways that have escaped linguistic theory. To get out of the doldrums, we need data that enable cross-linguistic comparison without pre-conceptualized meanings. We obtain those data through a new methodology for multilingual corpus research that we dub ‘Translation Mining’. As the perfect interacts with the grammatical categories past and present, its analysis requires a competition-based approach. Our corpus data show the alternations, so we improve on current analyses by developing a micro-typology that is grounded in translation equivalences between English, Dutch, German, French and Spanish. Perfects vary not only in sentence-level meanings, but also in narrative and non-narrative discourse use. As a theory of temporal structure in non-narrative discourse is lacking, we build a temporal extension of commitment-based dynamic theory to analyze the corpus data of perfects in context. The predictions of a corpus-based analysis need to be checked on an additional corpus and squared with native speakers, so we carry out a corpus experiment and native speaker surveys. The outcome of the programme will be (1) a semantic map of the perfect, (2) a dynamic account of its non-narrative nature, and (3) a unified truth-conditional and dynamic semantics of its sentence-level and discourse-level meanings. The results will be made available for translators, computational linguists and typologists.

1. General description of the Proposed Research

1.1 Main aim and delimitation

Many languages encode information on the location of a situation in time (tense) and its internal make-up (aspect) in verbal morphology (overview in Binnick 2012). The analysis of these verb forms is the domain of tense-aspect theories. Our main aim is to provide a semantics of the PERFECT. Notwithstanding the extensive literature on the PERFECT, this goal has not been reached (Ritz 2012). Section 1.2 explains why. Section 1.3 shows how we overcome the limitations of current analyses. But first, we delimit the research.
In English, the **PERFECT** involves *have* + past participle. Its meaning has a ‘past’ component (event in the past), as well as a ‘present’ component (current relevance/result state holding now). (1) gives examples with their usual labels (Portner 2003):

\[(1) \quad \begin{array}{ll}
\text{a. Mary has visited Paris. (past visit relevant now)} & \text{[experiential perfect]} \\
\text{b. Mary has moved to Paris. (currently living in Paris)} & \text{[resultative perfect]}
\end{array}\]

Ideally, a linguistic analysis of the PERFECT accounts for the distribution of PERFECT forms and meanings, not only in English, but also in other languages. The problem is that the PERFECT is a diachronically and linguistically unstable category (Lindstedt 2000) and is subject to widespread cross-linguistic variation. We zoom in on the *have* PERFECT that Dahl & Velupillai (2013) trace back to a transitive possessive construction, and manifests itself in languages like English, Dutch, German, French and Spanish. The grammars of these languages are closely related, yet their tense-aspect systems are fundamentally different (presence/absence of progressive/non-progressive, perfective/imperfective). This guarantees sufficient room for variation. This proposal uses small caps (PAST/PRESENT/PERFECT) to refer to the abstract form, independently of the language.

### 1.2 Problem statement and central claim

**Existing analyses** come in two flavours: descriptive approaches with an emphasis on typological distribution (Dahl & Velupillai 2013) and theoretical approaches with an in-depth formal analysis of one, or just a few languages (Portner 2003 for English, Musan 2002 for German). The two approaches share the following three problems:

**Problem (i):** Theoretical and typological approaches are driven by sets of predefined usages, illustrated by prototypical examples like (1). Now take the French Passé Composé (PC). According to Lindstedt (2000), it does not qualify as a PERFECT, because it has a narrative use, and narration is characteristic of the category of the PAST. The English counterpart of (2a) in (2b) does indeed require the Simple Past:

\[(2) \quad \begin{array}{ll}
\text{a. Quand je l’ai vu (PC), il m’a dit (PC) bonjour.} & \text{[narrative perfect]} \\
\text{b. When I saw/*have seen him, he said/*has said hello.}
\end{array}\]

However, the Passé Composé is formed by *have* (or *be*) + past participle, and can be used in the contexts in (1), just like the Present Perfect, so the two verb forms share core formal and distributional properties. Predefined meanings suggest we know what the necessary and sufficient conditions are of the PERFECT, but for many concepts it is impossible to provide such a definition (Wittgenstein 1968). We have to capture the family resemblances between the French Passé Composé and the English Present Perfect for a full-fledged semantics of the PERFECT, but in order to grasp them, we need to broaden the empirical scope of the debate.

**Problem (ii):** Everyone agrees that the PERFECT combines PAST and PRESENT meaning components, yet neither theoretical nor typological approaches investigate the interactions of the PERFECT with other verb forms. A combined feature search on WALS shows that some languages have a PERFECT as well as a past/present tense, and a perfective/imperfective contrast, while others don’t have all or any of these forms. In a Saussurean perspective in
which *la grammaire est un système où tout se tient*, the presence or absence of other tense-aspect categories is likely to have an impact on the meaning and distribution of the PERFECT. In Ritz’s (2012) words, the PERFECT is the ‘shapeshifter’ of tense-aspect categories, that adapts its meaning to fit into a given system. The contrasts in (2) and (4, 5) (below) support the need for a competition-based analysis of PERFECTS in context.

**Problem (iii):** Typological and theoretical approaches focus on sentence-level meanings, and exclude the narrative perfect in (2a) as a contradiction in terms. Building on Partee (1984), Kamp & Reyle (1993), Lascarides & Asher (1993), studies on temporal discourse structure have focused on the narrative tense use associated with the PAST in (2b). But if the PERFECT is not narrative, what is its role in discourse? Based on *Switchboard* examples like (3), Nishiyama & Koenig (2010) show that the English Present Perfect can be used to introduce a new topic:

(3) A: *Have you seen* ‘Dancing with wolves’?
    B: Yeah, *I’ve seen* that, that was a really good movie.

However, we lack data on other languages, and the focus of Nishiyama and Koenig is not on building a theory of non-narrative discourse.

The central claim of this programme is that the investigation of time in translation overcomes the empirical limitations of existing analyses, and provides the foundation for a data-driven, competition-based semantics of the PERFECT that integrates the sentence- and discourse-level.

**1.3 Approach and innovation**

How does the investigation of time in translation help us to develop a semantics of the PERFECT? We proceed in four steps. The first step is to get empirical data enabling us to compare PERFECTS across languages. This constitutes the primary motivation for working with translation equivalents: translations provide us with form variation across languages in contexts where the meaning is arguably stable.

Manual research shows this is a promising strategy. The literary example in (4) illustrates how an English PERFECT appears when the state following a change holds in the present, even if the French source text expresses this with a PRESENT, maintained in the Dutch translation:

(4) a. Rien ne bouge. [PRESENT] *No et moi*, Delphine de Vigan, 2007 [French]
    b. Nothing has moved. [PERFECT] *No and me*, translation 2010 [English]

Such translation data can be used to test claims in the literature. The pattern in (4) supports Nishiyama & Koenig’s claim that the French Passé Composé lacks the continuative PERFECT reading we find in English, and extends it to Dutch.

A systematic way to collect data on variation is to use multilingual corpora. In a multilingual corpus, no language occupies a privileged position, so we also detect PERFECTS that are not in the source language like (4b). The resultative meaning constitutes the core of the PERFECT (cf.
1b), but (5) (from the subtitle corpus) shows that it can also be conveyed by the English Simple Past:

(5) a. In case you hadn’t noticed, we just got a confession. [PAST] [English original]

b. Falls es ihnen entging, er hat gestanden. [PERFECT] [German]

c. Si vous ne l’avez pas remarqué, on a des aveux. [PRESENT] [French]

The same meaning of a past event and a result with current relevance can be conveyed by a PAST, PERFECT or PRESENT. Note that pairwise comparison of languages would not have generated the English-French pair (5a/c), because no PERFECT is involved (cf. also 4a/c). So multilingual corpus research is crucial to obtain enough variation.

The second step involves replacing the pre-defined set of meanings by those extracted from PERFECT use in various languages. We need more data to build a usage-based theory, so we scale up the translation approach. We create large sets of multilingual tuples like (5) from different source languages through a new methodology we call ‘Translation Mining’ (technical details, examples, and early results elsewhere on the website). In this way, we extract PERFECT verb forms in all languages, and place them in a database with their matched translations in all other languages. Items in the database are 5-tuples of verb forms from five languages with at least one PERFECT, e.g <Pretérito_Perfecto, Voltooid_Tegenwoordige_Tijd, Passé_Composé, Simple_Past, Präteritum>. Wälchli & Cysouw (2012) provide us with a technique for quantitative analysis of data from multilingual corpora to draw semantic maps. Based on a Europarl dataset, de Swart et al. (2016) show that PERFECTS from the five languages under investigation are not related by simple subset/superset relations, but occupy distinct regions on the map. As the PERFECT maps are grounded in actual language use, there is no need for pre-conceptualized meanings, so this approach overcomes Problem (i).

In a third step, we develop a data-driven semantic model that integrates sentence-level and discourse-level meanings. Each 5-tuple in the database is visible as a point on the map. Clicking on a point connects us to the original example and its four translations, so we can explore the map and zoom in on the differences between examples that do or do not lead to a PERFECT in translation. The enriched dataset provides us with the empirical foundation for the competition-based approach we need to overcome Problem (ii).

As the Translation Mining method extracts PERFECT data in context, we can investigate tense use at both the sentence and the discourse level to overcome Problem (iii). Pilot corpus data confirm the forward-looking behaviour of the Dutch PERFECT as introducing a new topic in (6), similar to the English example in (3):

(6) A: Jongens, kom gauw kijken! Ik heb een babypanda gevonden!
B: Wauw, wat een schatje!

A: Guys, quickly, come look! I have found a baby panda!
B: Wow, what a cutie!
Taking our inspiration from Farkas & Bruce (2011), we propose that the embedding of the Present Perfect in (6A) raises an issue for future discourse. Subsequent discourse elaborates the event using other tense-aspect forms. Interestingly, the PERFECT can also be backward-looking, as in (7):

(7)  A: Hoe ging het?  A: How did it go?
B: Het is gelukt!  B: It has succeeded! (≈ we managed)

Unlike the anaphoric PAST in (2b), the backward-looking PERFECT in (7) is not narrative, but closes off the issue under discussion. If PERFECTS are used at the ‘edges’ of discourse sequences, where they link a past event from or to the utterance situation, we need to place special emphasis on the beginning and end of discourse sequences. Note also that we should not only be concerned with formal dialogue like in Europarl, but investigate informal conversation and literary texts as well.

Non-narrative PERFECTS are found across the board, but not all languages restrict tense use at discourse edges to the PERFECT, as the English Simple Past in (5a) shows. The unexpected non-narrative discourse use of the English Simple Past in (5) is the mirror image of the equally unexpected narrative discourse use of the French Passé Composé in (2a), so a comparative perspective is not only needed at the sentence-level, but at the discourse-level as well.

The fourth step closes the argumentative cycle by testing the model’s predictions against new corpus data and native speaker judgments.

2. Operationalization in three subprojects

We operationalize the research programme in three subprojects addressing the distribution of PERFECT, PAST and PRESENT forms and meanings across text types (PhD1), analysing the discourse role of the PERFECT in conversation and literary texts (PhD2), and proposing a compositional and dynamic semantics of the PERFECT (Post-Doc). PI and co-applicant synthesize the outcomes in interdisciplinary research that reaches out to translation studies, computational linguistics and typology.

2.1 Subproject 1 (PhD): A semantic map of the PERFECT

Subproject 1 focuses on the interaction of PERFECT meanings with sentence-internal features (verb classes, adverbials) and register (text types), and draws a semantic map of the PERFECT. A semantic map is a geographic layout that graphically represents how meanings of grammatical functions are related to each other (Haspelmath 1997, Aloni et al. 2010, Zwarts 2010). As the distance between points on the map reflects the degree of (dis)similarity between n-tuples, we need enough data points in translation equivalent contexts to generate the right amount of cross-linguistic variation.

In a first stage, we expand our current data set, and apply Translation Mining to 5,000 contexts in which at least one of the languages of our sample uses a PERFECT, a total of 25,000 individual language data points. We use the Europarl corpus to extract PERFECTS from
formal dialogue, and the subtitle corpora for informal conversation data. The DCEP corpus (the Digital Corpus of the European Parliament, including press releases, minutes, reports) provides us with informative data, and the Dutch Parallel Corpus (Dutch/English/French) as well as the Oslo Multilingual Corpus (Norwegian/English/French/German/Dutch) with data from literary texts. Different text types come with their own vocabulary and discourse configurations, so the wide dataset will reveal how much this variation affects the perfect grammar. For the Europarl and the DCEP corpus we include two further Germanic and two further Romance languages that allow us to further increase the precision of our semantic maps.

In a second stage, automatically extracted information concerning aspectual class and combinatorial restrictions on adverbs will be used to relate variation in perfect meaning to sentence-internal factors. De Swart (2014) shows that the aspectually sensitive adverb all happily combines with the Dutch perfect (8a), but not with the past (8b). In contrast, its English counterpart already is found with both perfect and past, as shown in (8c,d).

(8)  
c. He has already taken a bath.  d. He already took a bath.

Such contrasts are surprising, because the literature frequently uses adverbs to distinguish the main readings (narrative, continuative, hodiernal, experiential and resultative) (Portner 2003, Schaden 2009, Kamp & Rossdeutscher 2013, references therein). The translation equivalents provide us with stable meanings across languages, so we can put the criteria to the test, and fine-tune them wherever necessary. Obviously, we hope that our data driven approach will also lead to the discovery of new grammatical restrictions that went undetected under a pre-conceptualized meaning approach.

Beyond adverbs, we exploit the corpus data to test hypotheses from the literature about the competition between perfect and present/past meanings. For instance, the German past in (9c) is unexpected under Schaden’s (2009) claim that this tense form is incompatible with current reference, unlike its English counterpart in (9a):

(9)  
a. She’s the one who found the body and called it in. [English original]  
b. Zij heeft het gemeld. [Dutch]  
c. Sie fand die Leiche und gab Bescheid. [German]

The past fand in (9c) contrasts with the perfect hat gestanden in (5b) above. On the basis of just two examples, we cannot determine whether we are dealing with a translation error, free variation, or subtle differences in contextual meaning. Larger datasets should enable us to statistically wipe out occasional exceptions. Even so, a usage-based theory like ours needs to be complemented with corpus experiments that allow us to validate the findings based on our reference corpus on a new corpus. As we make claims about linguistic competence, the model’s predictions also need to be squared with native speaker intuitions. So we move beyond given data, and move from description and theory to prediction and experiments, a unique move in data-driven approaches.
In the third stage of the project we check the predictions our analysis makes on the basis of a new sample of 2000 contexts from the translation corpora to which we add the required annotations. We further manipulate the original and translations in examples like (1)-(8), and run surveys with native speakers to judge the acceptability of competing tense-aspect forms, their combination with deictic expressions and aspectually sensitive adverbs, and possibly other interactions that emerged from the data. Through the Digital Humanities Lab, we have access to internet-based survey platforms like Crowdflower and Lime Survey that allow us to run large scale surveys in a time- and cost-effective way.

The results of Subproject 1 are crucial to extend and stabilize the empirical coverage of the different labels for PERFECT meanings across languages. Thus the PERFECT map constitutes the starting point for a broader typological investigation complementing Dahl (2015).

2.2 Subproject 2 (PhD): Perfect use in discourse: from conversation to literary texts

Subproject 2 deals with the discourse conditions on PERFECT use. Theories on the temporal structure of discourse have focused on the narrative PAST in (2b) as crucial for text coherence (Partee 1984, Kamp & Reyle 1993). In contrast, the PERFECT in examples (3), (6), (7) serves to put an issue on the table or close it off. At the ‘edges’ of a discourse (beginning or end), the PERFECT relates a past event to or from the utterance time. This non-narrative discourse use is not restricted to the PERFECT (cf. the English PAST in 5a), and it doesn’t stop the PERFECT from also having a narrative use (cf. the French PERFECT in 2a). Subproject 2 aims at an empirical description and a semantic-pragmatic analysis of tense-aspect forms in dialogue and conversation, where we expect a mixture of narrative and non-narrative discourse.

In the first stage, the empirical description is grounded in PERFECT examples extracted from Europarl and the subtitle corpus. PERFECTS in questions and responses, tense use and negation in previous and follow-up discourse moves will be investigated, as well as descriptive and referentiality features and location in time/space as dimensions of the event to be elaborated.

In a second stage, PhD 2 will develop an account of the non-narrative discourse use of the PERFECT, combining insights from conversation analysis (Sidnell & Stivers 2012), inquisitive semantics (Ciardelli et al. 2013) and commitment-based discourse theories (Farkas & Bruce 2011). In recent literature, polarity particles (Roelofsen & Farkas 2015), wh-exclamatives (Chernilovskaya 2014), specific this indefinites (Prince 1981, Ionin 2006, Onea 2016) and sluicing (John ate something, but I don’t know what, Anderbois 2011, 2014) have been analysed as grammatical markers that operate dynamically at the semantics-pragmatics interface. So far, the phenomena addressed have involved propositional-level information and individual discourse referents. However, Anderbois (2014) suggests that the implicit Davidsonian event variable may also have inquisitive content, because it supports sluicing. A preliminary corpus study of English, Dutch, French and German with a temporal sluice (when) yields many examples like (10), which supports the need for an inquisitive treatment of the PERFECT:

(10) a. Ja, ich habe ihn gesehen, aber ich weiß nicht mehr wann. [German]
   Yes, I have him seen, but I know not anymore when

7
English is exceptional, in that we also find PAST forms in the antecedent of the sluice:

(11) Somebody sent it to me way back then, but I don't remember when or who.

(11) suggests that the English PAST is not only a narrative anaphoric tense (Partee 1984), but also has cataphoric (forward-looking) inquisitive features. A mixed anaphoric/cataphoric dynamic semantics should also be attributed to German and French PERFECTS because of their narrative use (cf. 2). In view of the differences in inquisitive content languages assign to PAST and PERFECT verb forms, the dynamic semantics of the PERFECT needs to be set up in such a way that it accounts for cross-linguistic variation.

In the third stage, PhD 2 will analyse literary examples from the Dutch Parallel Corpus and the Oslo Multilingual Corpus. Insofar as tense shift in conversational narrative shares features with conversation (Fludernik 2012), we expect the pragmatic analysis to extend to the literary data. However, in this stage we are also after the literary effects of the topic-shifting role of the PERFECT, as we find it for instance in the way we are drawn into the inner thoughts of the protagonist (12):

(12) a. C’est pour ça que j’ai pris ma décision: à la fin de cette année scolaire, le jour de mes treize ans, le 16 juin prochain, je me suiciderai. [M. Barbery, L’élégance du hérisson]
   b. That is why I’ve made up my mind: at the end of the school year, on the day I turn thirteen, the sixteenth of June, I will commit suicide. [The elegance of the hedgehog]

The forward-looking PERFECT in (12) requires the translator to connect the current utterance to the upcoming narrative structure. We need to move beyond corpus data to answer the question why the translator did not choose a Simple Past in (12), given its forward-looking use in (11); compare also German (5b) and (9c). We will systematically manipulate tense-aspect forms in original texts and their translations to evaluate their linguistic and literary qualities. Crowdsourcing techniques (cf. PhD project 1) will be exploited for surveys that target the general public, alongside questionnaires submitted to advanced translation students and professional translators.

In sum, subproject 2 extends existing insights on anaphoric tense use by analysing the way the PERFECT opens up a new issue, or closes it. The commitment-based dynamic semantic-pragmatics will provide new temporal discourse features for computational linguistics to use in discourse and dialogue annotation.

2.3 Subproject 3 (Post-Doc): Truth-conditional and dynamic semantics of the PERFECT

Subprojects 1 and 2 provide a micro-typology of the PERFECT at the sentence and discourse level. The post-doc is the theoretical gamemaker of the project, and investigates new questions the detailed investigation of translation equivalents raises for linguistic theory:

- Semantic maps visualize implicational meaning hierarchies; they tell us how the different meanings of the PERFECT hang together. In formal semantics, these hierarchies have hardly been studied. How can they be used as building blocks for a compositional semantics of the PERFECT?
- An inquisitive discourse semantics of the PERFECT accounts for its cataphoric potential. But in some languages, e.g. English, the PAST has an inquisitive semantics as well. What are the implications of cross-linguistic variation for dynamic semantic theory?

- The PERFECT is considered to be one of the hardest tense-aspect forms to analyze at the sentential level, and yet we add another level, \textit{viz.} discourse. What – if anything – links the two?

\textit{Implicational hierarchies}

Typologists and historical linguists have worked on the way different meanings of the PERFECT are connected (Lindstedt 2000, de Acosta 2011). Formal semanticists have avoided the question or contended themselves with pragmatic principles that derive all readings from a single one, ignoring subpatterns (Musan 2002). Inspired by the work on the semantic map of indefinites by Aloni et al. (2010), and work in diachrony and typology, the Post-Doc will work out the compositional semantic underpinnings of PhD1’s map of the PERFECT.

\textit{Inquisitiveness}

Inquisitive semantics offers a promising framework for the analysis of cataphoric tense use. There is room for improvement though, given the variation in inquisitive strength between the PAST in (11) and the PERFECT in (10). A similar instance of variation in inquisitive strength is found between regular indefinites and \textit{this} indefinites. Corpus work by Prince (1981) shows that referents introduced by indefinite \textit{this} (13a) are more likely to be picked up by anaphoric pronouns in subsequent discourse than indefinites introduced by \textit{a} (13b). The discourse in (13a) feels incomplete without the second sentence whereas (13b) does not:

(13)  
a. There was this pink truck. It was the most amazing thing I had ever seen.  
b. There was a pink truck.

If \textit{a} and \textit{this} have the same inquisitive content (Onea 2016), this difference remains a mystery. The problem is mirrored in the temporal domain: both tense forms arguably imply existential quantification over a Davidsonian event argument, yet the PERFECT (3, 6, 10) is more generally used than the PAST (11) to introduce a new topic.

The solution the Post-Doc will explore is that \textit{this} and the PERFECT are not simply inquisitive expressions but come with the additional semantic requirement to refer to another individual/eventuality in the discourse. The extra referentiality requirement is not constrained by directionality: the demonstrative can either refer forward (indefinite \textit{this} in 13a), or refer back (definite \textit{this} in 14):

(14)  
A man was walking in the park. Everyone else looked depressed but \textbf{this man} looked happy.

Clearly, this mirrors the forward/backward looking PERFECTS at the edges of a discourse sequence (beginning or end). As the need for a more fine-grained dynamic semantics is not
unique to the temporal domain, a unified analysis of *this* and the PERFECT constitutes a relevant contribution to inquisitive semantics.

*From sentence to discourse*

Compositional analyses of the PERFECT typically reduce the role of the auxiliary to that of tense (Pancheva & von Stechow 2004). But this eliminates the possibility of insightfully linking its sentential and discourse semantics. Sæbø (2009) proposes an alternative in which perfect HAVE is the same as the relational HAVE we find in ‘*John has a brother*’. The role of HAVE here is not merely to introduce tense but ensures that the relation included in brother takes John as its argument, leading to the reading that ‘There is an individual who is a brother of John’. Building on Sæbø, we argue that the same semantics underlies both the sentence and discourse use of perfect HAVE. On its sentence use, HAVE relates the subject to the eventuality included in the past participle. On its discourse use, HAVE relates eventualities across sentences. The dynamic analysis of HAVE proposed by Le Bruyn, de Swart & Zwarts (2016), formalizes the relational semantics of HAVE as involving anaphors/cataphors. In the same way as demonstrative THIS, perfect HAVE can thus be seen as coming with the requirement of referring back or forward. This analysis will be extended to perfect BE along the lines set out in the literature on BE and HAVE languages (Isačenko 1974).

In sum, the Post-Doc will develop a unified compositional and dynamic semantics of the perfect. In doing so, s/he makes sure that the translation findings of the PhDs are established as new landmarks in current linguistic theory.

*Selected references*


