

Temporal reference in discourse and dialogue

Martijn van der Klis, Bert Le Bruyn and Henriëtte de Swart¹

Abstract

Stories are traditionally told in the past tense, but in naturalistic conversation we often refer to the present and future next to the past. This contrast between discourse and dialogue can be used to gain insight into the semantics of the PRESENT PERFECT, a verb form that combines reference to a past event with current relevance. PERFECT use varies across Romance and Germanic languages, and we rely on a parallel corpus of translations of J.K. Rowlings' *Harry Potter and the Philosopher's Stone* to bring out the cross-linguistic variation. We take the dialogue parts in the novel (in which the characters talk among each other) as a proxy for naturalistic conversation, and compare tense use in dialogue to tense use in narrative discourse (the parts of the novel in which the narrator tells the story). We showcase a new computational methodology to automatically chart variation between six languages on semantic maps. We report on two main findings. First, languages generally do not differ in tense use in narrative discourse, and the PERFECT is completely absent. Second, in dialogue, we find competition in languages between the PRESENT, PERFECT, and PERFECTIVE PAST. We find that languages differ in the way they carve up the semantic space: French, Italian, German, and Dutch have a PERFECT that extends into the past domain, while the English PERFECT extends into the present domain.

Keywords: tense, aspect, PERFECT, discourse, dialogue, semantic maps.

1. Introduction

1.1 Stability and cross-linguistic variation in the PRESENT PERFECT

Romance and Germanic languages share a grammaticalized PRESENT and PAST tense (e.g. the *simple present* and *simple past* in English, and the *présent* and *passé simple* in French).² Furthermore, they share a construction that morphologically consists of a present form of the auxiliary HAVE/BE and a past participle. This HAVE-PERFECT, as Dahl & Velupillai call it, is exemplified in (1).

- | | |
|---|--------------------------------------|
| (1) I – I've won the House Cup for Gryffindor. | [English, <i>present perfect</i>] |
| Ich... ich hab den Hauspokal für Gryffindor gewonnen . | [German, <i>Perfekt</i>] |
| Ik – ik heb de afdelingsbeker veroverd voor Griffoendor. | [Dutch, <i>vtt</i>] |
| J'ai... j'ai fait gagner la coupe à Gryffondor. | [French, <i>passé composé</i>] |
| Io... ho appena fatto vincere a Grifondoro la coppa del campionato. | [Italian, <i>passato prossimo</i>] |
| Yo... he ganado la copa de la casa para Gryffindor. | [Spanish, <i>pret. perf. comp.</i>] |

We refer to the grammatical category in (1) as the (PRESENT) PERFECT. The PERFECT is usually taken to refer to a past event with current relevance (Portner 2003, Ritz 2012). This leads to a division of labour between the

¹ The order of authors is alphabetical. Martijn has been responsible for the data collection and has compiled the paper after extended group discussions. Bert and Henriëtte have proposed the original idea of looking into literary corpora for variation in tense use and have provided extensive feedback on earlier drafts.

² Throughout this paper, we use small capitals to signal form-based categories (e.g. PERFECT) and italics for language-specific instantiations of these categories (e.g. the English *present perfect*).

compound PERFECT verb form and the simple PRESENT and PAST tenses when it comes to temporal reference to events and states in the present and the past. This division of labour does not always play out in the same way across languages. We find PERFECTS across the board in the context of (1), but cross-linguistic variation in PERFECT use is exemplified in (2) and (3) below. In this paper, we are interested in stability and cross-linguistic variation in Germanic and Romance languages and aim to provide a micro-typology of six languages: three Germanic (English, German, and Dutch) and three Romance (French, Italian, and Spanish). The examples all come from J.K. Rowling's *Harry Potter and the Philosopher's Stone* and its official translations. More on the corpus below.

From the overall characterization of the PERFECT as referring to a past event with current relevance, we can take the cross-linguistic similarities in (1) for granted. However, we also find variation in the distribution of the PERFECT across Romance and Germanic languages. For example, it is well-known that the English *present perfect* does not combine with past time adverbials like *yesterday* and uses a *simple past* instead, as illustrated in (2). However, all other languages except Spanish allow a PERFECT in this case.

- | | |
|--|---|
| (2) Dumbledore gave me the day off yesterday ter fix it. | [English, <i>simple past</i>] |
| Dumbledore hat mir gestern dafür freigegeben . | [German, <i>Perfekt</i>] |
| Perkamentus heb me gisteren vrijaf gegeven om 't te regelen. | [Dutch, <i>vtt</i>] |
| Dumbledore m'a accordé un jour de congé hier pour le préparer. | [French, <i>passé composé</i>] |
| Ieri Silente mi ha dato una giornata di libertà per fabbricarlo. | [Italian, <i>passato prossimo</i>] |
| Dumbledore me dio libre el día de ayer para hacerlo. | [Spanish, <i>pretérito indefinido</i>] |

On the other hand, as (3) shows, when states continue to be the case into the present, English uses the *present perfect* in circumstances in which all other languages require a PRESENT.

- | | |
|---|------------------------------------|
| (3) How long have I been in here? | [English, <i>present perfect</i>] |
| Wie lange bin ich schon hier? | [German, <i>Präsens</i>] |
| Hoe lang lig ik hier al? | [Dutch, <i>ott</i>] |
| Ça fait combien de temps que je suis là? | [French, <i>présent</i>] |
| Da quanto tempo sono qui? | [Italian, <i>presente</i>] |
| ¿Cuánto tiempo hace que estoy aquí? | [Spanish, <i>presente</i>] |

The kind of cross-linguistic competition with the PRESENT on the one hand, and with the PAST on the other has been reported and modelled by various authors (Bybee, Perkins, & Pagliuca, 1994; de Swart, 2007; Rotstein, 2008; Schaden, 2009; Ritz, 2012; Dahl & Velupillai, 2013; van der Klis, Le Bruyn, & de Swart, 2019). The linguistic and typological literature mainly focuses on the cases in (2), where the PERFECT extends its use to refer to an event in the past, and thereby intrudes into the domain of the PAST tense. Interestingly, the competition between PAST and PERFECT in Romance is linked to the distinction between written and spoken language.

1.2 The role of the PERFECT in conversation

In French, a common finding is that the *passé composé* has replaced the *passé simple* in conversation (Engel, 1990; Gadet, 1999). The *passé composé* also appears in certain written texts, such as sports reports and obituaries (Labeau, 2015), but in storytelling, the *passé simple* remains the preferred tense form to move the narration forward. Much recent work on temporal reference in narrative discourse builds on Kamp & Rohrer (1983), who show that the *passé simple* introduces the main events, and the *imparfait* describes states in the background. This preference for the *passé simple* is not specific to literary fiction: the *passé simple* also appears in detectives and fairy tales. For Italian, we find similar claims in the literature. The *passato remoto* has almost disappeared from conversational use, although regional variation persists (Rohlf, 1966; Lepschy

& Lepschy, 1981; Harris, 1982). Novels are taken to reflect the contrast between written and spoken language in the distinction between narrative discourse and narration, where the narrative parts imply the *passato remoto*, and the *passato prossimo* appears in direct quotes (Bertinetto, 1986; Centineo, 2016). To the best of our knowledge, no claims for different ways of temporal reference in written and spoken language have been made with respect to Spanish.

The link between extended PERFECT use and the written/spoken language distinction is also present in the Germanic literature, but the main emphasis here is on regional variation. The *Perfekt* is the default verb form to refer to events in the past in the southern varieties of German, where the *Präteritum* is mainly used with a restricted set of verbs, including copula verbs and modals (Hennig, 2000). However, Löbner (2002) claims that in standard present-day colloquial German, the *Perfekt* can be used in the main clauses of narrative passages in the same way as the *Präteritum*.

The literature on the PERFECT in Dutch and English does not mention a connection with spoken language. However, previous research on the semantics and pragmatics of the English *present perfect* already reveals its sensitivity to the interaction between speaker and hearer in conversation. According to Portner (2003), the English *present perfect* is used to answer the question under discussion. For example, in (4), a *present perfect* answers the question of who to ask for an explanation of George Eliot's style.

- (4) A: We need to get an explanation of George Eliot's style. Who can we ask?
B: Mary **has read** *Middlemarch*.

Portner argues that B's response is part of a more elaborate Gricean reasoning: George Eliot wrote *Middlemarch*, and if someone reads an author's book, they understand their style, and hence Mary, having read *Middlemarch*, is the right person to ask. The *present perfect* that B uses points at a specific consequence of Mary's reading of *Middlemarch*, namely, that she understands George Eliot's style. B's use of the *present perfect* therefore requires conversational background.

Nishiyama & Koenig (2010) show that the English *present perfect* is not only used to answer a question under discussion, but can also introduce a new topic. Their examples stem from the Switchboard corpus, in which two participants talk on a given topic, and generally ask whether the other participant has shared a similar experience on the topic.

- (5) A: **Have you seen** *Dancing with Wolves*?
B: Yeah, I've seen that, that's uh, that was a really good movie.

The two pragmatic functions of the English *present perfect* that Portner and Nishiyama & Koenig describe complement each other, but their remarks are somewhat preliminary, and they are not worked out in more detail in subsequent literature. Portner and Nishiyama & Koenig do not discuss the distribution of labour between PERFECT, PAST and PRESENT verb forms for temporal reference, so we don't know when the PERFECT is preferred over a PAST or PRESENT tense or why. What is more, under the assumption that pragmatic effects apply universally, a pragmatic account of the PERFECT leads us to expect stability in PERFECT use across languages, but of course, we already know from the examples in (1) through (3) that this prediction is not borne out.

In sum, PERFECTS in German and Romance languages share a common core meaning that comes out in contexts like (1). At the same time, the cross-linguistic variation illustrated in (2) and (3) is real, and needs to be accounted for. One promising route to deal with the competition between PERFECT and PAST is to explore PERFECT use in conversation. However, systematic research into the cross-linguistic variation of tense use in colloquial language is challenging, because actual conversations between speakers are typically not

translated. As high-quality spoken parallel corpora are not available, as we will argue in section 2, we turn to a parallel corpus of translations of J.K. Rowling's *Harry Potter and the Philosopher's Stone*.

1.3 From conversation to dialogue and parallel corpus research

The controlled environment of a literary parallel corpus offers insight into variation, as translators tend to stay close to the original meaning, but are free to in their choice of form. The *Harry Potter* novel contain both narrative discourse (in which the narrator tells the story) and dialogue (in which the characters talk among each other), and in this paper we use discourse and dialogue as a proxy for the difference between written and spoken language. This set-up enables us to study ways of temporal reference across languages in discourse and dialogue.

Our predictions based on the literature discussed so far are the following. Narrative discourse is strongly associated with a past orientation (Fleischman 1990), so we expect PAST tense forms to dominate in the narrative parts of *Harry Potter*. The temporal structure of narrative discourse is driven by events that move the storyline forward, while states overlap with the last introduced events (Hinrichs, 1986; Kamp & Reyle, 1993; Partee, 1984). Romance languages grammaticalize the distinction between IMPERFECTIVE and PERFECTIVE PAST, and dynamicity decides on tense use: the PERFECTIVE PAST introduces the events, while the IMPERFECTIVE PAST describes states that make up the background of the story (Kamp & Rohrer, 1983). In Germanic languages, which lack a grammaticalized PERFECTIVE/IMPERFECTIVE contrast, we expect the SIMPLE PAST to be used for the introduction of the main events as well as the background states.

In the dialogue parts, the strong presence of the utterance situation predicts a key role for reference to the present time domain. We further expect a division of labour between FUTURE, PRESENT, PAST, and PERFECT. The PRESENT links to events that take place during conversation and introduces liveliness in the story. Earlier occurring events could be referenced with use of the PAST and PERFECT, and upcoming events invite the use of FUTURE tense form. These hypotheses are mostly based on common sense, because apart from Wide (2002) on the Icelandic *búinn* construction, we are not aware of much systematic research into tense use in dialogue, let alone the PERFECT.

Parallel corpus research requires a computational methodology to draw inferences from variation. Typically, in typology, variation is plotted on semantic maps (in terms of Haspelmath (1997)). Wälchli & Cysouw (2012) have proposed to automatically generate semantic maps from parallel corpus variation. They use multidimensional scaling for this purpose: this method allows to visualize and interpret relevant dimensions of cross-linguistic variation. While Wälchli and Cysouw worked on the lexical domain on motion verbs, in later work, they applied their methodology in the domain of grammar, specifically *iamatives* (Dahl & Wälchli, 2016).

In this paper, we present a version of this methodology that we dubbed *Translation Mining* (van der Klis, Le Bruyn, & de Swart, 2017). Our goal is to create semantic maps (Haspelmath, 1997) of tense use directly from annotated data. The interface we develop facilitates the interpretation of semantic maps generated by multidimensional scaling. In section 2, we describe our interface in more detail. From these semantic maps, we are able to derive the micro-typology of PERFECT use across Germanic and Romance languages that we are after.

In van der Klis et al. (2019) we have already showcased this computational methodology to detect cross-linguistic variation in PERFECT use in a parallel corpus that consisted of translations of Albert Camus (1942) *L'Étranger*. We found a subset relation ranging from liberal to strict perfect use across Germanic and Romance languages. In this hierarchy displayed in (6), French and Italian form one end of the spectrum, while

English forms the other end. The inversed subset relation (\supseteq) signals that languages on the left use the PERFECT in every case where languages on the right use the PERFECT, including additional cases in which languages on the right use the PAST. For example, Dutch uses the PERFECT in all contexts in which Spanish and English use the PERFECT, but allows the PERFECT also in cases like (2) above, with an overt past time adverbial.

- (6) *Hierarchy of PERFECT use in Germanic and Romance languages* (van der Klis et al., 2019)
French/Italian \supseteq German \supseteq Dutch \supseteq Spanish \supseteq English

The *Harry Potter* novel that we investigate in this paper differs from *L'Étranger* in that the novel is written in English, which has a relatively strict *present perfect* as opposed to the more liberal French *passé composé*. Also, whereas *L'Étranger* is written in diary style, the *Harry Potter* novel is written as a classic novel with an omniscient narrator and contains a strict distinction between dialogue (in which the characters talk among each other) and narrative discourse (in which the narrator tells the story). So, the *Harry Potter* novel allows us to compare tense use in these two components of storytelling. Finally, translators are not expected to have paid special attention to tense and aspect in their translations. *L'Étranger* is well known for its at the time shocking use of the *passé composé* as the literary tense in lieu of the *passé simple* (de Swart, 2007), and this may lead to a translation bias (Gellerstam 1986). In contrast, *Harry Potter and the Philosopher's Stone* is, apart from its magical elements, a straightforward, contemporary novel and thus lacks the potential confound we find in *L'Étranger*.

1.4 Structure of the paper

This paper is structured as follows. In section 2, we put forward our parallel corpus methodology by describing the corpus selection, the corpus preparation, the annotation phase, and the computational method to transform variation into semantic maps called *Translation Mining*. In section 3, we compare the tense use in narrative discourse and dialogue. We find that both components of storytelling differ strongly in their tense use. Also, languages tend to use the same tenses in narrative discourse but have varying tense use in dialogue. Then, in section 4, we zoom in on the cross-linguistic differences in tense use in dialogue by applying our *Translation Mining* methodology. We find corroborating evidence for the implicational hierarchy of PERFECT use in Romance and Germanic languages put forward by van der Klis et al. (2019) and repeated in (6) above. Section 5 provides avenues for further research and concludes.

2. Methodology

In this section, we describe our rationale for choosing a literary parallel corpus, the creation of the corpus, the annotation process and the computational methodology to generate semantic maps from annotated data.

2.1. Corpus selection

Several parallel corpora are already in existence, but we felt none of them suitable for our purpose of analyzing tense use in narrative discourse and dialogue. We discuss why we did not opt for the Bible corpus, the Europarl corpus, or the OpenSubtitles corpus, before introducing our own literary parallel corpus.

The Bible corpus (Christodouloupoulos & Steedman, 2015) seems especially suited for typological research given its number of available translations. In the domain of tense and aspect, several studies have been performed (Asgari & Schütze, 2017; Dahl & Wälchli, 2016). However, its register is unusual: its religious nature sets it apart from other texts (de Vries, 2007). Furthermore, the Bible itself has no clear original source text: some translations take Hebrew as its source, others start from the Greek text. Also, not all translations

are recent, and interpretations of the Bible vary over time (Stolz, 2007). Finally, dialogues are mostly absent from the corpus.

The Europarl corpus (Koehn, 2005; Tiedemann, 2012), a compilation of proceedings of the European parliament, is a corpus of spoken dialogue and contains 24 languages. Although of course, the corpus is limited to languages spoken in Europe, it does contain the languages of our interest. However, since speakers in the Parliament can speak in any European language, the direction of translation is not always clear. This issue has led to the creation of a directed version of the corpus, Euoparl Direct (Cartoni & Meyer, 2012). However, speakers are not bound to their native language, which potentially generates confounds on the source language as well. Furthermore, after 2004, due to limited translation capacity, translators of lesser spoken languages have been using pivot languages, instead of directly translating from the original language the speaker uttered their sentence in (Hoek, 2018, p. 82).

The Europarl corpus has a high register, atypical for colloquial speech. The average number of words of English sentences in the corpus is reported as high as 21.61, which is more than the average sentence length of 18.12 in the written parts of the British National Corpus (Bick, 2014). Long sentences allow for more freedom in translation, rendering a lot of translation pairs unusable for form-based comparison (van der Klis et al., 2017). Recently, a filter to find syntactically comparable translations pairs has been developed to remedy this latter point (Kroon, Barbiers, Odijk, & van der Pas, 2019). Finally, the Europarl corpus can not really be considered a corpus of dialogue, given that most utterances are prepared beforehand or read from paper, rather than spontaneously spoken. Also, utterances can be post-edited. Still, the tense use in Europarl is comparable to dialogue. For example, in French, we find no use of the *passé simple* in this formal register, but use of the *passé composé* instead (van der Klis et al., 2019). Furthermore, the sheer size of the Europarl corpus cancels out most noise.

A recently created parallel corpus is the OpenSubtitles corpus (Lison & Tiedemann, 2016), which is a compilation of subtitles of movies and television series. Given the number of available translations, the various source and target languages, informal register, and the usually abundant dialogue, the corpus seems like a good fit for our purposes. Levshina (2017), however, describes various caveats. First, the translations in this corpus are not created by professional translators and therefore are prone to translation errors. Also, translations may be based on other subtitles rather than directly transcribed from the language spoken in the movies or series. Furthermore, translators are potentially bound by a limitation on the number of characters that can appear on the screen, potentially leading to a preference for shorter forms (Díaz Cintas & A. Remael, 2014, pp. 202–204). In our case, this could create effects of translators using the shorter forms of PRESENT or PAST rather than the compound form that is the PERFECT. Still, statistical analysis reveals film subtitles are not fundamentally different from other varieties of spoken and written British and American English (Levshina, 2017). Similar considerations apply to another recent initiative, ParTy (Levshina, 2016), which next to movie subtitles contains subtitles to TED Talks.

Given the aforementioned caveats for existing corpora, we instead turned to a parallel corpus of novel translations. Novels, next to narrative discourse, contain representations of spoken dialogue in an informal register. Translations usually are of high quality and the direction of translation is clear. We therefore have a highly controlled environment in terms of translations.

In the next section, we describe how we turned the *Harry Potter* novel and its translations into a digital parallel corpus. Also, we show how we separated dialogue from narrative discourse.

2.2. Corpus creation

Our corpus consists of the original British English version of *Harry Potter and the Philosopher's Stone* by J.K. Rowling and translations into German, Dutch, French, Italian, and Castilian Spanish. The novels were first converted to digital text via optical character recognition. Then, annotators manually checked the OCR'ed files for spelling errors and aligned the translations on paragraph level with the English original. After this, the texts were sentence- and word-tokenized using *Uplug* (Tiedemann, 2003). Then, the texts were tagged for parts-for-speech using *TreeTagger* (Schmid, 1994). Finally, the source text and its translations were sentence-aligned using *HunAlign* (Varga et al., 2005).³ Bibliographical details for the *Harry Potter* translations are in Appendix A.

After the preprocessing stage, we split up the corpus in two parts: narrative discourse (in which the narrator tells the story) and dialogue (in which the characters talk among each other). In (7), we show an example of the automatic assignment of narrative discourse and dialogue, from the start of chapter 17 of the novel.

(7) It was Quirrell.

"You!" gasped Harry.

Quirrell smiled. His face wasn't twitching at all.

"Me," he said calmly. *"I wondered whether I'd be meeting you here, Potter."*

"But I thought – Snape –"

"Severus?" Quirrell laughed and it wasn't his usual quivering treble, either, but cold and sharp. *"Yes, Severus does seem the type, doesn't he? So useful to have him swooping around like an overgrown bat. Next to him, who would suspect p-p-poor st-stuttering P-Professor Quirrell?"*

In the above conversation, the parts in italics are between quotation marks and therefore are regarded as dialogue. All other parts are part of the narrative discourse. The division between narration and dialogue has been created automatically with a Python script that matches quotation marks. The script can be found on GitHub.⁴ During the annotation phase, that will be discussed in more detail in the next paragraph, annotators were asked to tick a checkbox when this script had made errors in either the English original or the translation.

Our dataset in this paper consists of the first and last chapters of *Harry Potter and the Philosopher's Stone* and its translations. To be able to better situate the examples in the upcoming sections 3 and 4, we provide a short synopsis of these two chapters. Chapter 1 sets the scene of the novel, introduces us to the Dursley family and sees the young Harry Potter being delivered to the Dursleys by professors Dumbledore, McGonagall, and gamekeeper Hagrid. In chapter 17, we see Harry confronting Voldemort, the aftermath with Dumbledore and shortly thereafter with Hermione and Ron, the decision on the results of the House Cup and the final goodbyes at the end of the school year.

In the next section, we describe the annotation process.

2.3. Annotation

In the annotation phase, annotators with proficiency in English first manually selected all finite, non-modal verb phrases from chapters 1 and 17. We disregarded non-finite forms (infinitives, gerunds, and participles) as these are not in direct competition with finite forms like the PERFECT. Modals were not included because

³ *HunAlign* was provided with a translation dictionary of realia in *Harry Potter* to slightly improve its sentence alignment.

⁴ <https://github.com/time-in-translation/dialog-detection>

of potential interactions of tense and modality. Imperatives, on the other hand, were kept in, as they form a unique feature of spoken language and are rare in ordinary written text (Bick, 2014, p. 109).

The selection of verb phrases was done in a web application dubbed *PreSelect*. Figure 1 displays the annotation interface. Annotators can select one or more verb phrases within a fragment, assign a tense (or another label), and provide comments on their selections. After submitting one or multiple selections, the annotators receive the next randomly chosen fragment.

English 17.xml - s52.2

' I - I've won the House Cup for Gryffindor . '

This fragment does not contain a target

All targets have already been selected in this fragment

Tense

future continuous

Other label

Other label

Comments

Comments

Figure 1: The *PreSelect* annotation interface. Annotators were shown sentences from the English original and were asked to select all finite, non-modal verb phrases.

In the second step, annotators with proficiency in the translated languages annotated the (verbal) translations of the verb phrases selected through *PreSelect* in the first step of the annotation. Annotators were shown the selected verb phrase in the original English version, aligned with the translation in the language they were proficient in. Annotators were asked to select the words that translate the selected verb phrase. Annotators were given checkboxes to mark problems with the selection in the original language or with the translation (i.e. a wrong selection or annotation during the *PreSelect* phase, a mismatch during the automatic sentence alignment, an incorrect automatic marking of dialogue or narration in the original fragment, a change in the sentence structure in the translation). Finally, annotators could add comments if necessary. After submitting their selection and annotations, annotators were redirected to the next randomly chosen fragment. Figure 2 displays the parallel annotation interface, dubbed *TimeAlign* (van der Klis et al., 2017). The source code of both *PreSelect* and *TimeAlign* is available via GitHub.⁵

⁵ <https://github.com/UUDigitalHumanitieslab/timealign>

English (original)

17.xml - s52.2

' I - I **ve won** the House Cup for Gryffindor . '

Spanish (translated)

17.xml - s52.2

Yo ... **he ganado** la copa de la casa para Gryffindor .

- The selected words in the original fragment do not form an instance of (a/an) *present perfect*
- This is a correct translation of the original fragment
- The selected words in the original fragment are incorrectly marked as *dialogue*
- The translated fragment is not in the same structure (dialogue/narrative) as the original fragment

Comments

Comments

Submit

Figure 2: The TimeAlign annotation interface. Annotators were shown the selected verb phrase in the English original, aligned with the translation in, in this case, Spanish. Annotators were asked to select the words that translate the English selection.

After aligning verb phrases and checking for potential errors in the dialogue attribution, annotators were asked to provide a tense-aspect label (TA-label) per selected verb phrase. Annotation of TA-labels was performed in an Excel document, that was later re-uploaded to the *TimeAlign* annotation interface. For the TA-labels, annotators could pick from a list of TA-labels per language. An excerpt from this list is provided in Table 1. The first column in Table 1 provides the higher-level tense-aspect category for each TA-label. Note that Germanic languages lack a grammaticalized distinction between PERFECTIVE and IMPERFECTIVE PAST, and hence lack a TA-label in these cells. Similarly, all languages except English and Italian lack a grammaticalized PROGRESSIVE form.

TA-category	English	German	Dutch	French	Italian	Spanish
PRESENT	<i>simple present</i>	<i>Präsens</i>	<i>ott</i>	<i>présent</i>	<i>presente</i>	<i>presente</i>
PRESENT PROGRESSIVE	<i>present continuous</i>	-	-	-	<i>stare (presente) + gerundio</i>	-
PAST	<i>simple past</i>	<i>Präteritum</i>	<i>ovt</i>	-	-	-
PERFECTIVE PAST	-	-	-	<i>passé simple</i>	<i>passato remoto</i>	<i>pretérito indefinido</i>
IMPERFECTIVE PAST	-	-	-	<i>imparfait</i>	<i>imperfetto</i>	<i>pretérito imperfecto</i>
PAST PROGRESSIVE	<i>past continuous</i>	-	-	-	<i>stare (imperf.) + gerundio</i>	-
PERFECT	<i>present perfect</i>	<i>Perfekt</i>	<i>vtt</i>	<i>passé composé</i>	<i>passato prossimo</i>	<i>pret. perf. compuesto</i>
PAST PERFECT	<i>past perfect</i>	<i>Plusquam-perfekt</i>	<i>vvt</i>	<i>plus-que-parfait</i>	<i>trapassato prossimo</i>	<i>pret. plus-cuamperfecto</i>
FUTURE	<i>simple future</i>	<i>Futur I</i>	<i>ottt</i>	<i>futur simple</i>	<i>futuro semplice</i>	<i>futuro imperfecto</i>

Table 1: A selection of nine tense-aspect-categories and their respective TA-forms per language.

Assigning the TA-label marked the end of the annotation phase. In the next section, we describe our methodology for analyzing annotated datasets.

2.4. Analysis through Translation Mining

To construct semantic maps directly from data extracted from parallel corpora, we apply an existing method in the lexical domain at the level of grammar. Wälchli & Cysouw (2012) proposed to generate semantic maps through lexical variation. In earlier work, we extended Wälchli and Cysouw’s methodology to the level of grammar and dubbed it *Translation Mining* (van der Klis et al., 2017).

The method works as follows. We first define a context as a tuple of assigned TA-labels for a (translated) verb phrase in all languages. For example, (1) := $\langle \textit{present perfect}, \textit{Perfekt}, \textit{vtt}, \textit{passé composé}, \textit{passato prossimo}, \textit{pret. perf. comp.} \rangle$, (2) := $\langle \textit{simple past}, \textit{Perfekt}, \textit{vtt}, \textit{passé composé}, \textit{passato prossimo}, \textit{pret. indefinido} \rangle$, and (3) := $\langle \textit{present perfect}, \textit{Präsens}, \textit{ott}, \textit{présent}, \textit{presente}, \textit{presente} \rangle$.

We then define a simple distance function d . Two contexts are considered most similar ($d=0$) if the TA-labels for each language match, and most dissimilar ($d=1$) if none of the TA-labels match. For example, between (1) and (2), $d=0.333$, as two out of six languages use a different tense (PAST instead of PERFECT). Similarly, between (1) and (3), $d=0.833$, as only English uses the *present perfect* in both examples. Finally, between (2) and (3), $d=1$, as none of the TA-labels match. The resulting distance matrix for examples (1) through (3) is shown in Table 2 below.

Context	(1)	(2)	(3)
(1)	-	0.333	0.833
(2)	0.333	-	1
(3)	0.833	1	-

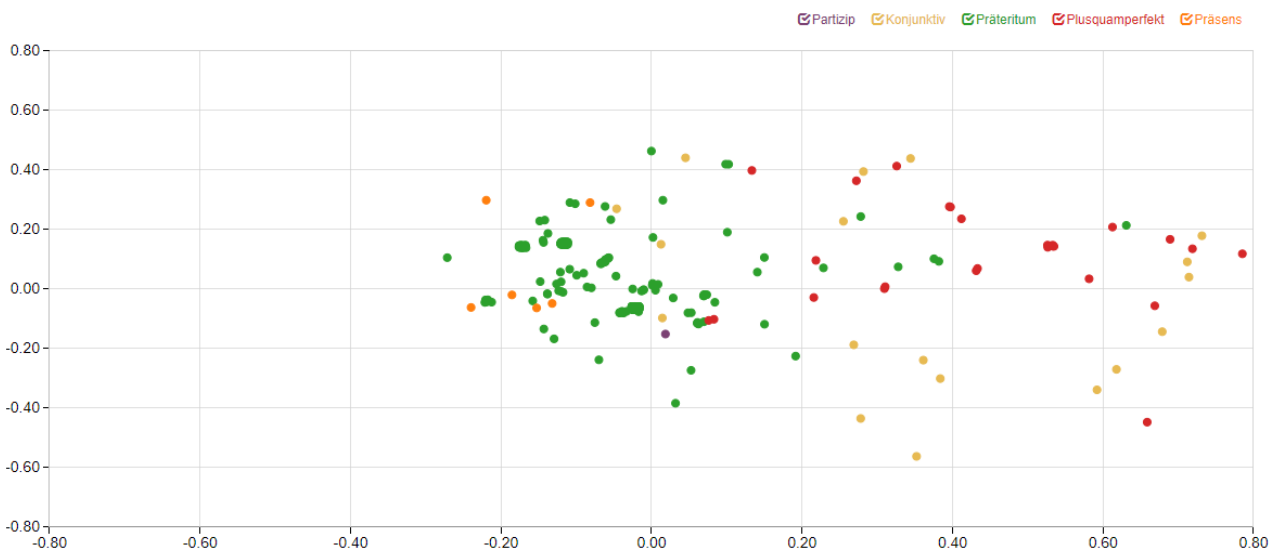
Table 2: Distance matrix based on the TA-label assignments of examples (1) through (3) in this paper.

When we apply this distance function to all contexts in the dataset, this process generates an n -dimensional distance matrix, with n the number of contexts. To visualize this distance matrix in a space of lower dimensions, one can use a dimensionality reduction algorithm like multidimensional scaling (MDS). The MDS algorithm tries to create a low-dimensional representation of the data, while still respecting distances in the original high-dimensional space.⁶ For example, a two-dimensional representation of distances between cities in Europe is a topographic map. However, as a consequence of dimensionality reduction, this representation disregards altitude and the curvature of the earth. Still, such a map did clearly help to navigate drivers through the continent before the recent development and omnipresence of GPS systems.

An online interface included in the *TimeAlign* software package allows us to create various MDS visualizations per corpus (in our terminology: *scenarios*), e.g. to only select the verb phrases appearing in dialogue, or only *present perfects* and their translations. The software not only shows the MDS solution but colours the data points according to the TA-labels of a chosen language, leading to an interpretable semantic map. Figure 3 shows the interface. The interface guides easier interpretation of the semantic maps by allowing users to visualize and filter the results per language, rather than providing one global solution.

MDS visualization (scenario *HP 1-17 narrative*)

Stress: 91.685



Filters

Language: German English Spanish French Italian Dutch Dimension on x-axis: 1 2 3 4 5 Dimension on y-axis: 0 1 2 3 4 5

Figure 3: The interface to the MDS solution. The filters on the bottom allow displaying the solution using the picked language, as well as to show other dimensions. The tenses displayed on the top provide a legend. Every dot can be clicked to see its underlying contexts.

⁶ In this paper, we refer to the outcome of MDS as a *solution* and its visual representation in our interface as a *semantic map*.

Furthermore, one of the advantages of MDS is that it is a *white box* method: all data points on the map can be traced back to (in our case) individual contexts. This allows to directly analyze the linguistic features of contexts that end up on parts of the map. Furthermore, potential outliers can be traced both visually and automatically. In the online interface, this is operationalized by drill-through functionality: users can click on points on the map, which will show the underlying contexts. For example, clicking on the right-most point in Figure 3 yields the context displayed in Figure 4 below. This allows a back-and-forth analysis between the visualization and the data.

Fragment overview



Source

English

declarative narration 17.xml past perfect

Where there should have been a back to Quirrell 's head , there was a face , the most terrible face Harry had ever seen .

Translations

German

Plusquamperfekt

Wo eigentlich Quirrells Hinterkopf hätte sein sollen , war ein Gesicht , das schrecklichste Gesicht , das Harry jemals gesehen hatte .

Spanish

subjuntivo pretérito pluscuamperfecto

Donde tendría que haber estado la nuca de Quirrell , había un rostro , la cara más terrible que Harry hubiera visto en su vida .

French

subjunctif passé

Derrière la tête de Quirrell , au lieu de son crâne , il y avait un visage , le visage le plus terrifiant que Harry eût jamais vu .

Italian

congiuntivo trapassato

Nel punto dove normalmente avrebbe dovuto trovarsi la nuca del professore , c' era un volto , il volto più orrendo che Harry avesse mai visto .

Dutch

vvt

Waar Krinkels achterhoofd had moeten zitten zat een gezicht , het vreselijkste gezicht dat Harry ooit had gezien .

Figure 4: Individual fragment overview, that allows seeing the source context and its translations, as well as the annotated verbal translation of the source verb phrase and the assigned TA-labels.

Finally, the interface allows an option to include incomplete tuples, contexts which miss TA-labels for one or more of the languages. Incomplete tuples arise in the case of nominal periphrases by translators or mistakes in the automatic sentence alignment. In this paper, we stick to complete contexts, though.

In the next two sections, we show the results of applying our parallel corpus methodology to the two subsets of the dataset: narrative discourse and dialogue. In section 3, we first compare tense use in both components of storytelling. In section 4, we apply multidimensional scaling to the domain of dialogue.

3. Tense use in discourse and dialogue

As stories are traditionally told in the past tense, we expect PAST tense forms to be dominant in the narrative discourse parts of the novel. In the introduction, we predicted tense use in dialogue to display a division of labour between FUTURE, PRESENT, PAST, and PERFECT. In this section, we present our results of the annotation of all indicative, non-finite verb phrases in chapter 1 and 17 of *Harry Potter and the Philosopher's Stone* and its translations. The narrative discourse consists of 800 English verb phrases in total. Of these, 513 verb phrases have been aligned and annotated for tense in all languages. For dialogue, we find 586 verb phrases in total. Of these, 355 contexts have TA-labels in all languages. Figure 5 shows the distribution of the five major indicative tense-aspect categories in dialogue and narrative discourse. To simplify the picture

somewhat, remaining TA-labels are not displayed in the graphs. Furthermore, for the Germanic languages, we used the same colour assignment for PAST as for the Romance PERFECTIVE PAST.⁷

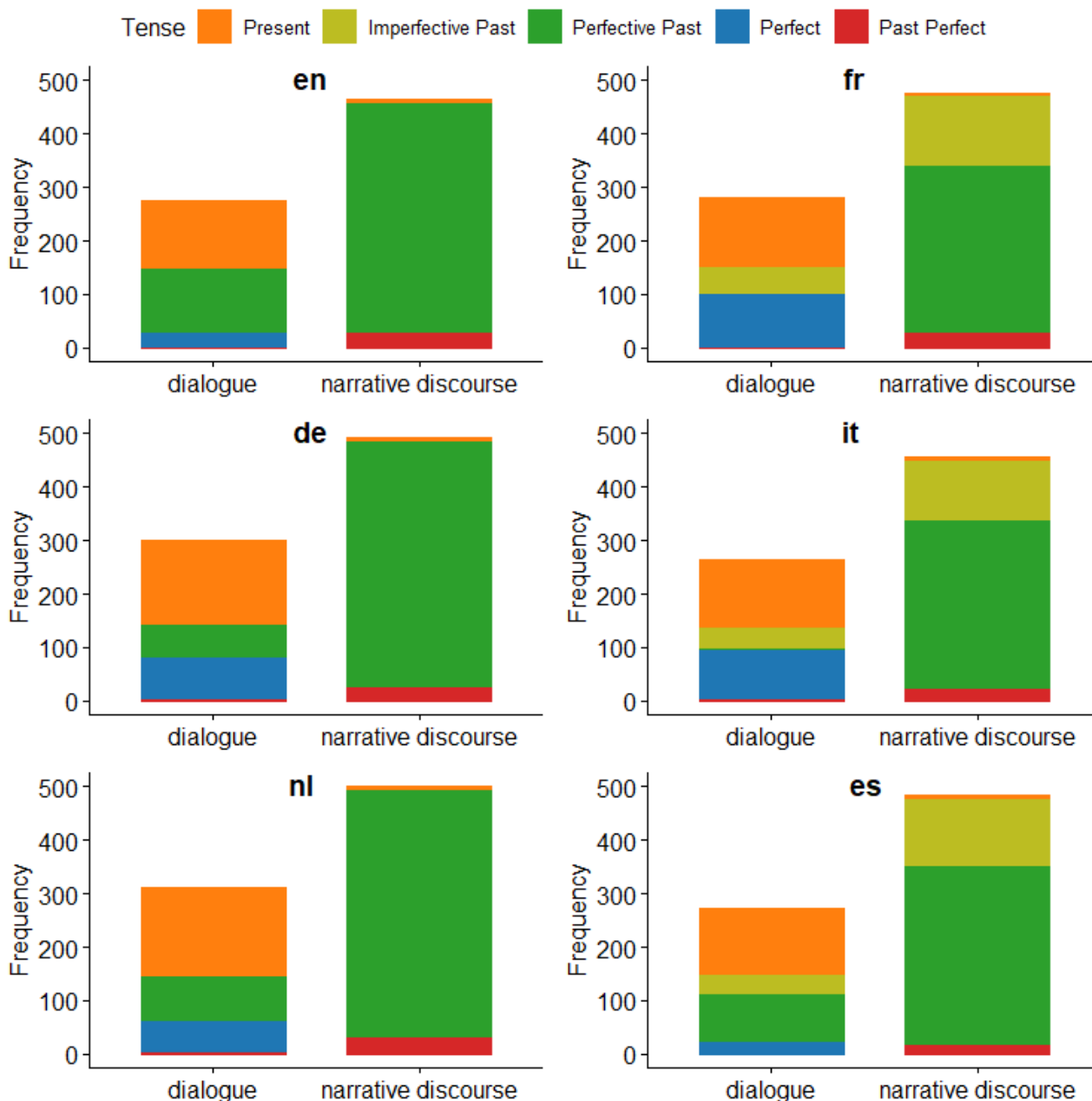


Figure 5: Tense use of PRESENT, PERFECTIVE PAST, IMPERFECTIVE PAST, PERFECT, and PAST PERFECT in dialogue and narrative discourse in the first and last chapter of Harry Potter and the Philosopher’s Stone. On the left side, the Germanic languages are displayed (English, German, and Dutch from top to bottom) and on the right side, the Romance languages are displayed (French, Italian, and Spanish).

The graphs visualize major distinctions in tense use between narrative discourse and dialogue, not only for languages like French, Italian, and German in which these distinctions have been made in the literature but for all other languages as well. French, for example, sees an alternation between the *imparfait* and the *passé*

⁷ Throughout the paper, we continue using the same color assignment as in Figure 5: orange for PRESENT, dark green for PERFECTIVE PAST, light green for IMPERFECTIVE PAST, blue for PERFECT, red for PAST PERFECT, etc.

simple in narrative discourse, but in dialogue, the *passé simple* is absent, and the *passé composé* is used instead.

Strikingly, we do not find PERFECT use in narrative discourse in any of the languages under investigation. All languages keep close to the view held in the literature about storytelling: the Germanic languages almost exclusively use the PAST, while the Romance languages grammaticalize the perfective/imperfective distinction.

In dialogue, on the other hand, we do find PERFECT use. However, the number of PERFECTS clearly differs per language. For example, if we compare the Romance languages, we see a lot of PERFECTS in French and Italian, but only a small amount of PERFECTS in Spanish. Similarly, while we see almost no PERFECTIVE PAST use in French and Italian, the number of PERFECTIVE PASTS in Spanish is sizable, although fewer than in narrative discourse.

If we shift our focus to the PRESENT, we see that while there is almost no PRESENT use in narrative discourse, we find that in dialogue, all languages use a similar number of PRESENTS. Of course, this does not yet imply that all languages have a shared space of PRESENT use, as we are taking a monolingual perspective in this section. In the next section, we take a cross-linguistic perspective to be able to tell whether languages use a PRESENT in similar contexts.

To show the statistical significance of differences in tense use between narrative discourse and dialogue, we can use tests for association per language. For English and French, we report the frequencies that were displayed in Figure 5 in table format in Table 3 and Table 4 respectively.

Formal structure	<i>simple present</i>	<i>simple past</i>	<i>present perfect</i>	<i>past perfect</i>	other	total
dialogue	126	120	27	2	80	355
narrative discourse	5	430	0	29	49	513

Table 3: Tense use in dialogue and narrative discourse in the English original of Harry Potter and the Philosopher’s Stone, chapters 1 and 17.

Formal structure	<i>present</i>	<i>imparfait</i>	<i>passé simple</i>	<i>passé composé</i>	<i>plus-que-parfait</i>	other	total
dialogue	130	50	0	99	3	73	355
narrative discourse	3	130	313	0	30	37	513

Table 4: Tense use in dialogue and narrative discourse in the French translation of Harry Potter and the Philosopher’s Stone, chapters 1 and 17.

In the statistical association tests, we remove the column with ‘other’ TA-label assignments. For English, in Table 3, we then find a significant association between formal structure and tense use: $\chi^2(3) = 308.87$, $p < 0.001$. The effect size is large (Cramér’s $V = 0.646$). For French, in Table 4, we also find a significant association between formal structure and tense use: $\chi^2(4) = 579.21$, $p < 0.001$.⁸ The effect size again is large (Cramér’s $V = 0.874$). For other languages, we also find a highly significant result. The results are summarized in Table 5 below.

⁸ Note that for French, the degrees of freedom is 4, given that the Romance languages have a grammaticalized distinction between IMPERFECTIVE PAST and PERFECTIVE PAST.

Language	df	χ^2	p	Cramér's V
English	3	308.87	< 0.001	0.646
German	3	523.33	< 0.001	0.811
Dutch	3	474.15	< 0.001	0.763
French	4	579.21	< 0.001	0.874
Italian	4	551.14	< 0.001	0.875
Spanish	4	304.32	< 0.001	0.634

Table 5: Results of the association tests between formal structure and tense use. All language show a significant association between formal structure and tense use.

In this section, we have seen that in narrative discourse, languages typically follow the theoretical perspective that has been put forward in the introduction. Germanic languages tell stories with a general-purpose PAST, while Romance languages alternate between PERFECTIVE and IMPERFECTIVE PAST (Kamp & Rohrer, 1983). However, tense use in dialogue significantly differs from tense use in narrative discourse. Furthermore, this section already hints at the variation between languages in PERFECT use in dialogue: languages differ in their frequency of PERFECT use, as Figure 5 has shown us. Also, the reported frequencies hint at a potential shared semantic space of PRESENT use. Hence, in the next section, to research these hypotheses, we take a cross-linguistic perspective at tense use in dialogue.

4. Cross-linguistic variation in tense use in dialogue

The global predictions advanced in the introduction about the differences in discourse and dialogue are borne out, as we see from the descriptive statistics presented in the last section. However, we were not yet able to ascertain whether languages use these tense-aspect categories similarly, or that they differ in tense use in some contexts. In this section, we set out to answer this question with an exploratory analysis of tense use in dialogue in *Harry Potter and the Philosopher's Stone* and its translations through the number of verb forms used in dialogue. In a second step, we analyze the cross-linguistic variation through multidimensional scaling.

4.1. Analysis through descriptive statistics

As stated in the previous section, through *PreSelect*, annotators found 586 verb phrases in total in the dialogue parts in the novel. Of these, 355 contexts were assigned TA-labels in all languages through *TimeAlign*. Table 6 shows the tallies per tense category.

TA-category	English	German	Dutch	French	Italian	Spanish
PRESENT	126	156	164	130	124	122
PRESENT PROG.	11	-	-	-	3	-
PAST	120	62	83	-	-	-
PERF. PAST	-	-	-	0	2	89
IMPERF. PAST	-	-	-	50	39	38
PAST PROG.	6	-	-	-	3	-
PERFECT	27	79	59	99	94	23
PAST PERFECT	2	4	6	3	4	1
FUTURE	20	14	9	11	23	13
other	43	40	34	62	63	69

Table 6: Number of occurrences for eight selected tense-aspect-categories in the dialogues of the novel.

While the PERFECT was absent in narrative discourse, it is prominently used in dialogue. Table 6, however, shows that languages do differ in their number of PERFECTS. Also, compared to the narrative discourse, there is increased use of the PRESENT. However, for the PRESENT, the tallies across languages do not differ that much, hinting at a shared space of PRESENT use between languages.

In Table 6Table 4, we also see frequent use of the FUTURE. The numbers between languages differ slightly, part of which is potentially explainable by the observation from the literature that German and Dutch allow a PRESENT to express more futurate uses (Hilpert, 2008, pp. 249–286). Finally, we find only a small amount of PAST PERFECT uses, especially when compared to narrative discourse.

Nevertheless, Table 6 does not tell us whether languages overlap in their tense use, or whether, for example, all English *present perfects* are translated with a German *Präsens*. Tuple frequencies allow insight into this. Table 7 shows the most frequent TA-tuples in dialogue.

English	German	Dutch	French	Italian	Spanish	#
<i>simple present</i>	<i>Präsens</i>	<i>ott</i>	<i>présent</i>	<i>presente</i>	<i>presente</i>	86
<i>simple past</i>	<i>Perfekt</i>	<i>vtt</i>	<i>passé composé</i>	<i>passato prossimo</i>	<i>pretérito indefinido</i>	28
<i>imperative</i>	<i>Imperativ</i>	<i>imperatief</i>	<i>impératif</i>	<i>imperativo</i>	<i>imperativo</i>	18
<i>simple past</i>	<i>Perfekt</i>	<i>ovt</i>	<i>passé composé</i>	<i>passato prossimo</i>	<i>pretérito indefinido</i>	13
<i>simple past</i>	<i>Präteritum</i>	<i>ovt</i>	<i>imparfait</i>	<i>imperfetto</i>	<i>pretérito imperfecto</i>	12
<i>simple past</i>	<i>Präteritum</i>	<i>ovt</i>	<i>passé composé</i>	<i>passato prossimo</i>	<i>pretérito indefinido</i>	10
<i>present perfect</i>	<i>Perfekt</i>	<i>vtt</i>	<i>passé composé</i>	<i>passato prossimo</i>	<i>pret. perf. comp.</i>	8

Table 7: Tuples of TA-labels with more than five occurrences in the dialogue dataset.

What immediately draws attention in Table 7 is the overlap in the use of the PRESENT: in 86 contexts, all languages use a PRESENT. Of the 126 *simple present* uses in English, 86 (68%) are translated with a PRESENT in all languages. Most of these contexts are factual statements with copula verbs *to be* or *to have*, like in (8) below.

(8) Context: Dumbledore has just said scars like Harry’s might come in handy. He explains:

I **have** one myself above my left knee which **is** a perfect map of the London Underground.

[English, *simple present*]

Ich selbst **habe** eine oberhalb des linken Knies, und die **ist** ein tadelloser Plan der Londoner U-Bahn.

[German, *Präsens*]

Ik **heb** er zelf ook eentje, net boven mijn linkerknie en dat **is** precies de plattegrond van de Londense metro.

[Dutch, *ott*]

Moi-même, j’en **ai** une au-dessus du genou gauche, qui **représente** le plan exact du métro de Londres.

[French, *présent*]

Anch'io ne **ho** una, sopra il ginocchio sinistro, che è una piantina perfetta della metropolitana di Londra. [Italian, *presente*]

Yo **tengo** una en la rodilla izquierda que **es** un diagrama perfecto del metro de Londres. [Spanish, *presente*]

In 28 contexts, we find uses similar to (2): French, Italian, German, and Dutch use a PERFECT, while English and Spanish resort to a PAST. We find two similar patterns in Table 7, one in which Dutch also patterns with English and Spanish on PAST use (13 contexts) and one in which only French and Italian use a PERFECT, and the other languages use a PAST (10 contexts). Of these three patterns, we show more examples when we analyze the generated semantics maps below.

Next, we find that languages generally overlap in use of the IMPERATIVE. There are 25 *imperatives* in the English original, of which 18 (72%) are translated with an IMPERATIVE in all languages. (9) shows an example.

- (9) **Look** in the Mirror and **tell** me what you see. [English, *imperative*]
Schau in den Spiegel, und **sag** mir, was du siehst. [German, *Imperativ*]
Kijk in de spiegel en **zeg** wat je ziet. [Dutch, *imperatief*]
Regardez dans le miroir et **dites-moi** ce que vous y voyez. [French, *impératif*]
Guarda nello specchio e **dimmi** che cosa vedi. [Italian, *imperativo*]
Mira en el espejo y **dime** lo que ves. [Spanish, *imperativo*]

Then, we find an imperfective pattern in 12 contexts, with the Romance languages using an IMPERFECTIVE PAST, and the Germanic languages all using their general-purpose PAST. This is reminiscent of the distribution of labour that we found in narrative discourse in the previous section, and is in line with the theoretical background covered in the introduction.

Finally, we find 8 contexts in which all languages use a PERFECT. In (1), we have already seen an example of PERFECT use for all language for recent events. Van der Klis (2018) reports that there is little to no variation across languages for referring to recent events. French and Spanish do have a grammaticalized construction, the *passé récent* and *pasado reciente* respectively, but the translators both opted for the PERFECT in (1).

Use of the PERFECT in all languages is furthermore triggered by sentence-level adverbials. For example, we find a durational *for*-adverbial in (10), and a quantificational adverbial with *never* in (11).

- (10) We've **had** precious little to celebrate *for eleven years*. [English, *present perfect*]
Elf Jahre lang **haben** wir herzlich wenig zu feiern **gehabt**. [German, *Perfekt*]
Tenslotte **hebben** we *de afgelopen elf jaar* bitter weinig te vieren **gehad**. [Dutch, *vtt*]
Nous n'**avons** pas **eu** grand-chose à célébrer *depuis onze ans*. [French, *passé composé*]
Per undici anni **abbiamo avuto** ben poco da festeggiare. [Italian, *passato prossimo*]
Hemos tenido tan poco que celebrar *durante once años...* [Spanish, *pret. perf. comp.*]
- (11) My dear Professor, I've **never seen** a cat sit so stiffly. [English, *present perfect*]
Mein lieber Professor, ich **habe** noch *nie* eine Katze so steif dasitzen **sehen**. [German, *Perfekt*]
M'n beste professor, ik **heb** nog *nooit* een kat zo stijfjes **zien** zitten. [Dutch, *vtt*]
Mon cher professeur, je n'**ai jamais vu** un chat se tenir d'une manière aussi raide. [French, *passé composé*]
Ma, mia cara professoressa, non **ho mai visto** un gatto seduto in una posa così rigida. [Italian, *passato prossimo*]
Mi querida profesora, *nunca* **he visto** a un gato tan tieso. [Spanish, *pret. perf. comp.*]

In Rathert (2004), a distinction is made between positional adverbs (like *yesterday* or *May 1, 1999*), durational (like *until, since, in, and for* (in (10))) and quantificational adverbs (like *once, twice, often, seldom, sometimes, and never* (in (11))). With respect to the PERFECT, we find that positional adverbs lead to differences in tense use: English and Spanish require a PAST in these cases, while the other languages can use a PERFECT, as (2) has already shown us. Durational and quantificational adverbs, however, do generally lead to PERFECT use in all languages.

In this paragraph, we have seen Table 6 and Table 7 already hint at potential overlapping use of tense-aspect categories (especially the PRESENT), but also cross-linguistic variation in reference to past events. In the next paragraph, we use multidimensional scaling to be able to better assess the potential overlapping use of tense-aspect-categories.

4.2. Analysis through multidimensional scaling

Table 6 has already provided us with a general idea of how languages carve up the temporal semantic space. However, what Table 6 does not show us, is whether languages line up in their tense use, or that e.g. English *simple present* uses generally translate into German *Präteritum* uses, even though the number of German *Präsens* uses is similar to that of the *simple present*. Table 7 has improved our cross-linguistic view but does not readily provide an visual overview the variation. Here, multidimensional scaling (MDS) provides a helpful tool, as it reduces high-dimensional variation to a visually interpretable lower-dimensional solution.

When we run the MDS algorithm on all dialogue contexts in the dataset, we first arrive at a solution that is hard to interpret. To signal the goodness-of-fit of a solution, MDS returns a stress value. Lower values indicate better fit. With all dialogue contexts find that a 5-dimensional multidimensional scaling solution reaches a stress value as high as 2.34, signalling a poor fit.

However, find that this somewhat high stress value for dialogue is strongly dependent upon the use of the subjunctive and conditional mood for translation in languages other than English and Dutch. If we remove all contexts that include at least one translation with conditional or subjunctive mood, we are left with 304 contexts. The global stress of this solution drops to 0.90. Figure 6 shows the MDS solution with the assigned English TA-labels, yielding a semantic map of tense use in English. The most prominent dimension of variation is displayed on the x-axis, while the second dimension is displayed on the y-axis.

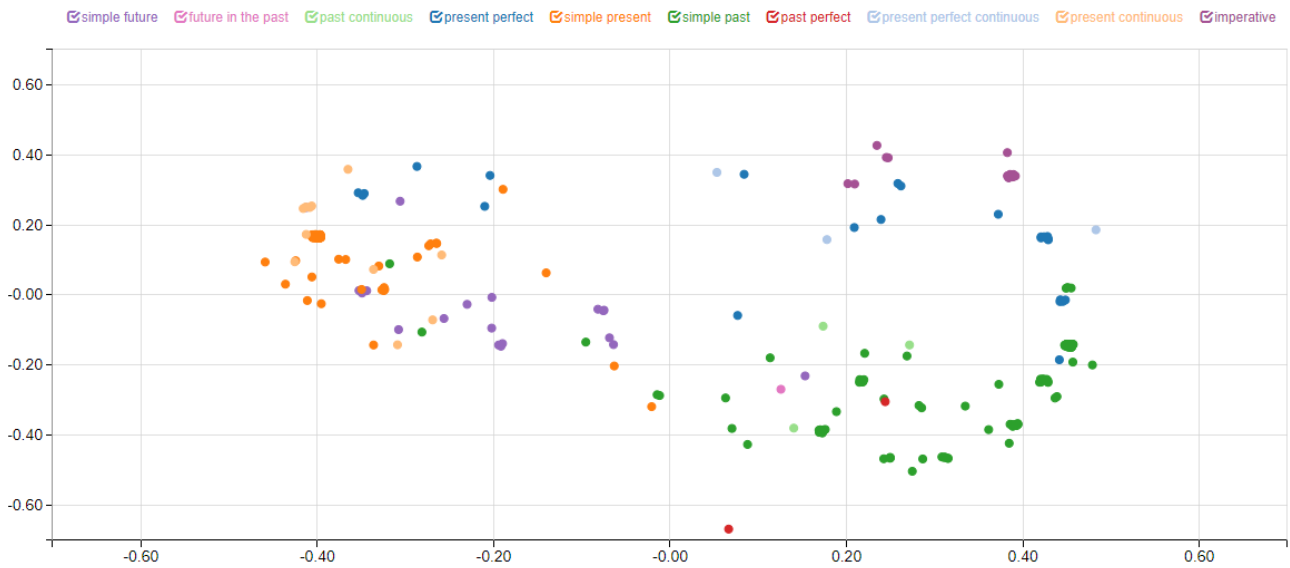


Figure 6: MDS solution for dialogue in chapters 1 and 17, coloured with the English TA-labels.

The semantic map for English in Figure 6 shows clustering for most TA-labels. Remember that, in our visualization, distinct colours refer to distinct TA-labels. For example, we see a cluster of *simple present* use (in orange) on the left side of the map and a cluster of *simple past* use (in dark green) on the right side of the map. Similarly, we find clusters of *simple future* and *imperative* use. However, we see that the *present perfect* appears in two small clusters: on the left, near the *simple present* cluster, and on the right, near the *simple past* cluster. Why this is, will become clear when we compare this temporal semantic map with the other languages.

Figure 7 and Figure 8 show the same MDS solution with assigned TA-labels for German and Dutch respectively. Later in this section, we introduce the Romance maps, but for now, let us first compare the semantic maps for the three Germanic languages.

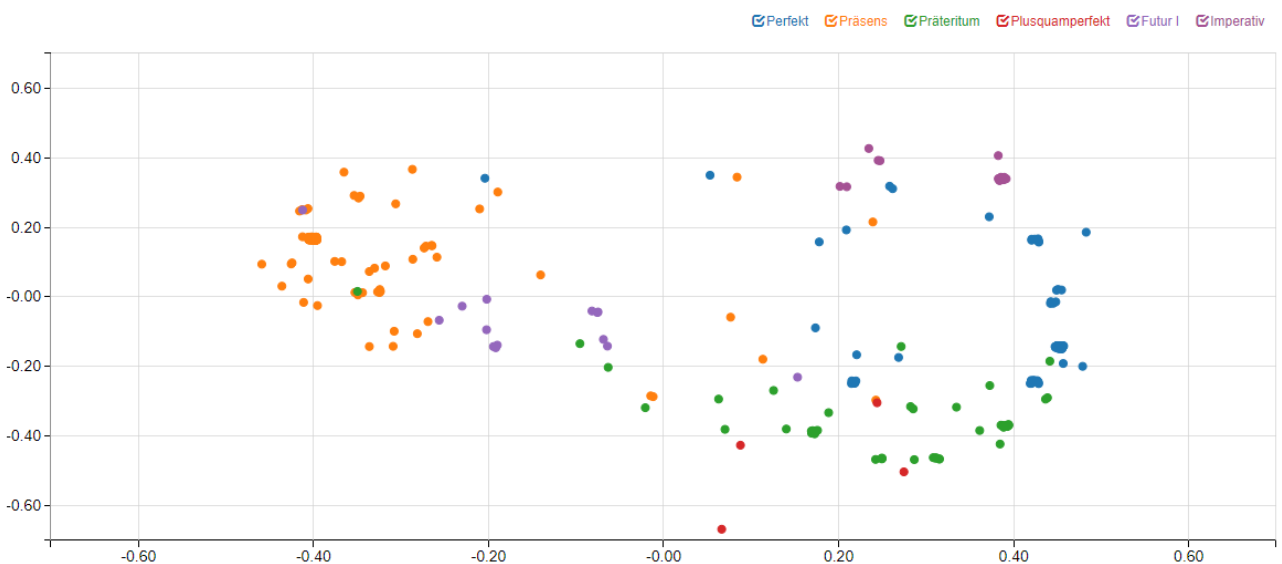


Figure 7: MDS solution for dialogue in chapters 1 and 17, coloured with the German TA-labels.

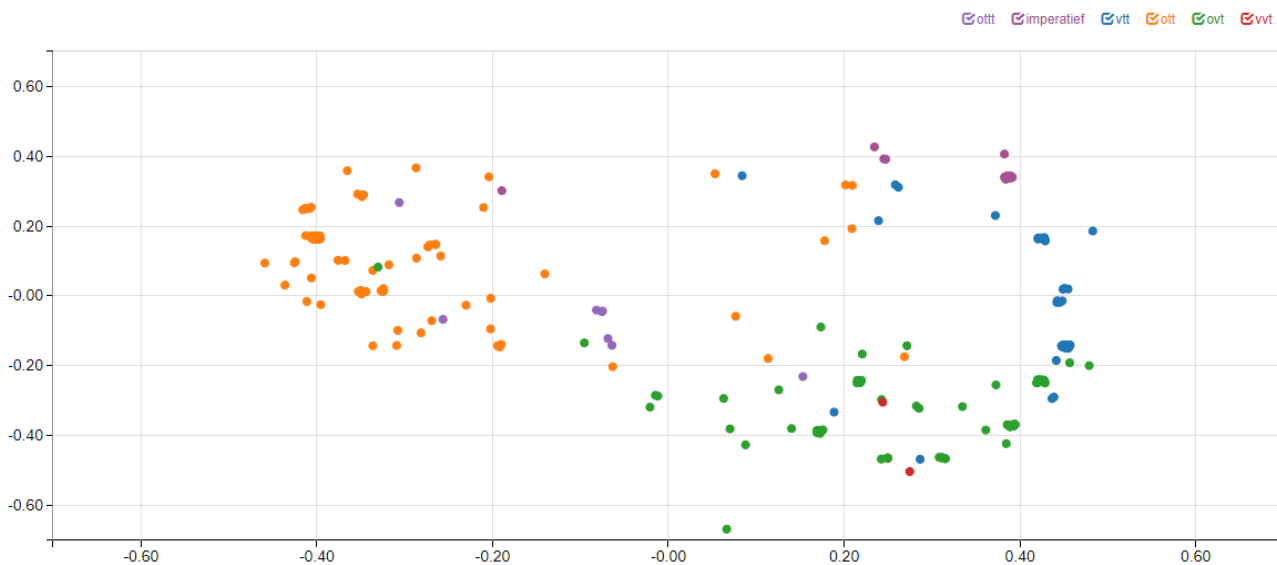


Figure 8: MDS solution for dialogue in chapters 1 and 17, coloured with the Dutch TA-labels.

When we compare the German solution in Figure 7 to the English solution in Figure 6, we see that the *Präsens* has completely taken over the *present perfect* cluster on the right of the map. In the Dutch map, in Figure 8, we see the same pattern for the *ott*. When we zoom in on these contexts, we find two types of contexts. First, we find context (3), the only example in our dataset of continuative *present perfect* use. In translations of (3), all languages used a PRESENT. Hence, MDS positions the contexts near contexts in which all languages use a PRESENT.

Another construction in which English opts for a PERFECT but other languages use a PRESENT is the resultative use in the construction *have got*, exemplified in (12). One can question whether this is grammaticalized construction rather than an actual PERFECT, especially since the past participle of *to get* is *gotten* instead of *got*. Nonetheless, we also find this resultative construction with *to leave* in (13). Here, German does use a *Perfekt* in translation, but similarly to (12), the focus is on the resulting state that persists into the here and now.

- | | |
|---|----------------------------|
| (12) Have a Chocolate Frog, I've got loads... | [English, present perfect] |
| Nimm einen Schokofrosch, ich hab ganze Wagenladungen davon... | [German, Präsens] |
| Vooruit, neem een Chocokikker, ik heb dozen vol... | [Dutch, ott] |
| Prenez donc un Chocogrenouille, j'en ai plein... | [French, présent] |
| Su, prendi una Cioccorana, ne ho a vagoni... | [Italian, presente] |
| Toma una rana de chocolate, tengo muchísimas... | [Spanish, presente] |
| (13) They're the only family he has left now. | [English, present perfect] |
| Sie sind die Einzigen aus der Familie, die ihm noch geblieben sind. | [German, Perfekt] |
| Ze zijn de enige familie die hij nog heeft. | [Dutch, ott] |
| C'est la seule famille qui lui reste désormais. | [French, présent] |
| Sono gli unici parenti che gli rimangono. | [Italian, presente] |
| Son la única familia que le queda ahora. | [Spanish, presente] |

Furthermore, we see the German *Präsens* and Dutch *ott* cluster taking over some of the *simple future* cluster. From the literature, we know the German and Dutch PRESENT exhibit more futurate uses than the English *simple present* (Hilpert, 2008). The semantic maps clearly show this option.

Finally, the German and Dutch PERFECT clusters overlap with the right *present perfect* cluster. Typical examples we find here are (1), (10), and (11): contexts in which all languages align on PERFECT use.

However, the German and Dutch PERFECT also extend into the *simple past* domain. What is more, the German *Perfekt* cluster extends further than the Dutch *vtt* cluster. This is reminiscent of the implicational hierarchy of PERFECT use that was established in (van der Klis et al., 2019). Drilling down to the data points, we indeed find similar cut-off points between PERFECT and PAST use. For example, in (14) the German *Perfekt* used for sequences of events, whereas both English and Dutch resort to a PAST.

- (14) Professor Quirrell **did** not **manage** to take it from you. I **arrived** in time to prevent that, although you were doing very well on your own, I must say. [English, simple past]
Professor Quirrell **ist** es nicht **gelingen**, dir den Stein abzunehmen. Ich **bin** rechtzeitig **dazugekommen**, um dies zu verhindern, obwohl du dich auch allein sehr gut geschlagen hast, muss ich sagen. [German, Perfekt]
Professor Krinkel **slaagde** er niet in om die af te pakken. Ik **arriveerde** net op tijd om dat te verhinderen, hoewel ik moet zeggen dat je het zelf ook niet slecht deed. [Dutch, ovt]

In (14), uttered by Dumbledore to explain how he managed to get hold of the *Philosopher's Stone*, we see a sort of mini-narrative: Dumbledore recalls his actions. English and Dutch then return to their tense use in narrative discourse, while German allows using a PERFECT in dialogue. In de Swart (2007), the observation that English and Dutch (and Spanish) resist PERFECT use in sequences of events has also been made.

With respect to the differences between Dutch and English, Le Bruyn, van der Klis, & de Swart (2019) provide an analysis. They show that dialogue is a combination of mini-narratives like (14) above and conversation between participants that they coin *pure* dialogue. In mini-narratives, tense use of English and Dutch aligns: here, we end up with the same distribution of tense use that we have seen in the previous section for narrative discourse. In pure dialogue, however, every eventive Dutch verb phrase is yielded with the *vtt*, whereas English tends to use a *simple past*. Stative verb phrases, on the other hand, are rendered with a PAST in both languages. (15) shows an example.

- (15) You **got** there? You **got** Hermione's owl? [English, *simple past* (twice)]
Sie **waren** da? **Hat** Hedwig Sie **erreicht**? [German, *Präteritum + Perfekt*]
Was u hier? **Hebt** u Hermeliens uil **ontvangen**? [Dutch, *ovt + vtt*]

In (15), Harry immediate response to (14), we see that the stative verb phrase (*being somewhere*) is translated with a PAST in Dutch and German, whereas the eventive verb phrase (*receiving something*) is rendered with a PERFECT instead.

To conclude, in this analysis of the Germanic semantic maps, we find there are overlapping areas of PRESENT, PAST, and PERFECT use. We also find competition between the FUTURE and the PRESENT, the PERFECT and the PRESENT, and the PERFECT and the PAST.

To better assess the difference between the form-based categories and their semantic interpretation, we introduce the notion *domain of competition*. Within a domain of competition, languages share a function (e.g. reference to past events and states) but potentially opt for different tenses (e.g. the *Präteritum* vs. the *Perfekt*). Figure 9 shows a visual interpretation of the competition between Germanic tenses. The three ellipses mark the three domains of competition: present, past, and perfect. The dashed lines show tenses taking over space from the domain of competition. We see that the English *present perfect* competes with the present, while the Dutch *vtt* and German *Perfekt* both compete with the past, albeit that the German

Perfekt takes over more space from the past. Furthermore, the English *simple future* competes with the present. Note that all these competition effects can be traced back to earlier literature, showing that our straightforward form-based analysis is able to create meaningful comparisons.

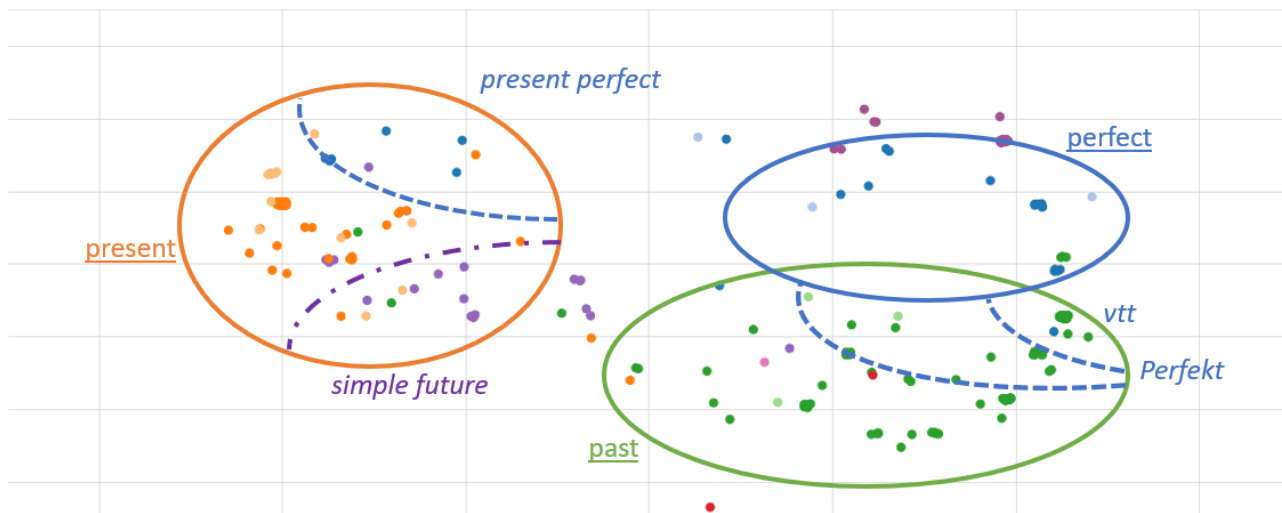


Figure 9: Idealization of the semantic maps for the Germanic languages. We focus on three areas of competition: present, past, and perfect. In the present area, we find competition with the English *simple future* and *present perfect*. In the past area, we find competition with the Dutch *vtt* and the German *Perfekt*. The three languages align on *PERFECT* use in the perfect area.

After having compared the Germanic languages, we are in a position to compare the Romance languages. We start with the French and Italian maps in Figure 10 and Figure 11 respectively.

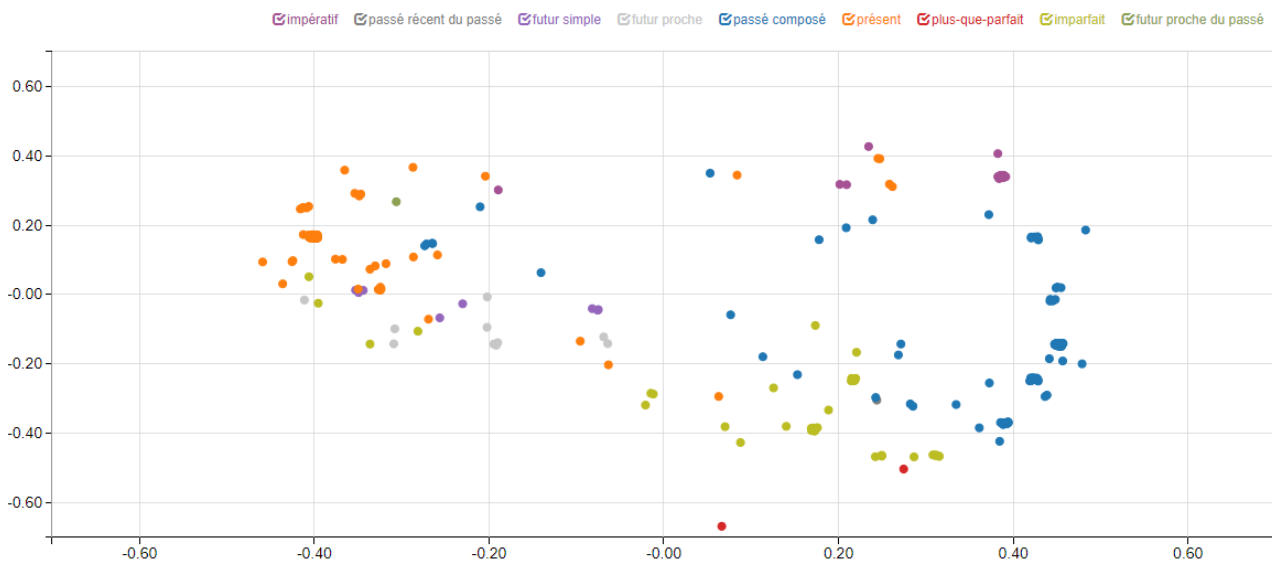


Figure 10: MDS solution for dialogue in chapters 1 and 17, coloured with the French TA-labels.

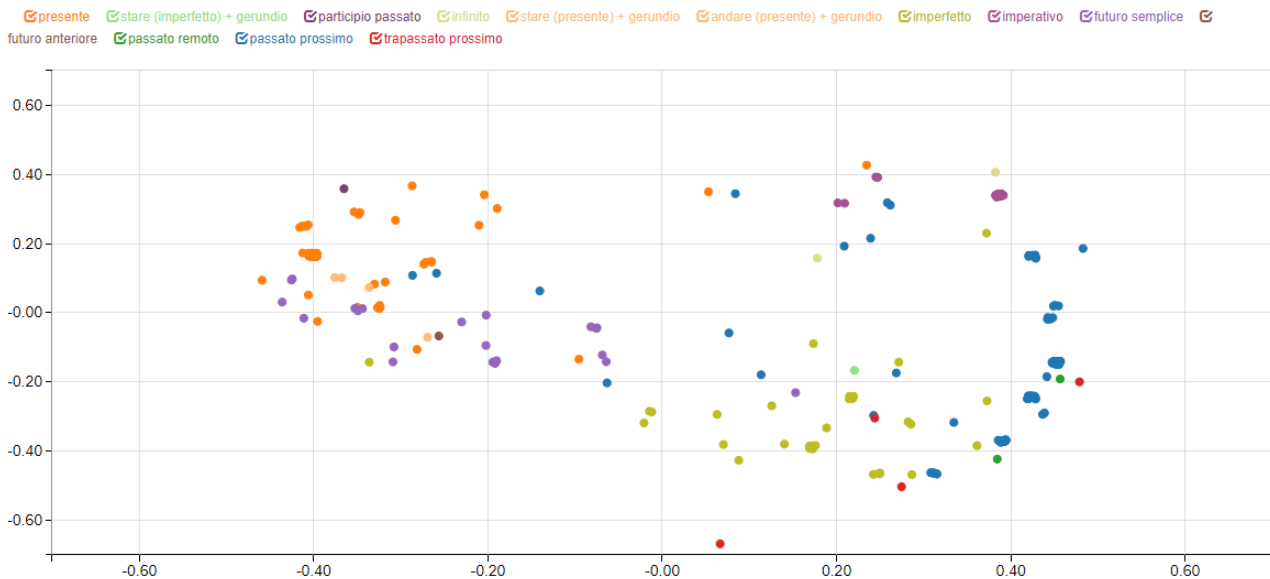


Figure 11: MDS solution for dialogue in chapters 1 and 17, coloured with the Italian TA-labels.

When comparing the French and Italian maps to the Germanic maps, we see that while the present area stays largely intact, the French and Italian PERFECT extends even further into the domain of competition that we in Figure 9 had interpreted as the past domain of competition. Apart from some minor differences, French and Italian overlap in their PERFECT use. The remainder of the past space turns lime green, signalling the use of the IMPERFECTIVE PAST. Again, the French *imparfait* and the Italian *imperfetto* overlap in their use.

When we consider the data points in which French and Italian use a PERFECT, but German does not, we primarily find cognitive state verbs like *to feel* in (16) below. All Germanic languages use a PAST in translation.

(16) Madam Pomfrey, however, **felt** it might not be very hygienic, and confiscated it.

[English, *simple past*]

Madam Pomfrey jedoch **meinte**, er sei vielleicht nicht besonders hygienisch, und hat ihn beschlagnahmt.

[German, *Präteritum*]

[H]elaas **leek** het madame Plijster niet erg hygiënisch en daarom heeft ze hem in beslag genomen.

[Dutch, *ott*]

Mais Madame Pomfresh **a trouvé** que ce ne serait peut-être pas très hygiénique et elle l'a confisqué.

[French, *passé composé*]

Ma Madama Chips non l'**ha giudicata** una cosa molto igienica, e quindi l'ha confiscata.

[Italian, *passato prossimo*]

Now, let us turn to the Spanish map, displayed in Figure 12. When we compare the semantic map to the French and Italian one, we find that the Spanish *pretérito imperfecto* overlaps with the French and Italian IMPERFECTIVE PAST. However, the Spanish *pretérito perfecto compuesto* does not extend into the past area of competition. Rather, the *pretérito indefinido* takes over this area completely. In that sense, Spanish is comparable to English: these languages share a strict PERFECT and use the PERFECTIVE PAST for cases in which French and Italian use the PERFECT. German and Dutch use a PERFECT in subsets of these areas.

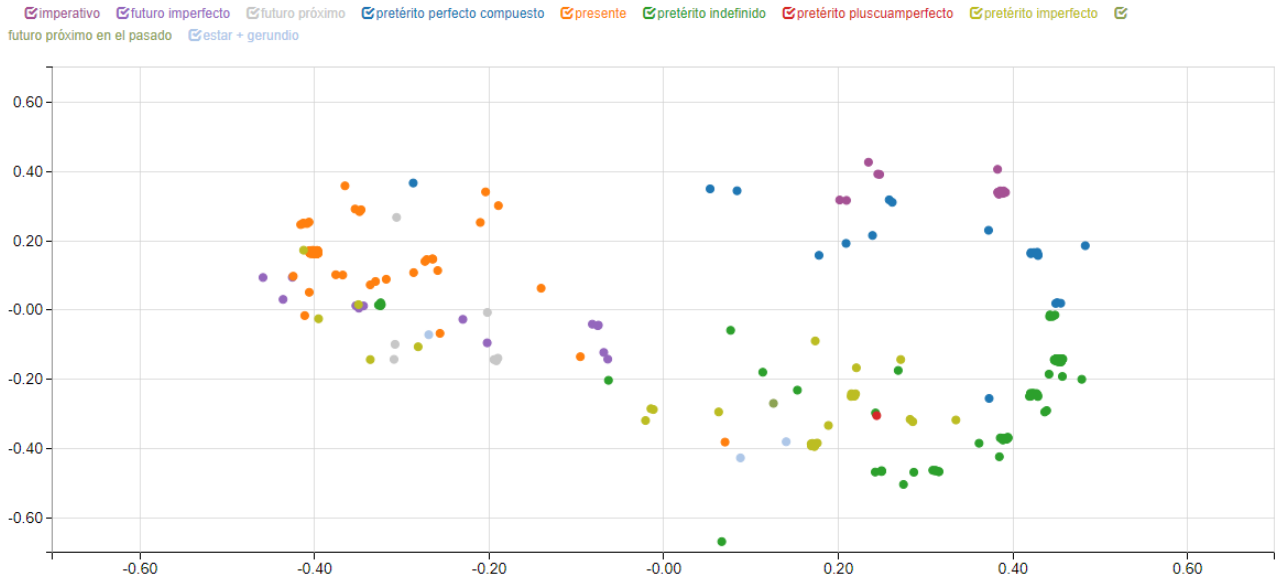


Figure 12: MDS solution for dialogue in chapters 1 and 17, coloured with the Spanish TA-labels.

From analyzing the Romance maps, it makes sense to further divide the past area of competition into a perfective past domain and an imperfective past domain. The Romance languages align on using an IMPERFECTIVE PAST in the imperfective past area of competition, while Germanic languages use their general-purpose PAST. We see however that the competition between tense forms takes place in the perfective past domain of competition. We reproduce most of the implicational hierarchy in (6) put forward by van der Klis et al. (2019). French and Italian have a very liberal PERFECT that has taken over the role of the PERFECTIVE PAST. Spanish and English on the other hand share a strict PERFECT with respect to reference to the past, but the English *present perfect* extends into the present domain of competition. German and Dutch take middle positions on the spectrum. In Figure 13, we find an idealization of this competition.

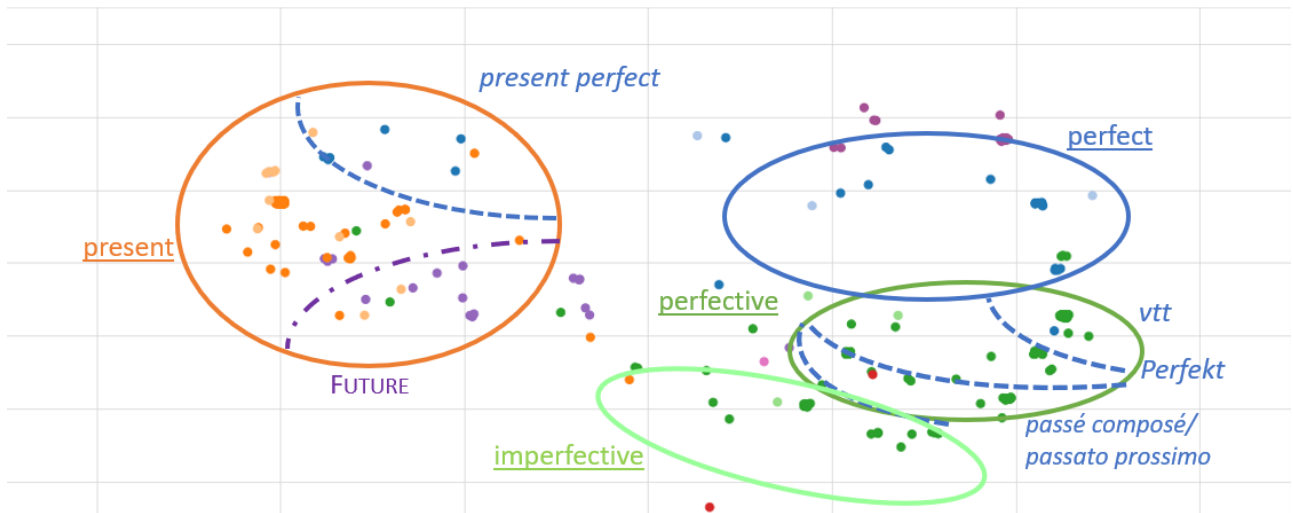


Figure 13: Idealization of the semantic maps for the Germanic and Romance languages. We focus on four areas of competition: present, perfective past, imperfective past, and perfect. In the present area, we find competition with the English present perfect, as well as competition with FUTURE forms all languages. In the perfective past area, we find competition with the Dutch, German, French, and Italian PERFECT. In both the imperfective past and the perfect area, we find no competition between languages.

The only subset relation from (6) we were not able to reproduce is the relation between Spanish and English. In our current dataset, we do not find evidence to conclude the Spanish *pretérito perfecto compuesto* extends into the past domain of competition. We propose the relation between Spanish and English as an avenue for further research.

Finally, we can interpret the dimensions of variation put forward by multidimensional scaling. We see that the MDS solution on the x-axis (i.e. the first dimension of the MDS solution) has made the present domain of competition at the left side of the semantic maps, and a past domain on the right side of the map. On the y-axis (i.e. the second dimension) the separation is less clear, but one could hypothesize IMPERFECT uses on the bottom of the solution are split from PERFECT uses at the top.

From this section and the previous, we conclude that Germanic and Romance languages do not differ in temporal reference in narration, but that there is a lot of variation in tense use in dialogue. The PERFECT plays an important role here: while the PERFECT is absent in narrative discourse, in dialogue the PERFECT is the main ingredient for competition between the PRESENT (for English) and the PAST (for all other languages). We also find a shared space of PERFECT use between languages. In the next section, we conclude our research and discuss some of its potential shortcomings.

5. Conclusion and discussion

A cross-linguistically validated account of the PERFECT requires a deeper insight into the dynamic role of this verb form. Based on preliminary remarks in the literature, we hypothesized differences in tense use between written and spoken language. Because cross-linguistic research on spoken language is hard to carry out, this paper uses the distinction between narrative discourse and dialogue as a proxy. In this paper, we have analyzed tense use in narrative discourse and dialogue in a range of Germanic and Romance languages. The data come from a parallel corpus of translations of *Harry Potter and the Philosopher's Stone*, which has been annotated for tense in six languages: three Germanic (English, German, and Dutch) and three Romance (French, Italian, and Spanish).

The prediction that tense use differs significantly between narrative discourse and dialogue is borne out for all languages. The Romance languages confirm distribution of labour between the PERFECTIVE and IMPERFECTIVE PAST that we know from the literature on temporal reference in narrative discourse. As expected, Germanic languages use their general-purpose (SIMPLE) PAST. PERFECT use is absent in narrative discourse, and PRESENT use is only found in a small number of cases. In contrast, dialogue has a more present orientation, which is marked by a large number of PRESENT uses, as well as the appearance of the IMPERATIVE, the FUTURE and the PERFECT.

In the second half of the paper, we used the *Translation Mining* methodology to detect cross-linguistic variation in tense use in dialogue. We found that languages mostly share tense use with respect to reference to present events, but differ in reference to past events. There is a shared space of PERFECT use that consists of events with current relevance or long-lasting states, but we see that the French, Italian, German, and Dutch PERFECT extend into the past domain of competition, whereas the English *present perfect* takes over some of the present area of competition. With these results, we were largely able to reproduce the implicational hierarchy of PERFECT use (van der Klis et al., 2019). Figure 14 summarizes our cross-linguistic perspective on the PERFECT.

← <u>past</u>		<u>present</u> →	
events in sequence	single events	core PERFECT: events with current relevance, long-lasting states	states continuing into the present
French, Italian, German			
	Dutch		
		Spanish	
		English	

Figure 14: The cross-linguistic perspective on the PERFECT that we arrive at in the dialogues of Harry Potter and the Philosopher's Stone. All languages share a common use of the PERFECT. French, Italian, and German share a PERFECT that fully extends into the past domain of competition. The Dutch vtt also extends toward the past but resists the use of the vtt in narrative sequences. The English present perfect on the other hand extends into the present domain of competition. The Spanish *préterito perfecto compuesto* does not extend into either domain of competition and can be considered a strict PERFECT.

We have argued a parallel literary corpus is a controlled environment to test claims on cross-linguistic variation as well as variation in tense use between written and spoken discourse. Still, our approach has its disadvantages. For example, dialogue in novels usually is idealized: there is usually no stalling by speakers or backchanneling by listeners. Furthermore, copyright restrictions are problematic. While we consider our research adhering to fair use policies, distribution of parallel corpora of popular novels might prove hard.

In terms of methodology, *Translation Mining* facilitated interpretation of the patterns by allowing a back-and-forth between overall patterns and drilling down to the data. Of course, usual caveats of parallel corpus research apply. For example, careful analysis of individual items is required to check for e.g. optionality or changes in the use of the verb, which in turn are responsible for the use of a different tense. However, since multidimensional scaling takes a perspective across multiple languages, these contexts still end up near interpretable sections of the semantic map. We therefore conclude our form-driven research methodology is suitable for detecting stability as well as cross-linguistic variation.

References

- Asgari, E., & Schütze, H. (2017). Past, Present, Future: A Computational Investigation of the Typology of Tense in 1000 Languages. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (Vol. 2, pp. 113–124). Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D17-1011>
- Bertinetto, P. M. (1986). *Tempo, aspetto e azione nel verbo italiano. Il sistema dell'indicativo*. Florence: Accademia della Crusca.
- Bick, E. (2014). The grammatical annotation of speech corpora. In T. Raso & H. Mello (Eds.), *Spoken Corpora and Linguistic Studies* (pp. 105–128). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/scl.61.04bic>
- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Cartoni, B., & Meyer, T. (2012). Extracting directional and comparable corpora from a multilingual corpus for translation studies. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)* (pp. 2132–2137).
- Centineo, G. (2016). Tense switching in Italian: the alternation between *passato prossimo* and *passato*

- remoto in oral narratives. In S. Fleischman & L. R. Waugh (Eds.), *Discourse Pragmatics and the Verb: The Evidence from Romance*. London: Routledge.
- Christodouloupoulos, C., & Steedman, M. (2015). A massively parallel corpus: the Bible in 100 languages. *Language Resources and Evaluation*, 49(2), 375–395. <https://doi.org/10.1007/s10579-014-9287-y>
- Dahl, Ö., & Velupillai, V. (2013). The Perfect. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <https://wals.info/chapter/68>
- Dahl, Ö., & Wälchli, B. (2016). Perfects and iamitives: two gram types in one grammatical space. *Letras de Hoje*, 51(3), 325–348. <https://doi.org/10.15448/1984-7726.2016.3.25454>
- de Swart, H. (2007). A cross-linguistic discourse analysis of the Perfect. *Journal of Pragmatics*, 39, 2273–2307. <https://doi.org/10.1016/j.pragma.2006.11.006>
- de Vries, L. (2007). Some remarks on the use of Bible translations as parallel texts in linguistic research. *STUF - Sprachtypologie Und Universalienforschung*, 60(2), 148–157. <https://doi.org/10.1524/stuf.2007.60.2.148>
- Díaz Cintas, J., & A. Remael, A. (2014). *Audiovisual Translation: Subtitling*. London/New York: Routledge.
- Engel, D. M. (1990). *Tense and text: a study of French past tenses*. London/New York: Routledge.
- Gadet, F. (1999). La langue française au XXe siècle: L'émergence de l'oral. In J. Chaurand (Ed.), *Nouvelle histoire de la langue française* (pp. 583–671). Paris: Seuil.
- Gellerstam, M. (1986). Translationese in Swedish novels translated from English. In L. Wollin & H. Lindquist (Eds.), *Translation Studies in Scandinavia* (pp. 88–95). Lund: CWK Gleerup.
- Harris, M. (1982). The “past simple” and the “present perfect” in Romance. In N. Vincent & M. Harris (Eds.), *Studies in the Romance verb: essays offered to Joe Cremona on the occasion of his 60th birthday* (pp. 42–70). London: Croom Helm.
- Haspelmath, M. (1997). *Indefinite Pronouns*. Oxford: Clarendon.
- Hennig, M. (2000). *Tempus und Temporalität in geschriebenen und gesprochenen Texten*. Tübingen: Niemeyer.
- Hilpert, M. (2008). *Germanic Future Constructions*. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/cal.7>
- Hinrichs, E. (1986). Temporal anaphora in discourses of English. *Linguistics and Philosophy*, 9(1), 63–82. <https://doi.org/https://doi.org/10.1007/BF00627435>
- Hoek, J. (2018). *Making sense of discourse*. Utrecht University. Retrieved from <https://dspace.library.uu.nl/handle/1874/370167>
- Kamp, H., & Reyle, U. (1993). *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-017-1616-1>
- Kamp, H., & Rohrer, C. (1983). Tense in Texts. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, Use, and Interpretation of Language* (pp. 250–269). Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110852820.250>
- Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. *MT Summit*, 5, 79–86.

- Kroon, M., Barbiers, S., Odijk, J., & van der Pas, S. (2019). A filter for syntactically incomparable parallel sentences. *Linguistics in the Netherlands*, 36, 147–161. <https://doi.org/10.1075/avt.00029.kro>
- Labeau, E. (2015). Il était une fois le passé simple... *Journal of French Language Studies*, 25(2), 165–187. <https://doi.org/10.1017/S0959269515000095>
- Le Bruyn, B., van der Klis, M., & de Swart, H. E. (2019). The Perfect in dialogue. *Linguistics in the Netherlands*, 36, 162–175. <https://doi.org/10.1075/avt.00030.bru>
- Lepschy, A. L., & Lepschy, G. (1981). *La lingua italiana: storia, varietà dell'uso, grammatica*. Milan: Bombiani.
- Levshina, N. (2016). Verbs of letting in Germanic and Romance languages: A quantitative investigation based on a parallel corpus of film subtitles. *Languages in Contrast*, 16(1), 84–117. <https://doi.org/10.1075/lic.16.1.04lev>
- Levshina, N. (2017). Online film subtitles as a corpus: an n -gram approach. *Corpora*, 12(3), 311–338. <https://doi.org/10.3366/cor.2017.0123>
- Lison, P., & Tiedemann, J. (2016). OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.
- Löbner, S. (2002). Is the German Perfekt a Perfect Perfect? In I. Kaufmann & B. Stiebels (Eds.), *More than Words: A Festschrift for Dieter Wunderlich* (pp. 369–391). Berlin: Akademie Verlag. <https://doi.org/10.1515/9783050081274-015>
- Nishiyama, A., & Koenig, J.-P. (2010). What is a perfect state? *Language*, 86(3), 611–646. <https://doi.org/10.1353/lan.2010.0014>
- Partee, B. H. (1984). Nominal and temporal anaphora. *Linguistics and Philosophy*, 7(3), 243–286. <https://doi.org/10.1007/BF00627707>
- Portner, P. (2003). The (temporal) Semantics and (modal) Pragmatics of the Perfect. *Linguistics and Philosophy*, 26, 459–510. <https://doi.org/10.1023/A:1024697112760>
- Ritz, M.-E. (2012). Perfect Tense and Aspect. In R. I. Binninck (Ed.), *The Oxford Handbook of Tense and Aspect* (pp. 881–907). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195381979.013.0031>
- Rohlf, G. (1966). *Grammatica storica dell'italiano e dei suoi dialetti*. Turin: Einaudi.
- Schaden, G. (2009). Present perfects compete. *Linguistics and Philosophy*, 32(2), 115–141. <https://doi.org/10.1007/s10988-009-9056-3>
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing*. Manchester.
- Stolz, T. (2007). Harry Potter Meets Le Petit Prince - On the Usefulness of Parallel Corpora in Crosslinguistic Investigations. *STUF - Sprachtypologie Und Universalienforschung*, 60(2), 100–117. <https://doi.org/10.1524/stuf.2007.60.2.100>
- Tiedemann, J. (2003). *Recycling Translations: Extraction of Lexical Data from Parallel Corpora and their Application in Natural Language Processing*. Uppsala University. Retrieved from <http://uu.diva-portal.org/smash/record.jsf?pid=diva2:163715>
- Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the 8th International*

Conference on Language Resources and Evaluation (LREC 2012) (pp. 2214–2218).

- van der Klis, M. (2018). Deriving anteriority in the perfect of recent past. *Linguistics in the Netherlands*, 35, 35–50. <https://doi.org/10.1075/avt.00003.kli>
- van der Klis, M., Le Bruyn, B., & de Swart, H. (2017). Mapping the Perfect via Translation Mining. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (Vol. 2, pp. 497–502).
- van der Klis, M., Le Bruyn, B., & de Swart, H. (2019). De la sémantique des temps verbaux à la traductologie: une comparaison multilingue de *L'Étranger* de Camus. In E. Corre, D.-T. Do-Hurinville, & H.-L. Dao (Eds.), *Linguistic approaches to Tense, Aspect, Modality, Evidentiality, based on the Novel L'Étranger ("The Stranger") by Albert Camus, and its Translations*. Amsterdam: John Benjamins Publishing Company.
- Varga, D., Halacsy, P., Kornia, A., Nagy, V., Nemeth, L., & Tron, V. (2005). Parallel corpora for medium density languages. In G. Angelova, K. Bontcheva, R. Mitkov, N. Nicolov, & N. Nikolov (Eds.), *Proceedings of the International Conference of Recent Advances in Natural Language Processing* (pp. 590–598). Borovets, Bulgaria.
- Wälchli, B., & Cysouw, M. (2012). Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics*, 50(3), 671–710. <https://doi.org/10.1515/ling-2012-0021>
- Wide, C. (2002). *Perfect in dialogue*. University of Helsinki.

Appendix A: Bibliographical information of used translations

- British English (original): Rowling, J.K. (1997). *Harry Potter and the Philosopher's Stone*. London: Bloomsbury.
- German: Rowling, J.K. (1997/1998). *Harry Potter und der Stein der Weisen*. (Klaus Fritz, Trans.). Hamburg: Carlsen Verlag.
- Dutch: Rowling, J.K. (1997/1998). *Harry Potter en de Steen der Wijzen*. (Wiebe Buddingh', Trans.). Amsterdam: de Harmonie.
- French: Rowling, J.K. (1997/1998). *Harry Potter à l'école des sorciers*. (Jean-François Ménard, Trans.). Paris: Éditions Gallimard Jeunesse.
- Italian: Rowling, J.K. (1997/1998). *Harry Potter e la Pietra Filosofale*. (Marina Astrologo, Trans.). Milan: Salani.
- Castilian Spanish: Rowling, J.K. (1997/1999). *Harry Potter y la piedra filosofal*. (Alicia Dellepiane Rawson, Trans.). Barcelona: Ediciones Salamandra.