The Gravitational Pull Hypothesis and imperfective/perfective aspect in Catalan translated and non-translated literary texts

Josep Marco Borillo (Universitat Jaume I, Spain)

Gemma Peña Martínez (Universitat Politècnica de València, Spain)

This article aims to test the Gravitational Pull Hypothesis on the imperfective/perfective aspect distinction in the language pairs English-Catalan and French-Catalan. It draws on the corresponding corpora in COVALT. The GPH posits three cognitive causes of translational effects: source or target language salience and connectivity. Different configurations of these causes, or factors, are expected to result in over- or under-representation of target language features. The imperfective/perfective aspect distinction was chosen as a testing ground for the GPH because it is morphologically marked in Catalan and French but not in English. That may give rise to different configurations of factors and, therefore, to different translational effects. It is predicted that the preterite, which conveys perfective aspect in Catalan, will be over-represented in Catalan translations from English as compared to translations from French and to Catalan non-translations. On the other hand, the imperfect, which conveys imperfective aspect, will be under-represented. Results confirm these predictions. For translations from French, both adherence to the patterns observed in Catalan non-translations and over-representation of the preterite are possible outcomes. Results lend support to the second alternative – over-representation of the preterite. These results highlight the importance of relying on frequency *and* other sources of evidence when formulating hypotheses in the framework of the GPH. Research from the field of second language acquisition proved particularly significant in this respect.

Keywords: Gravitational Pull Hypothesis; imperfective/perfective aspect; literary translation; COVALT; Catalan/English/French

1. Introduction

The aim of this article is to test out the revised Gravitational Pull Hypothesis (GPH) on the imperfective/perfective aspect distinction. The study draws on the

English- and French-Catalan sub-corpora (EN-CA and FR-CA, respectively) and the Catalan non-translation component (CA) in the *Corpus Valencià de Literatura Traduïda* ('Valencian Corpus of Translated Literature', COVALT). The outline of the article is as follows. Section 2 briefly presents the main tenets of Halverson's GPH and the evidence for it available so far. Section 3 deals with the notion of verbal aspect and more particularly with the imperfective/perfective aspect distinction, both in general and in the languages concerned in this study. Section 4 offers a detailed account of the methodology employed, and Section 5 presents and discusses results. Section 6 wraps up the study with some concluding remarks.

2. The Gravitational Pull Hypothesis

The GPH was put forward by Halverson (e.g. 2003, 2017) as an attempt to provide a cognitive account for so-called *translation universals* – more modestly referred to in recent years as features or properties of translation (see e.g. Lapshinova-Koltunski, 2015: 95). On a more concrete level, the hypothesis brings together two such properties, namely over- and under-representation of target language typical features. It posits three potential causes of translational effects: patterns of target language salience (factor 1 – magnetism); patterns of source language salience (factor 2 – gravitational pull); and patterns of connectivity, which reflect relationships between the source and target languages (factor 3 – connectivity). One effect is predicted for each potential cause, or factor. The effect of factor 1 will be over-representation; the effect of factor 2 will be over-representation too; and the effect of factor 3 may be over- or under-representation. Evidence has been found in the literature for both over- and under-representation, but it is clear that they are opposing features that cannot hold for all cases or all the time for a single case. Therefore, different configurations of the three factors mentioned will account for the prevalence of over- or under-representation for a given linguistic feature in a given language combination.

From a theoretical perspective, the GPH draws on two main sources: Langacker's Cognitive Grammar, and bilingual theory (see Halverson 2003 for a full account). Empirical research focusing on the GPH features studies by Halverson herself (2017), Hareide (2017a, 2017b), Vandevoorde (2020), Marco (2021), Lefer and De Sutter (2022) and a few others. While it is impossible to offer a thorough account of this body of research within the scope of this article, a brief summary is provided in the following paragraphs.

The first empirical study aiming to test the GPH on corpus data is Hareide's PhD dissertation, which later developed into two book chapters (2017a, 2017b). Hareide draws on two parallel corpora with the same target language, and focuses on an element or structure that can be regarded as unique (in Tirkkonen-Condit's sense, 2004) in one of the two language pairs but not in the other. The language combinations in her study were Norwegian-Spanish and English-Spanish, and the constructions chosen were the Spanish gerund and the *estar* + gerund construction. Results confirm the author's predictions in most but not all cases. Halverson (2017) pushed the theory forward by re-naming the three factors and using a mixed-methods approach in which salience was established on the basis of corpus data *and* data from other sources, such as elicitation tests and keylogging. Her aim was to test her hypothesis on the bilingual semantic networks

created by the different senses of the English verb *get* and its two most frequent equivalents in Norwegian. It was predicted that some senses of *get* would be over-represented whereas others would be under-represented, but only some of these predictions were confirmed by the data. Halverson (2017: 37) suggested as a possible explanation that "close cross-linguistic similarity leads to translational patterns that closely match the original English figures".

The two studies just referred to are, among other things, empirical tests of the hypothesis. But there are other pieces of research that use the GPH as an interpretative tool. A case in point is Vandevoorde (2020), who set out to explore semantic patterns in the area of inchoativity across the translation/non-translation divide. She identified a number of effects which could be subsumed under the concepts of shining-through, levelling out and normalization, and found a plausible explanation in one of the three factors in the GPH, namely connectivity. Another such study, though in a different direction, is Lefer and De Sutter (2022), which uses the three factors in the GPH as variables in a complex research model whose aim is to determine to what extent English concatenated nouns in European Parliament speeches are rendered with semantically equivalent terms both in translation and interpreting. Magnetism and connectivity are seen to impact translation solutions, whereas no evidence is found for gravitational pull. The study purports to be a step forward in the application of the GPH because it focuses on a linguistic phenomenon beyond the word level, and also because its "robust multifactorial statistics" (2022: 151) make it possible to tease apart the effects of each variable.

Finally, a host of studies on the GPH are currently being conducted by the COVALT group (Universitat Jaume I, Castelló, Spain) in the framework of a state-funded research project. Most of these studies are still in press or in preparation, and focus on such indicators as the passive voice, Light Verb Constructions, diminutive suffixes, deontic modality and verbal aspect, as in this article. These studies tend to concentrate on grammatical constructions, not lexical items, and are thus in line with Hareide's approach. Marco (2021) may serve as a token for this line of research. This book chapter aims to test the GPH on deontic modality in Catalan (as instantiated by modal indicator *caldre* and its alternatives) and draws on data from an English-Catalan and a French-Catalan corpus, together with a corpus of non-translations in the target language. Predictions about over- or under-representation of *caldre* based on different configurations of factors for each language pair are mainly confirmed.

3.   The imperfective/perfective aspect distinction

Aspect and tense are quite distinct verbal categories. Pérez Saldanya (2002: 2577) describes the difference as follows: "Aspect, as opposed to tense, is not a deictic category but a category related to the internal temporality of the represented situation, and more specifically to the way this situation is viewed […] and the part of this situation that is referred to".[1] One of the main aspectual distinctions to be found in many languages (Comrie, 1976; Binnick, 1991) is the one between imperfective and perfective aspect, which is spelt out by Binnick (1991: 157-158) as follows: "the perfective essentially presents the action as a whole, while the

---

[1] Unless otherwise indicated, all translations from languages other than English are our own.

imperfective essentially indicates only the lack of any such presentation. This may be interpreted either as indifference to wholeness or as a positive lack of it". Slightly different formulations of this basic distinction are provided by different authors (e.g. Wheeler *et al.*, 1999: 343; Diaubalick, 2019: 55).

Langacker (2008: 147), for his part, distinguishes between perfective and imperfective verbs. The former are "bounded in time", whereas the latter are not specifically bounded. Langacker's claim that there are perfective and imperfective verbs may not be an apt description of the state of affairs in Romance languages, where the imperfective/perfective distinction pertains to the verbal paradigm, not to verbs as lexical units, since most verbs can occur in both aspects.[2] However, it does capture a concept that intersects with verbal aspect, even if it is independent from it – lexical aspect. Lexical aspect is to a certain extent inherent to the verb's meaning, but it is more accurate to say that it stems from the verb and its arguments, rather than from the verb in isolation. The best-known account of lexical aspect is perhaps Vendler's (1967), who classified events into four types: states, activities, accomplishments and achievements. States and activities are atelic, i.e. they have no inherent end-point, whereas accomplishments and achievements are telic. States differ from activities in that they need no input of energy (they are stative, as in *My brother is tall*); activities are dynamic, with no end-point envisaged, as in *Yesterday I walked in the park*. On the other hand, accomplishments differ from achievements in that they have duration (e.g. *Cross the street*) whereas achievements are punctual (e.g. *The train arrived on time*).

This classification, according to Salaberry and Ayoun (2005: 7), is based on three semantic features: dynamicity, telicity and durativity. A kind of probabilistic relationship may be posited for each of these features and the imperfective/perfective aspect distinction (see e.g. Diaubalick, 2019: 63). Verbal predicates at the minus end of the dynamicity cline (i.e. those designating states) will tend to occur as imperfectives, whereas those at the plus end might not show any marked preference. Verbal predicates designating telic events (especially achievements, as they have no duration) will be protypically realised as perfectives. Finally, durativity will favour realisation as an imperfective, especially if there is no telicity (activities). Let it be clearly understood that these prototypical relationships are put forward as probabilistic tendencies, not bound by any sort of strict causation.

As seen in section 2, salience is all-important when it comes to hypothesis formulation against the backdrop of the GPH. However, in the light of what was said in the previous paragraph, it may not make sense to talk about differences in salience across the perfective/imperfective distinction. If some verbs tend to occur in the imperfective whereas others are prototypically realised as perfectives, the notion of salience would only make sense *within* a particular verb, not *across* verbs. However, without denying the general validity of the latter claim, there are reasons to assume that, in the Romance languages, the perfective is more salient than the imperfective across the board, as we will try to show in the following paragraphs.

The strongest evidence for the relative salience of the perfective can arguably be established in ontogenetic terms, as is furnished by research on the

---

[2] In fairness to Langacker, it must be added that his initial claim is later qualified (2008: 148): "the perfective/imperfective contrast is anything but a rigid lexical specification. While it is usual for a verb to have a primary classification as either perfective or imperfective, many verbs are comfortably used both ways".

acquisition of aspect (or of the verbal system) in L2 settings. There is a considerable body of literature on the subject, adequately summarized by Salaberry and Ayoun (2005) or Diaubalick (2019), amongst others. Salaberry and Ayoun (2005) list as many as six different hypotheses on the emergence and acquisition of aspect in an L2, which reflect as many underlying (and often not incompatible) theoretical approaches. We will here refer to two of them, which seem to converge to a great extent, as far as our interests are concerned.

The hypothesis that has attracted the most attention is perhaps the Lexical Aspect Hypothesis (LAH), which applies Vendler's categories to the study of aspect acquisition in L2 learners. As originally posited by Andersen (1991), the LAH proposes eight stages of development for the acquisition of L2 Spanish and

> predicts that perfective markers will appear first and spread from punctual verbs (when achievements are first marked with preterite in stage 2) to stative verbs, whereas the use of imperfective markers will appear later and spread from stative verbs (starting during stage 3) to punctual verbs. (Salaberry and Ayoun, 2005: 15)

The LAH has found empirical support, but a number of issues have been raised, which cannot be discussed here. An alternative account to the LAH is the Discourse Hypothesis, according to which "past tense verbal morphology is highly influenced by contextual factors above the sentence level such as text type, and, especially, narrative grounding" (Salaberry and Ayoun, 2005: 16). Narrative grounding accounts for the fact that, in a narrative, some events are foregrounded whereas others are part of the background. Foregrounded events move the story forward, whereas background events provide support information. These two levels of texture in a narrative have been linked to the cognitive concepts of *figure* and *ground* (e.g. Langacker, 2008: 58), which generate elements and structures with varying degrees of salience. Thus, according to Wallace (1982; in Comajoan, 2005: 43), "perfective, transitive, actional, and foreground verbal forms are more salient than nonperfective, intransitive, stative, and background verbal forms". Research on the acquisition of aspect by L2 learners from this theoretical perspective shows that the preterite emerges before the imperfect and is more common in the foreground. As learners progress, they start using the imperfect in the background.[3] As remarked by Salaberry and Ayoun (2005: 28), semantic and discourse factors overlap to a large extent, as telic events (i.e. accomplishments and achievements) tend to be foregrounded, whereas atelic events (states and activities) tend to be part of the background. So, insofar as this assumption is true, the LAH and the Discourse Hypothesis are fully compatible with each other.

In the research reported on in this paper, the imperfective/perfective aspect distinction was chosen as a testing ground for the GPH because it is grammaticalised in Catalan and French (and in the Romance languages in general) to a larger extent than it is in English (or in the other Germanic languages). This contrast is best seen in the past tenses.

Catalan has two simple tenses with past reference: the preterite, which can be synthetic (e.g. *cantà* 'he/she sang') or periphrastic (*va cantar* 'he/she sang'), and the imperfect (*cantava* 'he/she sang/was singing'). The preterite "combines

---

[3] A more fine-grained account than is possible here would reflect, though, all kinds of nuances. For such an account, see Salaberry and Ayoun (2005), Comajoan (2005) or Diaubalick (2019).

past time with perfective aspect; in expressing perfective aspect, the preterite contrasts with the imperfect" (Wheeler *et al.*, 1999: 343).

French also has two simple tenses with past reference: the *passé simple* (e.g. *chanta* 'he/she sang') and the imperfect (*chantait* 'he/she sang/was singing'). The *passé simple* is the tense prototypically used in *récit* (narrative; see e.g. Maingueneau and Salvador, 1995: 42). It has all but vanished from oral discourse, where the *passé composé* (*a chanté* 'he/she has sung/sang') is used instead. It presents an event from a synthetic viewpoint as bounded and not linked to the present. The *imparfait*, on the other hand, offers an internal view of the event with no consideration of temporal bounds. In that respect, it may be said to present a high degree of isomorphism with the Catalan imperfect. However, there is a very important difference between the French and the Catalan systems, which has to do with the ambiguity of the French *passé composé*. In Catalan, the verbal forms signalling perfective aspect are clearly distinguished from the one signalling perfect aspect (*ha cantat* 'has sung'), which "is an absolute hodiernal tense" (Comajoan, 2005: 40). The French *passé composé*, however, straddles those two meanings (the perfect and the perfective), thus blurring the distinction (see e.g. Grevisse and Goosse, 2008: 1092; Riegel *et al.*, 2004: 297).[4] We will take up this issue later in this section.

The perfective/imperfective distinction is absent from the English verbal system, where the only simple tense with past reference is the simple past. Comrie (1976: 25) argues that English has a separate habitual aspect (*used to*, as in *He used to be my friend*) and a separate progressive aspect (past continuous, as in *He was trying to be friendly*); otherwise there is only the simple past, with no indication of aspect. But it might be added that even the two aspects mentioned by Comrie (habitual and progressive, which are usually part of the imperfective in languages making the perfective/imperfective distinction) are not encoded to the same degree in English as the imperfective in the Romance languages. In the latter they are *morphologically* encoded, whereas such forms as *used to* and the past continuous in English might be regarded as no more than verbal periphrases in other languages.

The contrastive issue arising from the previous brief descriptions is explained by Wheeler *et al.* (1999: 348-349) as follows with regard to the English-Catalan pair: "The aspectual distinction is elusive for English speakers, mainly because the English simple past-tense form ('I went', 'I gave', etc.) can cover both". And its implications for translation are spelt out at length by Ainaud *et al.* (2003: 172):

> when faced with a simple past, a translator into Catalan is always forced to choose one of two possible aspectual interpretations – the imperfective or the perfective one. The linguistic context usually makes it clear whether the action or state in question is viewed from a perfective or an imperfective standpoint.

Then the authors go on to mention the three elements that help disambiguate aspectual meaning: lexical aspect itself, temporal expressions co-occurring with

---

[4] "[L]e passé composé a une double valeur, puisqu'il peut marquer l'accompli du présent ou concurrencer le passé simple pour dénoter un fait passé" (Riegel *et al.*, 2004: 297).

the verb and arguments assigned to the verb (under the shape of subject and complements, amongst others). In spite of all these clues, some ambiguity may survive in verbal actions, and the translator needs to cope with it; and the differing degrees of salience of each aspect across the board may lead the translator to choose the most salient one more often than they might when *writing* in the target language, as opposed to translating.

It is quite revealing that in French-Catalan translation textbooks the imperfective-perfective aspectual distinction is hardly an issue. Verdegal (2011), for example, provides a rich overview of problems facing the French-Catalan translator but does not even mention this aspectual distinction. It is implicitly assumed, then, that the degree of isomorphism between the two verbal systems in this respect is high enough to make an explicit approach unnecessary.

4. Methodology

As stated at the beginning, this study draws on the English-Catalan (EN-CA) and the French-Catalan (FR-CA) sub-corpora of COVALT – a multilingual corpus initially made up of the translations into Catalan of narrative works originally written in English, French, and German published in the autonomous region of Valencia from 1990 to 2000, together with their corresponding source texts. A comparable component of Catalan non-translations (CA) was later added. COVALT currently also includes three parallel corpora with Spanish as target language and a comparable component of Spanish non-translations.[5] Table 1 provides details on the size of the corpus components used in this study. All the parallel corpora are sentence-aligned, lemmatized and annotated for part of speech. FreeLing was used to lemmatize and pos-tag the Catalan components of the corpora, whereas TreeTagger was used for the remaining languages (Spanish, English, French and German).

Table 1. Size of the corpus components used in this study

COVALT CORPUS	
Component	Size (words)
CA	1,551,521
EN-CA (ST)	1,201,757
EN-CA (TT)	1,343,631
FR-CA (ST)	551,869
FR-CA (TT)	566,998

The main methodological inspiration for this piece of research is Halverson's work (most notably, Halverson, 2017), but we also draw on Hareide (2017a, 2017b), who used two comparable parallel corpora to test for the impact of different patterns of salience and connectivity. Since French (like Catalan, unlike English) does have the imperfective/perfective distinction in the past, possible differences between the two parallel corpora can throw light on the influence of factors 2 and 3. Factor 1 will remain stable in both parallel corpora,

---

[5] These corpora were compiled at the Translation and Communication Department, Universitat Jaume I (Castelló, Spain), and can be accessed for research purposes upon request (http://www.covalt.uji.es).

as the target language is the same. Corpus structure, then, is inspired by Hareide's work.

Particular hypotheses in the framework of the GPH need to take into account the relative salience of the elements constituting a (bilingual) semantic network as well as connectivity between those elements across languages. Salience (and entrenchment, for that matter) is often operationalized as frequency in empirical studies focusing on the GPH, as can be seen in both Halverson (2017) and Hareide (2017a, 2017b), and also in Marco (2021). In fact, this kind of operationalization is found more generally in cognitively oriented studies using corpora as a source of data, as witnessed by Schmid (2010: 116), who refers to the "considerable body of evidence from psycholinguistic experiments suggesting that frequency is one major determinant of the ease and speed of lexical access and retrieval". But the same author later on observes that the correlation between frequency and cognitive significance is far from unproblematic (see Marco 2021: 38). Schmid suggests distinguishing between absolute and relative frequency, the latter referring to "frequency of use with regard to a specific meaning or function, in comparison with alternative expressions of that meaning or function" (2010: 116). Oster and Tello (in preparation) also operationalize salience as frequency plus other factors, such as treatment in normative texts (grammars, translation textbooks, etc.), which may have impacted translator socialization. Since semantic networks are not directly accessible, their representation can draw on claims made in contrastive grammars, translation textbooks or any kind of relevant literature, but it also needs to take into account frequency data, whenever available. The following paragraphs provide that kind of data for Catalan, English and French.

Two different corpora were used to establish the relative frequency of the perfective and imperfective aspects in Catalan: the Catalan non-translation component in COVALT (CA) and the Catalan Web 2014 corpus in Sketch Engine, as such a combination might be expected to throw light on the matter of salience in a complementary way. The imperfect was found to be more frequent than the preterite in CA (52,572 vs 37,180 occurrences, with normalized frequencies at 33,880 and 23,960 per million words, respectively). However, the Catalan Web 2014 corpus in Sketch Engine (182,608,420 words) yielded very different results, with 1,401,098 matches for the preterite and 837,856 for the imperfect – normalized frequencies per one million words being 7,672.69 and 4,588.26, respectively.[6] Table 2 summarizes these data, which are indeed contradictory. The contradiction might be resolved by considering that the prevalence of the imperfect over the preterite is an artefact of our corpus. Since the imperfect tends to be associated with verbs denoting atelic events (states and activities) which are not foregrounded, it might well be the case that events of this kind outnumber telic events in CA.[7] This is largely confirmed by an individual analysis of the 25 most frequent verbs in CA, which shows that some of the top-ranking verbs on the list tend to occur as imperfectives. If matches for the

---

[6] The queries inserted were [tag="VMII.*|VSII.*"] for the imperfect, [tag="VMIS.*|VSIS.*"] for the simple past and [(tag="VAIP.+")&(lemma="anar")][pos="VMN.+|VSN.+"] for the periphrastic past. The number of matches thus retrieved were 837,856, 223,595 and 1,177,503, respectively. Queries are always provided in this study for the sake of reproducibility.

[7] It must be borne in mind that COVALT is exclusively made up of narrative texts, whereas the Catalan Web 2014 corpus includes texts from a wide array of genres, text types and registers. (We are indebted to the editors of this volume for this suggestion.)

imperfective and the perfective are added together, the verb *ser* ('to be'), for example, features 9,194 occurrences, of which 85% are imperfectives. The same is true of such other verbs as *tenir* ('to have', 88% of imperfectives), *poder* ('can', 80% of imperfectives) or *saber* ('know', 84% of imperfectives). Since those verbs are extremely frequent, their occurrences account for a large proportion of the total number of tokens in either aspectual construction. More than that, they may be said to skew the results, which might be very different indeed if the analysis was carried out on the basis of (lemmatized) types, not tokens. Therefore, the prevalence of the imperfective over the perfective in CA, in terms of frequency, must be balanced against: 1) the large number of tokens of verbs tending to occur as imperfectives; 2) the fact that perfectives outnumber imperfectives in the Catalan Web 2014 corpus in Sketch Engine; and 3) empirical research on the acquisition of aspect in an L2, bearing out the assumption that the perfective is more salient than the imperfective in Romance languages. All things considered, it is assumed in this study that the perfective is more salient than the imperfective in Catalan and will exert magnetism in translations into that language.

Table 2. Raw and normalized frequencies of the Catalan imperfect and preterite in CA and the Catalan Web 2014 corpus in Sketch Engine

| | CA | | Catalan Web 2014 | |
	Raw f	Norm f (1,000,000 w)	Raw f	Norm f (1,000,000 w)
Imperfect	52,572	33,880	837,856	4,588.26
Preterite	37,180	23,960	1,401,098	7,672.69

In English, the simple past, which conveys perfective aspect, is by far the most salient verbal form with past reference, as witnessed by data from both the ST component of EN-CA and the English Web 2020 corpus in Sketch Engine (36,561,273,153 words). In the former, the aggregate result of the three different queries[8] performed amounts to 82,444 matches (68,602.89 instances per million words). Just for the sake of comparison, when queried for the past continuous,[9] which is the main alternative to the simple past as regards verbal forms with past reference, the ST component of EN-CA yields 3,411 matches (2,838.34 per million words). When the same operation is performed on the English Web 2020 corpus, the queries for the simple past yield a total of 968,034,747 matches (26,477.05 instances per million tokens) and the query for the past continuous yields 36,807,877 matches (1,006.74 instances per million tokens). These data are summarised in Table 3. Differences are huge indeed. It is assumed, therefore, that the English simple past will exert gravitational pull in EN-CA translation.

Table 3. Raw and normalized frequencies of the simple past and the past continuous in the ST component of EN-CA and the English Web 2020 corpus in Sketch Engine

| | ST component of EN-CA | | English Web 2020 | |
	Raw f	Norm f (1,000,000 w)	Raw f	Norm f (1,000,000 w)
Simple past	82,444	68,602.89	968,034,747	26,477.05
Past continuous	3,411	2,838.34	36,807,877	1,006.74

---

[8] The English texts in COVALT are annotated with TreeTagger, which has different tags for the simple past of the verb *to be* [pos="VBD"], the verb *to have* [pos="VHD"] and all other verbs [pos="VVD"].

[9] The query inserted was [word="was|were"][]{0,3}[pos="VVG|VBG|VHG"].

In French, the picture that emerges is not so clear. In the ST component of FR-CA in COVALT (as in CA, and perhaps for the same presumed reasons), the imperfect is more frequent than the *passé simple*: the former features 17,842 occurrences (32,330.14 per million words) and the latter 11,339 (20,546.54 per million words). But, apart from these two forms, account must be taken of the *passé composé*, which yields 2,090 matches (3,787.13 per million words).[10] However, as argued above, it is impossible to tell apart instances signalling perfect aspect (more or less equivalent to the English present perfect) from those signalling perfective aspect (as in the English simple past) unless a manual analysis is performed. Therefore, it is impossible to know what contribution the *passé composé* makes to perfective aspect as a whole. The number of matches for the *passé composé* is small when compared to both the *imparfait* and the *passé simple*, which suggests that its contribution cannot be too large. On the other hand, if we suspect (as was assumed in the case of CA) that this distribution is an artefact of our corpus and resort to a large general corpus of French, such as the French Web 2017 corpus in Sketch Engine (5,752,261,039 words), the problem remains. Again, in this corpus the *imparfait* prevails over the *passé simple*, the former featuring 29,803,042 occurrences (5,181.1 per million words) and the latter 16,788,184 (2,918.53 per million words). And then, the *passé composé* occurs 75,484,119 times (13,122.51 per million words), but it is impossible to determine how many of these instances are actual perfectives.[11] These data are summarised in Table 4. Therefore, it is extremely difficult to decide whether either of the two aspects (perfective or imperfective) is likely to exert gravitational pull in translation from French. On the one hand, the perfective is claimed to be the most salient aspect in Romance languages in general, as seen in Section 3; on the other, it is hard to establish relative salience on the basis of frequency data. In view of all this, it will be assumed in this study that neither the imperfective nor the perfective aspect exerts gravitational pull in FR-CA. At any rate, any gravitational pull effect stemming from the frequency distribution across the imperfective/perfective distinction in French, as described in this paragraph, would be minor when compared to the strong pull expected from the simple past in EN-CA.[12]

Table 4. Raw and normalized frequencies of the *imparfait*, the *passé simple* and the *passé composé* in the ST component of FR-CA and the French Web 2017 corpus in Sketch Engine

	ST component of FR-CA		French Web 2017	
	Raw f	Norm f (1,000,000 w)	Raw f	Norm f (1,000,000 w)

---

[10] The queries inserted for the *imparfait* and the *passé simple* were [pos="VER:impf"] and [pos="VER:simp"], respectively. As to the *passé composé*, the following query was inserted: [pos="VER:pres"& lemma="avoir"][]{0,3}[pos="VER:pper"]|[pos="VER:pres"&lemma="être"][]{0,3}[pos="VER :pper"]. Queries use the TreeTagger tagset for French.

[11] The queries inserted in the French Web 2017 corpus in Sketch Engine are [tag="V..I.*"] for the *imparfait* and [tag="V..S.*"] for the *passé simple*. For the *passé composé* a long string was inserted in order to retrieve instances of both transitive (with *avoir* as auxiliary) and intransitive (with *être* as auxiliary) verbs: ([[(tag="V.IP.*")&(lemma="avoir")][]{0,3}[tag="V.P.*"])|([[(tag="V.IP.*")&(lemma="être")][] {0,3}[tag="V.P.*"]). But the latter half of the formula yields both true and false positives – e.g. *est sortie* (the true *passé composé* of an intransitive verb) and *est mangée* (the present indicative passive of a transitive verb). That adds to the noise and creates still more indeterminacy in the matches – a difficulty that can only be solved through manual sifting, as already remarked.

[12] We are indebted to the editors of this volume for the latter observation.

Imparfait	17,842	32,330.14	29,803,042	5,181,1
Passé simple	11,339	20,546.54	16,788,184	2,918.53
Passé composé	2,090	3,787.13	75,484,119	13,122.51

Two facts receive special attention in the semantic network involving the forms conveying imperfective and perfective aspect in Catalan and their matching English forms. Firstly, an English simple past can be construed as a perfective or an imperfective event and translated into Catalan as a preterite or an imperfect, respectively; but the link between the simple past and the Catalan preterite is assumed to be stronger than that between the simple past and the Catalan imperfect. And secondly, "English constructions like 'I used to go', 'I was going', and 'I would (habitually) go' will almost always correspond to the Catalan imperfect" (Wheeler *et al.*, 1999: 346). Figure 1 is an attempt to represent that network. Thicker box lines signal salience, either in the source language (gravitational pull) or in the target (magnetism). Thicker lines between boxes signal strong connectivity, as opposed to thinner ones, which represent the opposite.
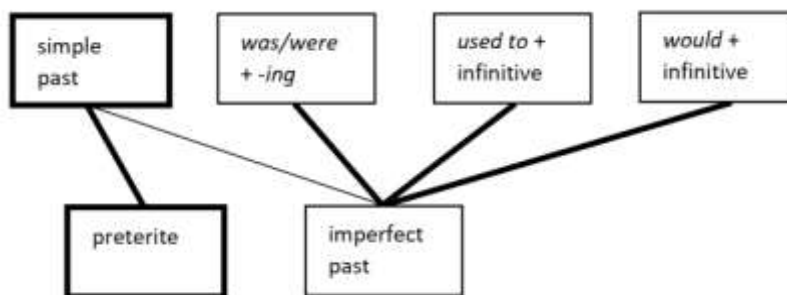

Figure 1. Network for the Catalan preterite and imperfect past and their English triggers

On the other hand, the network involving the two simple past tenses in Catalan and their matching French forms must perforce look different, as French also makes the aspectual distinction in the past, and the values assigned to the imperfective (realized by the *imparfait*) and the perfective (realized by the *passé simple* or the *passé composé*) are very similar to the values assigned to those aspects in Catalan. Therefore, a much higher degree of overlap is to be expected between the French and the Catalan systems. The bilingual network might look like that depicted in Figure 2.
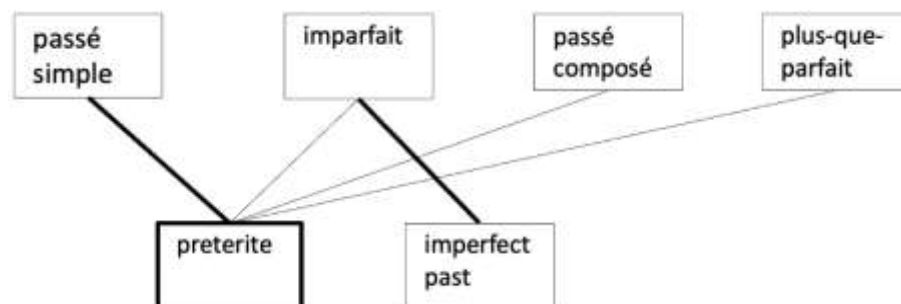

Figure 2. Network for the Catalan preterite and imperfect past and their French triggers

Such networks enable us to make the following predictions about the frequency of the two Catalan verbal constructions:

1. The Catalan preterite will be over-represented in translations from English as compared to Catalan non-translations – and as a corollary the imperfect will be under-represented.
2. The frequencies of the Catalan preterite and imperfect in translations from French will not differ significantly from the frequencies of those two verbal tenses in Catalan originals if strength of connectivity between the *imparfait* and the imperfect, on the one hand, and between the *passé simple* and the preterite, on the other, prevails over magnetism of the preterite in Catalan. Alternatively, the Catalan preterite will be over-represented if magnetism prevails over connectivity patterns.
3. The Catalan preterite will be over-represented in translations from English as compared to translations from French – and as a corollary the imperfect will be relatively under-represented.

The rationale behind these hypotheses has been carefully spelt out earlier in this section, but let us try and summarize it in terms of the three factors in the GPH. In EN-CA all three factors are assumed to pull towards over-representation of the perfective, when compared to CA: magnetism of the preterite in Catalan, gravitational pull of the simple past in English, and (allegedly) strong connectivity between both verbal constructions. In FR-CA, on the contrary, only one factor pulls towards over-representation of the preterite – magnetism. Gravitational pull does not seem to be at play, and connectivity patterns would tend to align FR-CA with CA, thus assuaging any bias towards over- or under-representation. The third hypothesis is a logical consequence of the first and the second put together: in EN-CA all three factors are assumed to pull in the same direction whereas in FR-CA magnetism and connectivity are expected to pull in opposite directions and gravitational pull is assumed to be in abeyance.

Once our particular hypotheses have been formulated in the light of the GPH, corpus querying and analysis of results comprise the following steps:

1. Data retrieval with CQPweb from EN-CA and FR-CA (starting from the TT end). The target components of these two parallel corpora are queried for verbal forms conveying perfective and imperfective aspect. Query matches are thinned (i.e. a sample of the matches is randomly selected by CQPweb on the basis of a given number or a percentage) to 1% for EN-CA and to 5% for FR-CA (the difference in thinning being due to different corpus sizes), and the random samples are manually analysed. The sample size is 396 instances of the imperfect and 397 of the preterite for EN-CA and 742 instances of the imperfect and 581 of the preterite for FR-CA. False positives are weeded out at this stage. It is unnecessary to query CA because it has already been done previously with a view to determining salience in the target language. Thinning of matches in CA proceeded along the same lines as in EN-CA - 1%, with a total of 526 instances of the imperfect and 372 of the preterite.
2. Quantification and testing for significance. The results obtained in the previous step are compared to those previously obtained for CA. All relevant differences (i.e. between CA and EN-CA, between CA

and FR-CA, and between EN-CA and FR-CA) are tested for statistical significance.

3. Identification of the main triggers for the preterite and the imperfect in the ST components of both parallel corpora. The random sample obtained in step 1 of the parallel concordances yielded by both EN-CA and FR-CA is manually analysed in order to classify ST triggers – e.g. simple past, past continuous, *passé simple*, *imparfait*, etc. In step 1, only the TT component of the parallel concordances was taken into account, whereas step 3 focuses on the classification of ST triggers. Misalignments are weeded out at this stage.

4. Data retrieval (simple past and other possible triggers for English, *passé simple* and *imparfait* for French) with CQPweb from EN-CA and FR-CA (starting from the ST end). The ST triggers identified in step 3 are entered as queries in CQPweb in order to identify the main TT matches of those triggers. As in step 1, query matches for the main triggers of the Catalan imperfect and preterite are randomly thinned for manual analysis. The 82,444 matches for the simple past in EN-CA are thinned to 1%, the resulting sample size being 825 instances. The 17,842 matches for the *imparfait* and the 11,339 for the *passé simple* in FR-CA are thinned by 5%, the resulting sample sizes being 567 and 892, respectively. (The *passé composé* is not included in the queries because it carries little weight as a ST trigger for the preterite, accounting for just 4.11% of the cases.) Then the TT matches of those three ST triggers are classified.

5. Determining strength of connectivity patterns between ST and TT forms. A formula is used to quantitatively determine strength of connectivity between the Catalan verbal constructions conveying perfective and imperfective aspect and their English and French main triggers.

6. Hypothesis verification and refinement.

Section 5 provides the results of corpus data analysis for each of these steps, as well as a discussion of those results.

5. Results and discussion

The results of step 1 are shown in Table 5. The prevalence of the imperfect in CA is neutralized in EN-CA and partly neutralized in FR-CA, where the gap between the two aspects is narrower.

Table 5. Raw and normalized frequencies of the Catalan imperfect and preterite in CA, EN-CA and FR-CA

	CA		EN-CA		FR-CA	
	Raw f	Norm f (1,000,000 w)	Raw f	Norm f (1,000,000 w)	Raw f	Norm f (1,000,000 w)
Imperfect	52,572	33,880	39,616	29,480	14,960	26,380
Preterite	37,180	23,960	39,718	29,560	11,561	20,390

These results were tested for significance through the chi-square test, which measures differences in a contingency table, not across individual values (e.g.

separately for the imperfect or the preterite). The null hypothesis assumes that rows and columns in a contingency table are independent; if the significance threshold is reached, the null hypothesis is rejected and rows and columns are considered dependent. The results of the test are as follows: $\chi2=1,267.509$ and $p<0.001$ for CA/EN-CA, $\chi2=39.470$ and $p<0.001$ for CA/FR-CA, and $\chi2=333.389$ and $p<0.001$ for EN-CA/FR-CA (with a significance threshold of 3.84 for df=1 and $p\leq0.5$). Values are extremely high and they confirm hypotheses 1 and 3, as well as the second alternative of hypothesis 2: magnetism seems to prevail over connectivity patterns. Moreover, they reveal the size of the difference in the distributions across corpora: the gap between CA and EN-CA is indeed huge, almost four times as wide as that between EN-CA and FR-CA, which is still very large; and CA and FR-CA are much closer to each other, but the gap is still wide enough for the difference to be highly significant.

Steps 3, 4 and 5 in the methodology outlined above are instrumental in determining the degree of connectivity between the imperfective and perfective forms in Catalan translation and their triggers in the English and French source texts. Firstly (step 3), those triggers must be identified through parallel corpus analysis, and their relative frequency calculated. Tables 6, 7, 8 and 9 offer raw and relative frequencies for each ST trigger. The headings of the first two columns in the tables are self-explanatory, but the third is perhaps not. Drawing on Schmid's (2010) *attraction-reliance method*, Halverson (2017: 30ff) introduced two statistical measures with a view to gauging the strength of translation relationships between items in a parallel corpus: source concentration and target concentration. Source concentration is "the percentage of all occurrences of a TL item that are translations of a specific SL item" (Halverson 2017: 30), whereas target concentration is "the percentage of a set of translations of an SL item that is comprised by a given TL item" (Halverson 2017: 30). What the figures in the tables make abundantly clear is that, for the two aspectually marked Catalan forms in both language pairs, there is a ST form with a high or very high concentration: the simple past for both the imperfect and the preterite in EN-CA, the *imparfait* for the imperfect, and the *passé simple* for the preterite in FR-CA. Source concentrations range from 67.77% to 91.41%. This means that, whenever an imperfect or a preterite occurs in Catalan translations from English and French (in the COVALT corpus), there is a high probability that the ST trigger is x – a kind of statement that cannot be made when there is more dispersion, i.e. when source concentration is more evenly distributed among several items.

Table 6. ST triggers of the Catalan imperfect in EN-CA

ST trigger	Raw f	Source concentration
simple past	265	67.77%
-ing	22	5.63%
no verb	22	5.63%
past continuous	21	5.37%
prepositional phrase	12	3.07%
adjective	7	1.79%
past participle	6	1.53%
past perfect	6	1.53%
simple present	6	1.53%
would + infinitive	6	1.53%
noun phrase	5	1.28%
other	13	3.32%

TOTAL	391	

Table 7. ST triggers of the Catalan imperfect in FR-CA

ST trigger	Raw f	Source concentration
imparfait	581	83.48%
passé simple	19	2.73%
indicatif présent	19	2.73%
gérondif	13	1.87%
imparfait subjonctif	10	1.44%
participe présent	10	1.44%
plus-que-parfait	10	1.44%
infinitif présent	6	0.86%
syntagme prépositionnel	6	0.86%
conditionnel présent	5	0.72%
other	17	2.44%
TOTAL	696	

Table 8. ST triggers of the Catalan preterite in EN-CA

ST trigger	Raw f	Source concentration
simple past	351	91.41%
-ing	13	3.39%
past perfect	6	1.56%
no verb	5	1.30%
past continuous	2	0.52%
simple present	2	0.52%
other	5	1.30%
TOTAL	384	

Table 9. ST triggers of the Catalan preterite in FR-CA

ST trigger	Raw f	Source concentration
passé simple	444	79.29%
imparfait	26	4.64%
passé composé	23	4.11%
indicatif présent	23	4.11%
plus-que-parfait	13	2.32%
participe présent	8	1.43%
infinitif présent	5	0.89%
other	18	3.21%
TOTAL	560	

Important as source concentration is in establishing strength of connectivity, it is only one half of the equation. The other half consists in querying the parallel corpora from the source end in order to identify the main TT matches of the top-ranking ST triggers of the Catalan imperfect and preterite, and calculating their target concentration (step 4). Tables 10, 11 and 12 show the main TT matches for the simple past in EN-CA and for the *imparfait* and the *passé simple* in FR-CA, respectively. The picture that emerges here is different for EN-CA and FR-CA, as expected. In FR-CA, target concentration on a single item is high both for the *imparfait* (the Catalan imperfect accounting for over 89% of translation solutions) and the *passé simple* (where the preterite is the translation solution in over 93% of the cases). In EN-CA, TT matches for the simple past are mainly distributed

across the preterite and the imperfect (in that order), which account for over 51% and 37% of the cases, respectively.

Table 10. TT matches for the simple past in EN-CA

TT match	Raw f	Target concentration
preterite	357	51.22%
imperfect	264	37.88%
perfect	15	2.15%
pluperfect indicative	13	1.87%
suppression	10	1.43%
imperfect subjunctive	8	1.15%
no verb	8	1.15%
infinitive	7	1.00%
present indicative	5	0.72%
other	10	1.33%
TOTAL	697	

Table 11. TT matches for the *imparfait* in FR-CA

TT match	Raw f	Target concentration
imperfect	603	89.07%
preterite	24	3.54%
imperfect subjunctive	12	1.77%
conditional	12	1.77%
present indicative	8	1.18%
pluperfect indicative	8	1.18%
other	10	1.48%
TOTAL	677	

Table 12. TT matches for the *passé simple* in FR-CA

TT match	Raw f	Target concentration
preterite	368	93.16%
imperfect	17	4.30%
other	10	2.53%
TOTAL	395	

Now that we have the two halves of the strength of connectivity equation, we need to bring them together. As argued elsewhere (Marco, 2021: 42 and ff.), a formula is needed to operationalize strength of connectivity, as source and target concentration offer complementary but partial views on this factor. The formula used here, which was called Unidirectional Translation Correspondence in a previous study (Marco, 2021: 43), draws on Altenberg's (1999) concept of mutual correspondence.[13] Altenberg's concept is intended to measure the strength of the translation relationship between an item A in a given language and an item B in a different language in a parallel bi-directional corpus. It captures the number of times that A is translated as B and B as A in proportion to the overall number of times that A and B occur in the source texts. Since our corpora are not bi-directional, the formula was adapted as follows:

$$(Ab + Ba) \times 100$$

---

[13] We are indebted to Sandra Halverson (personal communication) for this suggestion.

––––––––––––––

(At + Bs)

where Ab and Ba = the number of times A is the translation of B and B is translated as A (it will be the same figure, of course), and At + Bs = the total number of occurrences of A in TT and of B in ST. Ab and Ba must be the same figure when the whole corpus is taken into account, but this need not be the case when a random sample of all query matches is analysed. In our data (sampled by the thinning utility in CQPweb), for example, the simple past features 697 times in the ST component of EN-CA, and the imperfect occurs 391 times in the TT component. Since the imperfect is triggered by the simple past 265 times, and the simple past is rendered as the imperfect 264 times, the formula applies as follows: (265+264) x 100 / (697+391) = 48.62%. The results of applying the same formula to the other three pairs of verbal forms are 65.49% for the simple past/preterite in EN-CA, 86.23% for the *imparfait*/imperfect in FR-CA and 85.02% for the *passé simple*/preterite in FR-CA. These results show on the one hand strong connectivity, verging on one-to-one correspondence, between the *imparfait* and the imperfect, and between the *passé simple* and the preterite in FR-CA, and on the other a higher degree of connectivity between the simple past and the preterite than between the simple past and the imperfect in EN-CA.

To sum up, different configurations of factors account for varying degrees of support for each of the three hypotheses. Hypothesis 1 is fully confirmed by the data. The alleged magnetism of the preterite in Catalan, the strong gravitational pull exerted by the simple past in translation from English and the relatively high degree of connectivity between the English simple past and the Catalan preterite (or at least the fact that connectivity between those two forms is remarkably higher than that between the simple past and the Catalan imperfect) all pull towards over-representation. As regards hypothesis 2, the alleged magnetism of the preterite in Catalan is again on the side of over-representation of the perfective aspect; but connectivity data pulls towards a distribution that would more or less overlap with that of CA (Catalan non-translations), as Unidirectional Translation Correspondence values for the *imparfait/*imperfect and the *passé simple/*preterite are high. (Let it be recalled that we were forced to exclude gravitational pull of the perfective and the imperfective as a factor in translation from French because of the indeterminacy of corpus results if they are not manually analysed. This is a clear limitation of this study which needs to be addressed in future.) Results show that the balance is tipped in favour of over-representation of the preterite and under-representation of the imperfect; or, in other words, magnetism prevails over connectivity in this particular configuration of factors. Finally, hypothesis 3 is confirmed as a logical result of hypotheses 1 and 2 put together. In a comparison between translations from English and translations from French, magnetism remains neutral. Therefore, possible differences must be explained by the configuration of the remaining two factors, gravitational pull and connectivity. In EN-CA, the simple past exerts strong gravitational pull, whereas in FR-CA this factor was excluded for the reasons just explained. At any rate, neither the French *passé simple* nor the *imparfait* are likely to exert such strong pull, as they compete with each other, whereas in English the simple past is overwhelmingly salient. As to connectivity, a high UTC is observed in FR-CA between isomorphic forms, i.e. between the *imparfait* and the Catalan imperfect and between the *passé simple* and the Catalan preterite. In EN-CA, on

the other hand, the simple past (which covers, to a great extent, the ground jointly covered in Catalan by the imperfect and the preterite) shows a higher UTC with the preterite than the imperfect. As a result, both factors (gravitational pull and connectivity) pull towards over-representation of the preterite and under-representation of the imperfect in EN-CA as compared to FR-CA.

6.   Concluding remarks

The last paragraph of Section 5 summarizes the main findings in this study as regards confirmation (or otherwise) of the three hypotheses put forward. Those findings spark off a number of reflections.

Firstly, the GPH is largely confirmed by the results of this study. This needs to be highlighted as hypotheses on translated language can only be proved or disproved on the basis of evidence accumulation. The results of this study must first and foremost be seen as a modest contribution to that (arguably collective) task, briefly outlined in Section 2.

Secondly, it is remarkable how evidence from the field of L2 acquisition and the findings in this study converge on the imperfective/perfective aspect distinction. The research reviewed in Section 3 (even if only a broad overview was possible) was unambiguous as to the salience of the perfective, when compared to the imperfective, in Romance languages. This was part of the input for the hypotheses then formulated; and the results clearly validate the salience of the perfective. This is most clearly visible in FR-CA, where a high degree of isomorphism between the aspectual systems of French and Catalan might have pulled towards non-significance of the differences in aspect distribution between FR-CA and CA. But the opposite was true: magnetism pulled towards over-representation of the preterite in spite of high connectivity. It may be concluded that multi-disciplinarity is always a good thing, and, more particularly, that studies on language acquisition have a lot to offer whenever the GPH is put to the test.

Finally, a number of limitations of this study must be pointed out. The first one is corpus size in the case of FR-CA. With just over half a million words in each component, this parallel corpus might be said to be too small for hypothesis validation. There is not much that can be said to counter such an objection. Our initial aim when building the COVALT corpus was to include all Catalan translations published in the region of Valencia between 1990 and 2000 of narrative texts originally written in English, French and German, with no consideration of corpus size. A larger corpus would certainly exhibit more variety, especially of source texts as regards temporal background. Whether that would impact the balance between perfective and imperfective aspects remains to be seen.

The second limitation of the study is arguably its monofactorial nature – the fact that it concentrates on what De Sutter and Lefer (2020: 5) call *translation status*, i.e. the translation/non-translation distinction. These authors make a strong case for multifactorial corpus-based translation studies, since differences in patterns observed between translations and non-translations may stem from factors other than translation status – as they convincingly demonstrate for optional *that* in complement clause constructions in English. However, translation status is the only variable we were interested in for the purposes of this study, as it is the variable upon which the GPH hinges for its very formulation. The GPH

assumes differences across the translation/non-translation divide insofar as specific configurations of magnetism, gravitational pull and connectivity favour such differences. Therefore, those three factors are seen as explanatory – they account for possible difference, or similarity, as the case may be. De Sutter and Lefer identified other variables, apart from translation status, which were inherent to the construction under scrutiny (e.g. length of the full complement clause, distance between the matrix-clause verb and the complement clause subject, etc.). But in the case of perfective vs. imperfective aspect, the main inherent variable that comes to mind is lexical aspect, which cannot be established on the basis of verbs in isolation but verbal predicates, i.e. taking account of the verbs and their arguments in individual occurrences. Therefore, analysing lexical aspect would involve tagging all corpus matches for that variable – a daunting task indeed, as almost 5,000 concordances were manually analysed in this study. At any rate, including lexical aspect as a variable points towards a possible extension of the research reported on in this paper. Unless lexical aspect proves relevant, the only possible explanatory variable for differences between translations and non-translations, we would argue, is translation status.

Acknowledgements

References

Ainaud, J., Espunya, A. and Pujol, D. 2003. *Manual de traducció anglès-català*. Vic: Eumo.

Altenberg, B. 1999. Adverbial Connectors in English and Swedish: Semantic and Lexical Correspondences. In *Out of Corpora: Studies in Honour of Stig Johansson*, H. Hasselgård and S. Oksefell (eds), 249–268. Amsterdam: Rodopi.

Andersen, R. 1991. Developmental Sequences: The Emergence of Aspect Marking in Second Language Acquisition. In *Crosscurrents in Second Language Acquisition and Linguistic Theories*, T. Huebner and C.A. Ferguson (eds), 305–324. Amsterdam and Philadelphia: John Benjamins.

Binnick, R.I. 1991. *Time and the Verb. A Guide to Tense & Aspect*. Oxford: Oxford University Press.

Comajoan, L. 2005. The Acquisition of Perfective and Imperfective Morphology and the Marking of Discourse Grounding in Catalan. In *Tense and Aspect in Romance Languages*, M.R. Salaberry and D. Ayoun (eds), 35–77. Amsterdam and Philadelphia: John Benjamins.

Comrie, B. 1976. *Aspect*. Cambridge: Cambridge University Press.

De Sutter, G. and Lefer, M.A. 2020. On the Need for a New Research Agenda for Corpus-based Translation Studies: A Multi-methodological, Multifactorial and Interdisciplinary Approach. *Perspectives* 28(1): 1–23.

Diaubalick, T. 2019. *La adquisición del sistema verbal español por aprendices alemanes y el papel del aspecto gramatical. Una comparación entre los tiempos del pasado y los tiempos del futuro*. Tübingen: Narr.

Grevisse, M. and Goosse, A. 2008. *Le bon usage*. Brussels: De Boeck Duculot (14th edition).

Halverson, S. 2003. The Cognitive Basis of Translation Universals. *Target* 15(2): 197–241.

Halverson, S. 2017. Developing a Cognitive Semantic Model: Magnetism, Gravitational Pull, and Questions of Data and Method. In *Empirical Translation Studies. New Methods and Theoretical Traditions*, G. de Sutter, M.A. Lefer and I. Delaere (eds), 9–45. Berlin: Mouton de Gruyter.

Hareide, L. 2017a. The Translation of Formal Source-language Lacunas. An Empirical Study of the Over-representation of Target-Language Specific Features and the Unique Items Hypothesis. In *Corpus Methodologies Explained. An Empirical Approach to Translation Studies*, M. Ji, M. Oakes, L. Defeng and L. Hareide (eds), 137–187. London and New York: Routledge.

Hareide, L. 2017b. Is There Gravitational Pull in Translation? A Corpus-based Test of the *Gravitational Pull Hypothesis* on the Language Pairs Norwegian-Spanish and English-Spanish. In *Corpus Methodologies Explained. An Empirical Approach to Translation Studies*, M. Ji, M. Oakes, L. Defeng and L. Hareide (eds), 188–231. London and New York: Routledge.

Langacker, R. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Lapshinova-Koltunski, E. 2015. Variation in Translation: Evidence from Corpora. In *New Directions in Corpus-based Translation Studies*, C. Fantinuoli and F. Zanettin (eds), 93-113. Berlin: Language Science Press.

Lefer, M.A. and De Sutter, G. 2022. Using the Gravitational Pull Hypothesis to Explain Patterns in Interpreting and Translation: The Case of Concatenated Nouns in Mediated European Parliament Discourse. In *Mediated Discourse at the European Parliament: Empirical Investigations*, M. Kajzer-Wietrzny, A. Ferraresi, I. Ivaska and S. Bernardini (eds), 133–159. Berlin: Language Science Press.

Maingueneau, D. and Salvador, V. 1995. *Elements de lingüística per al discurs literari*. València: Tàndem.

Marco, J. 2021. Testing the Gravitational Pull Hypothesis on Modal Verbs Expressing Obligation and Necessity in Catalan through the COVALT Corpus. In *Empirical Studies in Translation and Discourse*, M. Bisiada (ed), 27–52. Berlin: Language Science Press.

Oster, U. and Tello-Fons, I. In preparation. Between source language constructions and target language expectations. Translating passive construal.

Pérez Saldanya, M. 2002. Les relacions temporals i aspectuals. In *Gramàtica del català contemporani*, vol. 3: *Sintaxi*, J. Solà, M.R. Lloret, J. Mascaró and M. Pérez Saldanya (dirs), 2567–2662. Barcelona: Empúries.

Riegel, M., Pellat, J.C. and Rioul, R. 2004. *Grammaire méthodique du français*. Paris: Presses Universitaires de France (3rd edition).

Salaberry, M.R. and Ayoun, D. 2005. The Development of L2 Tense-Aspect in the Romance Languages. In *Tense and Aspect in Romance Languages*, M.R. Salaberry and D. Ayoun (eds), 1–33. Amsterdam and Philadelphia: John Benjamins.

Schmid, H.-J. 2010. Does Frequency in Text really Instantiate Entrenchment in the Cognitive System? In *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches*, D. Glynn and K. Fischer (eds), 101–133. Berlin: Walter de Gruyter.

Tirkkonen-Condit, S. 2004. Unique items – over- or underrepresented in translated language? In *Translation universals. Do they exist?*, A. Mauranen and P. Kujamäki (eds), 177–184. Amsterdam: John Benjamins.

Vandevoorde, L. 2020. *Semantic Differences in Translation*: *Exploring the Field of Inchoativity*. Berlin: Language Science Press.

Vendler, Z. 1967. *Linguistics and Philosophy*. Ithaca, New York: Cornell University Press.

Verdegal, J. 2011. *La pràctica de la traducció francès-català*. Vic: Eumo.

Wallace, S. 1982. Figure and Ground: The Interrelationships of Linguistic Categories. In *Tense-Aspect: Between Semantics and Pragmatics*, P. Hopper (ed), 201–223. Amsterdam and Philadelphia: John Benjamins.

Wheeler, M.W., Yates, A. and Dols, N. 1999. *Catalan: A Comprehensive Grammar*. London and New York: Routledge.