

## Chapter 2

# Testing the Gravitational Pull Hypothesis on modal verbs expressing obligation and necessity in Catalan through the COVALT corpus

Josep Marco

Universitat Jaume I

This chapter aims to test Halverson's Gravitational Pull Hypothesis (GPH) on the Catalan modal verb *caldre*. Data from two comparable parallel sub-corpora from the COVALT corpus (English-Catalan and French-Catalan) are used to that end. However, *caldre* only serves as a starting point for hypothesis formulation and data analysis, as it is part of the wider network of modal verbs and verbal periphrases conveying obligation or necessity in Catalan. On the basis of networks of obligation and necessity in Catalan, English and French, three hypotheses are put forward: 1) *caldre* will be under-represented in the English-Catalan subcorpus when compared to Catalan non-translations; 2) *caldre* will be either over- or under-represented in the French-Catalan sub-corpus when compared to Catalan non-translations; and 3) *caldre* will be significantly more frequent in the French-Catalan than in the English-Catalan sub-corpus. Results confirm hypotheses 1 and 3, and for hypothesis 2 the scales are tipped in favour of over-representation, mainly as a result of strong connectivity between *caldre* and its French source text triggers. Connectivity, in the present study, seems to be favoured by formal similarity, which comes in two forms: syntactic isomorphism and phonological/graphological similarity.

### 1 Introduction

The main aim of this chapter is to test out Halverson's Gravitational Pull Hypothesis (2003; 2010; 2017) on the Catalan modal verb *caldre* in two compara-



ble parallel sub-corpora from the COVALT corpus: English-Catalan and French-Catalan. The Gravitational Pull Hypothesis (GPH) was first put forward as an attempt to account for different hypotheses about translated language by anchoring them in cognitive linguistics and bilingualism. The Catalan modal verb *caldre* is arguably a suitable testing ground for the hypothesis because it may be considered a unique item (UI) in the English-Catalan language pair but not in the French-Catalan. Focusing on an item that fulfils that condition in that kind of environment (two comparable parallel corpora) is one of the methods suggested (Hareide 2017) to put the GPH to the test. However, *caldre* will only serve as a starting point in hypothesis formulation and corpus data analysis. It is part of the wider network of modal verbs and verbal periphrases conveying obligation or necessity in Catalan; therefore, other nodes in that network in the languages involved will be brought under scrutiny.

The outline of the study is as follows. §2 will present the main tenets of the GPH. §3 will provide an account of the main modal verbs and periphrases used to convey obligation and necessity in Catalan, English and French. §4 will spell out the aims of the study, the particular hypotheses to be tested on corpus data, together with the assumptions on which they are based, and the different steps into which the method followed can be broken. §5 will present data from the two parallel sub-corpora and the component of Catalan non-translations, which will be analysed and discussed. Finally, §6 will offer some conclusions.

## 2 The Gravitational Pull Hypothesis

Halverson's Gravitational Pull Hypothesis (GPH) aims to account for different hypotheses about translated language by anchoring them in cognitive linguistics. As is well-known, Baker (e.g. 1993) postulated a number of so-called *translation universals*, or features of translated language purported to occur independently of the language pair involved and the particular factors defining the translation situation (brief, genre, register, gender, ideology, etc.). One of the postulated features was over-representation of typical target language (TL) elements in translated text. This claim may be said to be akin to Toury's law of growing standardisation, which states that "in translation, source-text textemes tend to be converted into target-language (or target-culture) repertoemes" (Toury 1995: 267–268) – textemes being special (perhaps unique) textual relations and repertoemes, conventional elements of the target repertoire. On the other hand, Tirkkonen-Condit (2002; 2004) argued a few years later that typical TL elements or structures tended to be under-represented (not over-represented) in translated

## 2 *Testing the Gravitational Pull Hypothesis on verbs in Catalan*

language when their use was not triggered by a formal equivalent or counterpart in the source language (SL). These elements lacking correspondence in the other member of a particular language pair were called *unique items* – even if uniqueness in this case must be understood as a relative concept concerning two particular languages, not in the absolute sense of a given item existing only in one human language and being unheard of in the rest. How can these two opposing claims be true? On the face of it, the truth of one would preclude the truth of the other. However, both have been attested and are supported by (at least partial) evidence.

In this context, the main tenet of Halverson's GPH is that features of translated language (including over- and under-representation of TL typical items) can be explained on the basis of general characteristics of human cognition. Cognitive linguistic theory claims that linguistic units are integrated into higher-order structures, such as schematic networks, made up of nodes and links between nodes. Networks are characterised by asymmetry: some nodes are more salient than others. The most salient elements are usually the high-level schema (because of its high level of generality) and the prototype (understood as the best representative of a category). On the other hand, empirical research on bilingualism has identified two layers of representation in the bilingual brain: the lexical memory, where the forms of spoken and written lexical items are stored, and the conceptual level, where word meanings are stored. Links between words in different languages are set up through connections either at the conceptual or the lexical level, and such connections may rest upon total or partial overlap. These are then the two basic elements of human cognition, which are brought to bear on translational behaviour: degree of cognitive salience (of particular nodes in a network) and degree of overlap between nodes and networks across languages, which will favour connectivity or otherwise. The theory merges concepts from cognitive grammar (especially Langacker 1987) and bilingualism (De Groot 1992a,b). The implications of all this for translation are spelt out by Halverson (2003: 218) as follows:

The basic idea is straightforward: in a translation task, a semantic network is activated by lexical and grammatical structures in the ST. Within this activated network, which also includes nodes for TL words and grammatical structures, highly salient structures will exert a gravitational pull, resulting in an overrepresentation in translation of the specific TL lexical and grammatical structures that correspond to those salient nodes and configurations in the schematic network.

In a later work (Halverson 2010), some problems with the 2003 original model were identified. The main one was theoretical: the GPH as initially formulated conflated two separate factors: the content of the schematic network and the specific pattern of connectivity, which need to be teased apart for a more valid explanation of translational effects. The model was accordingly revised to posit three possible cognitive causes of translational effects (Halverson 2010: 356; Hareide 2017: 192): patterns of prototypicality in the target language, conceptual structures or the representation of the source language item, and patterns of 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. A further revision of the GPH model was proposed by Halverson in 2017. The term *gravitational pull* was now reserved for translational effects stemming from the representation of SL items (factor 2), i.e. for “the cognitive force that makes it difficult for the translator to escape from the cognitive pull of highly salient representational elements in the source language” (Halverson 2017: 14). Salience in the target language (factor 1) “may be more clearly captured by the metaphorical term *magnetism*” (2017: 14). And the third factor is called *connectivity* and defined as “the nature and strength of links between elements in a bilingual’s two languages” (2017: 14). As in earlier versions of the model, Halverson (2017: 28) stresses the fact that it is not possible at present to predict how salience patterns and connectivity interact, whether the different factors act independently or jointly etc. Like any scientific theory, the GPH is open to refinement and modification on the basis of evidence.

Pending availability of more evidence, the choice of this hypothesis as a framework can only be justified on theoretical grounds. Firstly, it may be said to be a move away from binary formulations towards a more complex, multi-factorial analysis. Corpus-based studies of features of translated text typically set out to test a particular hypothesis (e.g. simplification, explicitation, normalisation) in isolation on a given set of data, so that the hypothesis is confirmed or refuted on a binary, yes/no basis. The GPH, in contrast, brings together several causes and attempts to find out the effects brought about by a particular configuration of such causes and the interaction between them. And secondly, it is rooted in cognition – cognition being one of the two possible causes of translation features proposed so far (Halverson 2017: 10): “there are two main approaches taken to the problem of explaining translational patterns. These two are socially and cognitively oriented, respectively”. Malmkjær (2008: 57) had gone further than that and suggested that the term *universal* (which is less and less used nowadays in Malmkjær’s absolute sense) be reserved for phenomena that can be cognitively

explained. The reader is referred to Halverson (2017) for a more detailed illustration of the two approaches to the explanation of translation properties.

### 3 Modal verbs expressing obligation or necessity in Catalan, English and French

There are two major verbal constructions expressing obligation or necessity in Catalan: *haver de* + infinitive and *caldre*. Of course the language has many other resources to convey these meanings, but these two are fully grammaticalised – in fact, Badia i Margarit (1994: 611) refers to them as “grammatical formulas”. The former is a modal periphrasis and shares some features of syntactic behaviour with other modal periphrases. According to Gavarró & Laca (2002: 2714), it can convey both necessity (interpreted as epistemic) and obligation (interpreted as deontic). *Caldre* + infinitive (one of the possible constructions in which *caldre* can occur), on the other hand, is excluded by these authors (2002: 2710) from the list of Catalan modal periphrases on grounds of syntactic behaviour, even if earlier grammarians (cf. Badia i Margarit 1994) had treated it as such.

Be that as it may, there seems to be general agreement that *caldre* is a modal verb (see e.g. Payrató 2002: 1192; Rigau 2005). Rigau sees *caldre* as belonging to the category of relative impersonal verbs, i.e. verbs used impersonally in that they refer to a person who does not feature in the sentence as agent but recipient (Rigau 2005: 242). This verb, together with similar ones belonging to the same category, follows the syntactic pattern of such Latin verbs as *licet* (‘it is licit/permitted’) or *oportet* (‘it is proper/necessary’, ‘it behoves’). These verbs were only used in the third person and took two kinds of complements: an infinitive or a subordinate clause introduced by *ut* (similar to a *that*-clause). Relative impersonal verbs exist in all Romance languages, even if their syntactic behaviour shows some variation. The list provided by Rigau (1999: 324) includes (relevantly to our purposes, as will be seen) French  *falloir*. *Caldre* comes from Latin *calēre* (‘to be hot’) – hence the sense of urgency, of necessity. It has cognates in such neighbouring languages as Occitan and Aragonese, and it used to have them in medieval Spanish, Old French and Old Italian (Rigau 1999: 331) as well. Syntactically, it may take three kinds of complements: an infinitive, a subordinate clause introduced by *que* (i.e. a *that*-clause) and a noun phrase. It may also take a zero complement, with any of the three types of complements just mentioned left implicit. Examples (1–4) illustrate these four patterns, respectively.

- (1) Cal tenir molta força per moure aquesta taula.  
‘It takes a lot of strength/a very strong person to move this table.’

- (2) No cal que t'amoïnes tant.  
'You needn't worry so much.'; more literally, 'it is not necessary that you worry so much.'
- (3) Et caldrà molta paciència si vols convèncer-lo.  
'It will take you a lot of patience if you want to persuade him.'
- (4) Pots fer servir el meu cotxe, si cal.  
'You can use my car if necessary.'

To sum up, *haver de* + infinitive is a modal periphrasis that can convey both epistemic and deontic meaning. It can occur both in personal (as in 5) and impersonal (as in 6) constructions.

- (5) Ha de lliurar el CV si vol que l'entrevisten.  
'He must/has to submit a CV if he wants to be interviewed.'
- (6) S'ha de tenir en compte que jo no hi era.  
'It must be taken into account that I was not there.'

*Caldre* is a modal verb, but grammarians do not agree on the kind of modality it conveys, whether epistemic or deontic. It is almost exclusively used in impersonal constructions. It is here assumed, since there is no evidence to the contrary in the literature, that the meaning of *caldre* remains stable across the four constructions it occurs in. No meaning variation is observed depending on the kind of complement it takes.

As to modality in English, various accounts have been provided on the basis of widely differing theoretical assumptions. Cognitive accounts start from the basic epistemic vs. deontic distinction, even if Langacker (1991: 272) claims that this distinction "is not always easy to maintain", as most English modals can be used both epistemically and deontically. Langacker draws on Talmy and Sweetser to suggest that "the English modals are best analyzed in terms of force dynamics" (1991: 273). Force-dynamic values are applicable either to the domain of social interaction (deontic modality) or reasoning (epistemic modality). Radden & Dirven (2007) take a much more comprehensive view, which can only be briefly summarised here. These authors define modality as "an assessment of potentiality, depending either on the speaker's judgement of the reality status of a state of affairs (epistemic modality) or on the speaker's attitude towards the realisation of a desired or expected event (root modality)" (Radden & Dirven 2007: 246). Under root modality three sub-types are subsumed: deontic, intrinsic and disposition. Deontic modality is typically realised in two ways: obligation

## 2 Testing the Gravitational Pull Hypothesis on verbs in Catalan

(e.g. *You must be home by nine*) and permission (e.g. *You may leave now if you wish*). Intrinsic modality is concerned with “intrinsic qualities of a thing or circumstances” (Radden & Dirven 2007: 246), as in *Pros and cons must be weighed up before a decision is made*. Finally, disposition modality encompasses the notions of ability, propensity and willingness (2007: 246). Radden & Dirven (2007) make another distinction that cuts across the previous one, that between compelling and enabling modalities – the two main paths of grammaticalisation leading from lexical to deontic and epistemic meanings. Compelling modalities “involve a compelling force; they comprise obligations, prohibitions and intrinsic and epistemic necessities” (Radden & Dirven 2007: 247), whereas enabling modalities convey possibilities, abilities and permissions.

The focus of the present chapter is on the first group, once epistemic necessity has been removed, i.e. on obligation (whether positive or negative) and intrinsic necessity, because those are the senses present in the Catalan modals *caldre* and *haver de*. (Epistemic necessity is conveyed by a different modal verb, *deure*.) These modalities are expressed by “the central modals *must*, *need (to)* and *should* and the semi-modals *ought to*, *have to* and *have got to*” (Radden & Dirven 2007: 247). Differences among these verbs are set up on the basis of two criteria: the source of the compelling force and the degree of strength. The source of the compelling force may be the speaker (subjective) or external circumstances (external). And according to its degree of strength, the compelling force may be strong, neutral or weak. Obligation is always subjective, with *must* and *have got to* as strong and *should* and *ought to* as weak indicators of modality. Intrinsic necessity is external, with *have (got) to* and *must* as strong, *need to* as neutral and *should* and *ought to* as weak indicators of modality. The central (in the sense of most commonly used) modal verbs for these categories would be *must* for strong obligation, *have (got) to* for strong intrinsic necessity and *should* for weak obligation or intrinsic necessity. Radden & Dirven (2007: 249) further add that shifts in the system of compelling modals have occurred in American English due to democratisation and colloquialisation, to the extent that *must* has become much less common than *have (got) to*.

Catalan *caldre* conveys both obligation and intrinsic necessity. In French, these meanings are mainly conveyed by the verbs *devoir* and *falloir*. According to Lewis (2015: 159), “*devoir* is said to be more solemn or more insistent than *falloir*, while *falloir* is more often used in ‘subjective contexts’ where *devoir* might be interpreted as epistemic or as expressing futurity”. Both are polysemic in that they can express obligation, whether from an internal or an external source (i.e. the source of the obligation may be either the speaker or otherwise, respectively),

and “non-deontic necessity” (what we have referred to here as intrinsic necessity). Moreover, *devoir* can also convey epistemic necessity (like English *must* or Catalan *deure*) and futurity. Beyond their semantic values, Lewis emphasises the syntactic differences between the two verbs (2015: 158–159): “deontic *devoir* typically takes a human subject while *falloir* can only be used with dummy subject *il*”. In other words, while the former occurs in personal constructions, the latter is impersonal, like *caldre*. Furthermore, *devoir* is regarded as more formal than *falloir*.

## 4 Aim and methodology

As explained at the beginning, the main aim of this chapter is to test out the GPH on the Catalan modal verb *caldre* in two comparable parallel sub-corpora from the COVALT corpus: English-Catalan and French-Catalan. The corpus used will be both parallel and comparable, as data will also be retrieved from a component of Catalan non-translations. *Caldre* is used as a starting-point for the analysis because it may be said to be a unique item for the English-Catalan but not for the French-Catalan language pair. English does not have a syntactic counterpart for *caldre*, as possible candidates, such as the expression *it + take + X* (as in *It takes a lot of courage to rise to that challenge*), are not frequent or grammaticalised to the same extent as *caldre* is.<sup>1</sup> French, on the other hand, has the verb *falloir*, as seen above, which also conveys obligation or necessity and typically occurs in impersonal constructions. As seen in §3, both *caldre* and *falloir* fall under the category of relative impersonal verbs, which cuts across all Romance languages, and take the same kind of complements to a large extent – the only difference being that *falloir* cannot take a noun phrase as a complement. They share the other two complements (infinitive and *that*-clause), and that is the basis of their syntactic similarity. Formal similarity comes under many guises. The most obvious one is phonological or graphological similarity, especially when it concerns two words with a common origin, e.g. English *hound* and German *Hund*. But there may be formal similarity at other levels, such as that of syntax. Catalan

---

<sup>1</sup>The query [lemma=“it”][lemma=“take”] in the ST component of the English-Catalan sub-corpus in COVALT yields 44 matches, 15 of which are false positives, the remaining 29 often featuring a time complement, as in *It took three days to...* The normalised frequency (f) of this construction is 0.024 per 1,000 words. Just for the sake of comparison, the normalised frequency of *must* as an indicator of obligation and intrinsic necessity is 0.55 – over 20 times as high as the frequency of *it + take*. (This value is based on a projection of the results yielded by the manual analysis of a random sample of 300 instances, out of the total 993 matches found for the query [lemma=“must”].)

## 2 Testing the Gravitational Pull Hypothesis on verbs in Catalan

*caldre* and French *falloir* are not cognates, but they share two syntactic patterns in addition to their semantic common ground. The rationale behind using an item that is unique for a certain language pair but not for another is the same as in Hareide (2017), which serves here as a source of methodological inspiration: the basic contrastive fact around which the study pivots may well give rise to different configurations of factors related to salience and connectivity that may impact translation outcomes.

One of the pre-requisites for this methodology is having two parallel corpora that can be regarded as comparable in all relevant respects, i.e. textual genre, date and place of publication, and type of readership. The COVALT corpus fulfils such a requirement. COVALT (Valencian Corpus of Translated Literature) is a multilingual corpus 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. The English-Catalan sub-corpus comprises 36 English source texts, amounting to 1,201,757 words, and their corresponding target texts in Catalan (1,343,631 words). The French-Catalan sub-corpus comprises 21 French source texts, amounting to 551,869 words, and their corresponding target texts (566,998 words). COVALT also includes non-translated components for both target languages, Catalan and Spanish. The Catalan non-translated component is a set of narrative works originally written in Catalan intended to be comparable to the translated component in all relevant respects: place of publication (Valencian Community), date of publication (1990-2000), language (Catalan) and genre (narrative fiction). The non-translated component amounts to 1,551,521 tokens. 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>).

Before formulating hypotheses, we need at least some basic information on the relative salience of the main verbal indicators of obligation and intrinsic necessity in the three languages involved in this study. Since salience is operationalised as frequency (as will be seen later on), corpus data will be used when suitable. Grammars tell us that both *caldre* and *haver de* + infinitive are central as regards the expression of obligation and necessity in Catalan, and the same applies to *must* and *have (got) to* for English; but which member of the pair is the more frequent? In the component of Catalan non-translations in COVALT, *caldre* occurs 735 times, with a normalised frequency of 0.47 per 1,000 words, whereas *haver de* + infinitive features 1,924 occurrences, with a normalised frequency of 1.24 per 1,000 words. *Haver de* + infinitive is about 2.5 times as frequent as *caldre* (in

terms of normalised frequency) and we may assume, therefore, that it is more salient.

In the case of English, it would not make much sense to compare corpus frequencies because the meanings of *must* and *have to* do not overlap to such an extent as *caldre* and *haver de* in Catalan. As seen in §3, *must* is the central verb for strong obligation and *have (got) to* for strong intrinsic necessity (Radden & Dirven 2007). We can rely on these assumptions in order to formulate hypotheses. It must also be borne in mind that the absence of obligation is usually conveyed by the negative forms of *have to* and *need (to)* – not by the negative form of *must*, which expresses negative obligation, i.e. prohibition.

As to French, we saw above that the main modal or semi-modal verbs conveying obligation and intrinsic necessity are *devoir* and *falloir*. As in the case of English, it would not make much sense here to compare corpus frequencies of these two verbs (e.g. in the ST component of the French-Catalan sub-corpus in COVALT) because their meanings overlap only to a certain extent. *Devoir* is more polysemous than *falloir*, as it also conveys epistemic necessity and futurity, and it would be necessary to discard these meanings manually. Lewis (2015) reports on a previous study by Labbé & Labbé (2013) according to which *falloir* is much more frequent than *devoir* in spoken and literary French, whereas the opposite is true for a corpus of presidential speeches. Lewis claims that her own results from a corpus-based analysis of political speeches in English and French are consistent with Labbé and Labbé's findings.

We may therefore assume that: a) *haver de* + infinitive is a more salient indicator of obligation and intrinsic necessity than *caldre* in Catalan; b) the main *prima facie* equivalents of *caldre* (and *haver de*) in English (*must* and *have to*) and French (*falloir*) are also salient in their respective modality networks; and c) patterns of connectivity between *caldre* and those *prima facie* equivalents will be stronger for French than for English. The first two assumptions were justified in the previous paragraph. The third assumption is based on the formal similarity between *caldre* and *falloir*, i.e. on their syntactic overlap (explained above), which is not paralleled by *caldre* and any of its English equivalents. On the basis of these assumptions, the following three hypotheses can be formulated:

1. *caldre* will be under-represented in the English-Catalan subcorpus when compared to Catalan non-translations, as neither factor 1 (magnetism) nor factor 3 (high degree of connectivity) will be at play – factor 2 (gravitational pull) being the only factor that might pull towards over-representation;

## 2 Testing the Gravitational Pull Hypothesis on verbs in Catalan

2. *caldre* will be either over- or under-represented in the French-Catalan sub-corpus when compared to Catalan non-translations, depending on which factor prevails (gravitational pull and a high degree of connectivity will pull towards over-representation whereas magnetism will pull towards under-representation);
3. *caldre* will be significantly more frequent in the French-Catalan than in the English-Catalan sub-corpus, as over-representation will be favoured by two factors (gravitational pull and a high degree of connectivity) in the former and only one (gravitational pull) in the latter.

The method employed to verify these hypotheses will consist of the following steps:

1. data retrieval with CQP (Corpus Query Processor), a tool that allows to query corpora on the basis of regular expressions containing words, lemmas and part-of-speech tags. Both the translated components of the English- and French-Catalan sub-corpora, and the Catalan non-translated component will be queried on the lemma *caldre*;
2. manual sifting in order to tell apart true from false positives. Corpus queries usually yield matches that do not conform to the criteria the analyst had in mind. If false positives are not removed, the data on which quantification draws will be distorted;
3. quantification + testing for significance. Raw and relative frequencies of *caldre* in the three components mentioned in step 1 will be established and tested for significance;
4. searching for triggers (i.e. ST segments matching the query word) of *caldre* in the English and French STs;
5. searching for TT segments matching the main triggers of *caldre*. Query matches will be thinned if their number proves unmanageable. Thinning is the standard method used by CQP for random sampling, and it can be based on a raw figure or a percentage;
6. manual sifting (again), in order to tell apart true from false positives;
7. establishing degrees of connectivity between ST and TT items. The measure to be used for that purpose will be introduced below;

8. repeating the whole process for *haver de* (the main alternative to *caldre* in Catalan, as seen above) in the English-Catalan and French-Catalan sub-corpora and the Catalan non-translated component.

As these steps suggest, for the big picture to emerge as regards patterns of salience and connectivity in the two language pairs it is necessary to go beyond the initial pivot of the study (*caldre*) and look at the main nodes in the monolingual and bilingual networks of which *caldre* is a part. This kind of analysis is extremely time-consuming. The *big* picture may not be the *full* picture, but it is hoped it will include enough relevant information not only to test the hypotheses but also to understand why they are confirmed or refuted.

Before moving on to results and discussion, the thorny question of the relationship between frequency, on the one hand, and salience and connectivity, on the other, must be addressed. Schmid (2010) poses the question in the most explicit possible manner when he wonders whether frequency in text instantiates entrenchment in the cognitive system. Entrenchment is defined as “the degree to which the formation and activation of a cognitive unit is routinized and automated” (Schmid 2010: 115). It is fostered by repetitions of cognitive events. Schmid (2010: 116) 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”, and goes on to argue that, since speed of access and retrieval correlates with routinisation, “this indeed supports the idea that frequency and entrenchment co-vary” (Schmid 2010: 116). But this is not as straightforward as it seems.

Drawing on previous authors, Schmid (2010: 116) claims that “it is not frequency of use as such that determines entrenchment, but frequency of use with regard to a specific meaning or function, in comparison with alternative expressions of that meaning or function”. The former type of frequency is called *absolute* and the latter *relative*. Schmid observes that, even though the correlation between frequency and cognitive significance is far from unproblematic, cognitively-oriented corpus linguists “try to correlate the frequency of occurrence of linguistic phenomena (as observed in corpora) with their salience or entrenchment in the cognitive system” (Schmid 2010: 101). Indeed, it seems difficult to proceed otherwise. Schmid’s caveats are very much in place in methodological terms, but he provides no alternative to frequency as an operationalisation of salience and entrenchment, as no direct access to the cognitive system seems to be available at present. Halverson advocates a mixed-methods approach with different types of data (elicitation data and analysis of keystroke logs) in addition to corpus data, but, regardless of the type of data under scrutiny, both salience and entrenchment are operationalised as frequency. The same procedure will be

followed here, even though most analyses (as in Halverson 2017) will be based on relative rather than absolute frequency.

## 5 Results and discussion

The lemma *caldre* was inserted in the query box of CQPweb for the three relevant sub-corpora: English-Catalan (EN-CAT), French-Catalan (FR-CAT) and Catalan non-translations (NTR). Query matches were manually checked and the number of false positives found to be rather low: 7 (out of 386 hits) for English-Catalan, 9 (out of 524 hits) for French-Catalan and 50 (out of 785 hits) for Catalan non-translations. All false positives are related to the contraction *cal(s)*, meaning ‘at somebody’s (house)’ and the adjective *calent*, meaning ‘hot’. Once these unwanted matches have been removed, results are as shown in Table 2.1.

Table 2.1: Query results for *caldre* in English-Catalan, French-Catalan and Catalan non-translations (f = normalised frequency per 1,000 words)

	n (words)	Query matches	f
<b>Translations from English</b>	1,343,631	379	0.28
<b>Translations from French</b>	566,998	515	0.91
<b>Catalan non-translations</b>	1,551,521	735	0.47

The figures for normalised frequency per 1,000 words strongly hint at significant differences across corpora. The log-likelihood (LL) test was applied to each pair of corpora and the differences turned out to be extremely significant in all three cases, with LL values at 70.33 for EN-CAT/NTR, 121.59 for FR-CAT/NTR and 299.56 for EN-CAT/FR-CAT.<sup>2</sup> The implications of these results for the three hypotheses formulated in the previous sections can be spelt out as follows:

1. *caldre* is under-represented in English-Catalan translations, when compared to Catalan non-translations;
2. *caldre* is over-represented in French-Catalan translations, when compared to Catalan non-translations, in accordance with one of the two possibilities foreseen in hypothesis 2;
3. *caldre* is significantly more frequent in French-Catalan translations than in English-Catalan translations.

<sup>2</sup>The critical value of the log-likelihood test is 3.84 for a 95% level of confidence (i.e. for a p value of <0.05) and 6.63 for a 99% level of confidence (p<0.01). Therefore, any LL value lower than 3.84 indicates that differences do not reach the threshold of statistical significance.

Thus, hypotheses 1 and 3 are confirmed, and for hypothesis 2 the scales are tipped in favour of over-representation, which suggests that gravitational pull and a high degree of connectivity between *caldre* and its French triggers prevail over the relatively low magnetism posited for this verb in the Catalan modality network (in comparison with *haver de* + infinitive).

But at this point we know nothing yet about connectivity patterns between *caldre* and its triggers, as we have only looked at the translated component of the parallel corpora, not at the bilingual concordances. Bilingual concordance analysis for each parallel corpus is expected to provide: a) a list of ST triggers for *caldre*; b) the source concentration for those triggers; c) starting from the ST pole, a list of matching TT segments for the main triggers of *caldre*; d) the target concentration of those TT segments; e) a quantitative measure of the degree of connectivity between *caldre* and its ST triggers, based on source and target concentration. Let us see how this unfolds step by step, first for EN-CAT and then for FR-CAT. But before looking at results we need to dwell on the concepts of source and target concentration.

Schmid (2010) put forward two statistical measures to gauge the interaction between nouns and different kinds of shell-content constructions. One of these measures was the so-called *attraction-reliance method*. If we take, for instance, the construction Noun + *that* + clause, we may be interested in calculating the strength of the relationship between the noun *fact* and that construction (i.e. *the fact that...*). The attraction-reliance method allows us to do just that by calculating first the frequency of *fact* in that construction in proportion to the total frequency of the construction (attraction) and then the frequency of *fact* in that construction in proportion to the total number of occurrences of the noun in the corpus (Schmid 2010: 107). The attraction-reliance method “captures to some extent the intuition that some nouns are more important for certain constructions than others, and that some constructions are more important for certain nouns than others” (Schmid 2010: 111). Halverson (2017: 30ff) draws on Schmid’s method to introduce two statistical measures intended to gauge 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). Both measures are expressed as percentages. There is no need to provide examples here as plenty of them will come up in what follows.

Table 2.2 shows the ST triggers for *caldre* in EN-CAT both in terms of raw frequency and source concentration. Since the list of trigger types was rather long, triggers with fewer than 10 occurrences were grouped under “Other” for

## 2 Testing the Gravitational Pull Hypothesis on verbs in Catalan

Table 2.2: ST triggers for *caldre* in EN-CAT (n = raw frequency, s.conc = source concentration)

	n	s.conc
need	58	15.30
∅	56	14.78
have to + inf	47	12.40
must + inf	42	11.1
infinitive	32	8.44
other solutions with <i>need</i>	22	5.80
be/become necessary	16	4.22
should + inf	10	2.64
other	94	24.80
misalignments	2	0.52
<b>Total</b>	<b>379</b>	<b>100</b>

the sake of convenience. That is why this category yields such a comparatively large figure. It includes such heterogeneous triggers as imperatives, *ought to* + infinitive, *-ly* adverbs, *require/be required*, *it + take*, *want*, *have got to* + infinitive, and several others. ∅ accounts for triggers with no overt expression of obligation or necessity. The figures for source concentration are relatively low in all cases, which means that no single ST trigger is responsible for the activation of a large percentage of occurrences of *caldre*. The three triggers with source concentration values higher than 10% (apart from ∅ and “Other”) are *need*, *have to* + infinitive and *must* + infinitive, and they range from 11.1% to 15.30%. This suggests low connectivity, as assumed at the stage of hypothesis formulation, but only from the perspective of source concentration. We need to look at the main triggers for *caldre* in order to have the full picture of connectivity patterns.

Table 2.3 shows the TT matching segments of the three main ST triggers for *caldre* (*need*, *have to* + infinitive and *must* + infinitive) in EN-CAT both in terms of raw frequency (n) and target concentration (t.conc). When the number of hits for these three triggers was deemed manageable, all results were manually analysed, as in the case of *need*; when the number was deemed too high for manual analysis, results were thinned, as in the cases of *have to* + infinitive and *must* + infinitive.

The query for *need* (as a verb) yielded 227 matches, with *necessitar* (‘need’) as the top-ranking match with a high target concentration (48.02%). *Caldre* comes second with a target concentration of 24.67%. The query for *have to* + infinitive yielded 523 matches, which were thinned to 250. These 250 were manually sifted

Table 2.3: TT matches for *need*, *have to* and *must* in EN-CAT (n = raw frequency, t.conc = target concentration)

	need		have to		must	
	n	t.conc	n	t.conc	n	t.conc
<b>caldre</b>	56	24.67	16	6.67	15	7.46
<b>necessitar</b>	109	48.02	–	–	–	–
<b>fer falta</b>	21	9.25	–	–	–	–
<b>haver de</b>	14	6.17	166	69.17	142	70.65
<b>no modality</b>	–	–	28	11.67	14	6.97
<b>other</b>	22	9.69	26	10.83	26	12.93
<b>misalignments</b>	5	2.20	4	1.66	4	1.99
<b>Total</b>	227	100	240	100	201	100

and 10 of them were seen to convey meanings other than obligation or intrinsic necessity and consequently removed. Analysis of the remaining 240 hits shows that the top-ranking TT match for *have to* + infinitive is by far the modal periphrasis *haver de* + infinitive, with a high target concentration of 69.17%. That means that *have to* + infinitive is translated as *haver de* + infinitive in over two thirds of the cases. *Caldre* is a poor match for *haver de* + infinitive, with a target concentration of just 6.67%. A similar picture emerges for *must* + infinitive. This query yielded 993 results, which were thinned to 300. These were again manually sifted and 99 of them were discarded because they were instances of *must* conveying strong possibility (i.e. epistemic modality), not obligation or intrinsic necessity. Manual analysis of the remaining 201 instances shows *haver de* + infinitive as the top-ranking match for *must* + infinitive, with a high target concentration value of 70.65%, with *caldre* again a poor second with a target concentration of merely 7.46%.

To sum up, the source concentration of English ST triggers for *caldre* is never too high (15.30 for *need*, 12.40 for *have to* + infinitive, 11.10 for *must* + infinitive), and nor is the target concentration of *caldre* as a Catalan TT match for its English triggers (24.67 for *need*, 6.67 for *have to* + infinitive, 7.46 for *must* + infinitive). But how can these two measures, source and target concentration, be brought together under a single formula that operationalises degree of connectivity, or strength of translation relationships, between items across the two components of a parallel corpus? Both Schmid's attraction-reliance method and Halverson's adaptation in the form of source and target concentration are conceived as measures offering complementary views on connections between two items, but no

## 2 Testing the Gravitational Pull Hypothesis on verbs in Catalan

suggestions for combining these measures are offered. A possible way of bringing them together is through an adaptation of Altenberg's (1999) concept of Mutual Correspondence.<sup>3</sup> The 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 is defined as "the frequency with which different (grammatical, semantic and lexical) expressions are translated into each other" and formulated as follows:

$$\frac{(A_t + B_t) \times 100}{(A_s + B_s)}$$

where  $A_t$  and  $B_t$  = the number of times the compared items (A and B) are translated into each other, and  $A_s + B_s$  = the total number of occurrences of the compared items in the source texts. Since the situation is different here, as the corpus we are using is parallel but not bi-directional, the formula is adapted as follows:

$$\frac{(A_b + B_a) \times 100}{(A_t + B_s)}$$

where  $A_b$  and  $B_a$  = 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  $A_t + B_s$  = the total number of occurrences of A in TT and of B in ST. Moreover, a different name needs to be found, as using the term *mutual* for a translation relationship that is not bi-directional may be misleading.<sup>4</sup> I suggest the alternative term *Unidirectional Translation Correspondence* (UTC), which has the twofold advantage of drawing a parallel with Altenberg's term through the preservation of *correspondence* and explicating the unidirectional nature of the translation relationship.<sup>5</sup> Let us take the pair *caldre/need* as an example. If *need* is translated as *caldre* 58 times, since

---

<sup>3</sup>I would like to thank Sandra Halverson (personal communication) for suggesting this option.

<sup>4</sup>I am indebted to Sandra Halverson (personal communication) for this suggestion.

<sup>5</sup>A different possibility might have been the use of Dyvik's (e.g. 2002) *semantic mirrors method*, which allows the analyst to establish translation correspondences across languages by generating "images" of one word in the other language and then proceeding the other way around with a view to setting up (partly overlapping) semantic fields in both languages. Vandevoorde (2020) put the method to good use with the help of sophisticated statistics-based visual representations. But I can see two reasons for not using it in my research. Firstly, Dyvik's method is intended for use with bi-directional corpora, whereas mine are unidirectional. And secondly, while Vandevoorde aims at the visual representation of semantic fields (more particularly, the field of inchoativity in Dutch translated and non-translated language), my aim is to test a hypothesis on a particular modal indicator. True, in order to do that I need to look at other items in the network, especially as onomasiological salience can only be determined by comparing frequencies of synonyms and near-synonyms. But taking account of the whole semantic field of obligation/necessity in the three languages involved falls outside the scope of my study.

*caldre* occurs 379 times in the Catalan TTs and *need* 227 in the English STs, the UTC of *caldre* and *need* in the English-Catalan sub-corpus will be as follows:  $(58+58) \times 100 / (379+227)=19,14\%$ . The results of applying the same formula to the other two pairs are 10.66% for *caldre/have to* and 8.04% for *caldre/must*.<sup>6</sup>

These figures clearly suggest that the degree of connectivity (operationalised as UTC) between the Catalan modal verb *caldre* and its three main ST triggers in the English-Catalan sub-corpus of COVALT is rather low, which (together with the relatively low magnetism of *caldre*) accounts for its under-representation. The results in Table 2.3 also suggest that *haver de* + infinitive, which shows a high target concentration as a TT segment matching *have to* + infinitive and *must* + infinitive, is likely to display a high degree of connectivity with those two triggers. At the stage of network modelling prior to hypothesis formulation in §4 it was established that *haver de* + infinitive is about 2.5 times as frequent as *caldre* in Catalan non-translations, which suggests that the former is more salient than the latter as an indicator of obligation and necessity. For a full comparison between the two, we now need to look at the source concentration of the main triggers of *haver de* + infinitive with a view to determining the UTC of *haver de* + infinitive and each of these triggers. The whole process carried out for *caldre* must be repeated for *haver de*.

Table 2.4 shows the results for *haver de* + infinitive in the three sub-corpora. As seen above, *haver de* + infinitive is much more frequent than *caldre* in NTR (1.24 vs. 0.47 in normalised frequency per 1,000 words), and the same is valid for EN-CAT (1.55 vs. 0.28) and FR-CAT (1.04 vs. 0.91). Differences are huge indeed in the first two cases, but not so much in FR-CAT. However, when the log-likelihood test is applied, they turn out to be significant in all cases, with LL values at 1,303.55 for EN-CAT (extremely significant), 550.98 for NTR (extremely significant) and 5.36 for FR-CAT (significant at  $p<0.05$ ).

As to degree of connectivity between *haver de* + infinitive and the ST triggers analysed above (*need*, *have to* + infinitive and *must* + infinitive), we already have data for queries in the English-to-Catalan direction. The next (and last) step will

---

<sup>6</sup>On the basis of intuition alone I should have thought that there is no correlation between ST triggers and the four constructions *caldre* can occur in. However, this intuition needed to be confirmed by corpus data. A second manual analysis of the bilingual concordances for *caldre* shows that its distribution across types of construction is not symmetrical, as it occurs 223 times with an infinitive, 74 with a *that*-clause, 45 with a noun phrase and 37 with a zero complement. In relative terms, that amounts to 58.84%, 19.53%, 11.87% and 9.76%, respectively. If this analysis is replicated for each individual trigger (*need*, *must* + infinitive, *have to* + infinitive, etc.), frequency distributions do not exactly match the one just given, but differences are not marked enough to suggest a correlation between the two variables (type of trigger and type of construction *caldre* occurs in).

## 2 Testing the Gravitational Pull Hypothesis on verbs in Catalan

Table 2.4: Query results for *haver de* + infinitive in English-Catalan, French-Catalan and Catalan non-translations (f = normalised frequency per 1,000 words)

	n (words)	Query matches	f
<b>Translations from English</b>	1,343,631	2,088	1.55
<b>Translations from French</b>	566,998	592	1.04
<b>Non-translations</b>	1,551,521	1,924	1.24

be to insert *haver de* + infinitive as query and to look at its ST triggers in order to determine their source concentration for the Catalan modal periphrasis. Table 2.5 offers such information. The 2,088 hits for *haver de* + infinitive were thinned to 250, four of which were manually discarded. On the basis of the remaining 246 matches, the source concentration of ST triggers of *haver de* + infinitive is found not to be very high in any case; that of *must* + infinitive is 20.73% and that of *have to* + infinitive is 17.48%. That means that the occurrence of *haver de* + infinitive in translations from English is not largely dependent on any particular trigger. But, as seen above, the target concentration of *haver de* + infinitive as a TT match for *have to* and *must* is very high. The Unidirectional Translation Correspondence value is 30.95% for *haver de/must* and 27.72% for *haver de/have to*, which is considerably higher than the UTC values for *caldre* and its main ST triggers. Therefore, the connectivity patterns of *haver de* + infinitive with its main ST triggers are stronger than those of *caldre* with its main triggers. That,

Table 2.5: ST triggers for *haver de* in EN-CAT (n = raw frequency, s.conc = source concentration)

	n	s.conc
<b>no modality</b>	57	23.17
<b>must</b>	51	20.73
<b>have to</b>	43	17.48
<b>should</b>	28	11.38
<b>be + inf</b>	12	4.88
<b>other</b>	52	21.14
<b>misalignments</b>	3	1.22
<b>Total</b>	246	100

together with its higher salience, makes *haver de* + infinitive a likelier match than *caldre* for English items conveying obligation or intrinsic necessity.

The data for *caldre* and its triggers retrieved from the French-Catalan corpus are much more straightforward. Hypothesis 2 predicted that *caldre* would be either over- or under-represented in FR-CAT as compared to NTR because gravitational pull and a high degree of connectivity would pull towards over-representation whereas magnetism would pull towards under-representation. Hypothesis 3 predicted that the frequency of occurrence of *caldre* in FR-CAT would be higher than in EN-CAT because over-representation would be favoured by two factors (gravitational pull and a high degree of connectivity) in the former and only one (gravitational pull) in the latter. Hypothesis 3 was confirmed, and for hypothesis 2 over-representation was the case, which suggests that gravitational pull and a high degree of connectivity prevail over the relatively low magnetism of *caldre*.

Table 2.6 shows results for the ST triggers of *caldre* in FR-CAT both in terms of raw frequency and source concentration. The 515 hits for *caldre* in FR-CAT were thinned to 250 and manually analysed. The top-ranking trigger is by far the modal verb *falloir*, with a high source concentration of 68.4%. None of the remaining triggers individually reaches the value of 10%. That means that, when *caldre* occurs in FR-CAT, its occurrence is triggered by *falloir* in over two thirds of the cases. Data for the translation relationship between *falloir* and *caldre* from the source pole are shown in Table 2.7.

The query for *falloir* yields 607 matches, which are thinned to 200 and manually sifted. Two are manually discarded and, for the remaining 198 instances, *caldre* is by far the best represented Catalan match for *falloir*, with a high target concentration of 59.1%, with *haver de* + infinitive a poor second at 10.10%. This

Table 2.6: ST triggers for *caldre* in FR-CAT (n = raw frequency, s.conc = source concentration)

	n	s.conc
<b>falloir</b>	171	68.4
<b>no modality</b>	17	6.8
<b>devoir</b>	12	4.8
<b>other</b>	46	18.4
<b>misalignments</b>	4	1.6
<b>Total</b>	250	100

## 2 Testing the Gravitational Pull Hypothesis on verbs in Catalan

Table 2.7: TT matches for *falloir* in FR-CAT (n = raw frequency, t.conc = target concentration)

	n	t.conc
<b>caldre</b>	117	59.1
<b>haver de</b>	20	10.10
<b>no modality</b>	11	5.55
<b>other</b>	34	17.17
<b>unclear</b>	10	5.05
<b>misalignments</b>	6	3.03
<b>Total</b>	198	100

suggests strong translation links between *falloir* and *caldre* from both perspectives – a suggestion confirmed by their UTC, which stands at 62.74%. Connectivity patterns between *caldre* and its main French trigger, *falloir*, are very strong. That seems to be the main reason for over-representation of *caldre* in FR-CAT, together with salience of *falloir* in the French modal network for obligation and necessity, which was established on the basis of previous studies (Labbé & Labbé 2013; Lewis 2015).

For the analysis based on FR-CAT to be parallel in all respects to that based on EN-CAT, it would now be the time to look at the ST triggers of *haver de* + infinitive in FR-CAT. However, in EN-CAT that step was justified by the fact that *haver de* + infinitive was better represented as a target match for *have to* + infinitive and *must* + infinitive than *caldre*, whereas the case is otherwise for *falloir* in FR-CAT, with *caldre* as the top-ranking target match and *haver de* + infinitive with a relatively low target concentration of 10.10%. Therefore, it is not necessary to perform that query, which would probably show a higher source concentration of *devoir* (the other major verb conveying obligation and intrinsic necessity in French) than was the case with *caldre*.

It may be in place at this point to recapitulate the results of the corpus analysis reported on in this section. It was initially established that *caldre* is less salient in the TL than its main alternative in the obligation/intrinsic necessity network, *haver de* + infinitive. Therefore, magnetism can only be expected to play a minor role in the creation of translation effects. Even so, it is over-represented in FR-CAT as compared both to NTR and EN-CAT (hypotheses 2 and 3). This may be accounted for by strong connectivity between *caldre* and *falloir* (attested by the data) and, perhaps, by the gravitational pull of *falloir* (not tested for but re-

flected in the literature). On the other hand, connectivity between *caldre* and its main English ST triggers (*need*, *have to* + infinitive and *must* + infinitive) is low, which, added to the relatively low magnetism of *caldre*, results in its under-representation in EN-CAT. *Haver de* + infinitive, on the contrary, is over-represented in EN-CAT. This may be accounted for by the relatively high salience of *haver de* and the relatively strong connectivity between *haver de* and two of its ST triggers (*must* and *have to*).

## 6 Conclusions

The case of *caldre* shows that connectivity may tip the scales in favour of over- or under-representation. Through the use of two comparable parallel corpora with the same target language, TL salience is controlled for, as there is no reason to think that a certain TL item will be more salient in one corpus than in the other. Connectivity, in the present study, seems to be favoured by formal similarity, which comes in two forms. The first is syntactic isomorphism. In the French-Catalan combination, both *caldre* and *falloir* are mainly used in impersonal constructions that share two possible kinds of complements – infinitives and *that*-clauses. In the English-Catalan combination, *haver de* + infinitive can be used in both personal and impersonal constructions; and, whenever *must* + infinitive, *have to* + infinitive or *need* are used in a personal construction, there is a higher degree of overlap with *haver de* than with *caldre*.

The second factor is phonological/graphological similarity: *haver de* and *have to* display that kind of similarity, which would seem to foster connectivity at a very basic level. They are not cognates, as Latin *habere* and the Proto-Germanic root of English *have* are not etymologically related; but they could easily pass for cognates on the basis of phonological/graphological similarity. Cognate status is often deployed as an independent variable in psycholinguistic experiments on word translation. De Groot (1992b) is a case in point. This author sets out to measure translation performance (operationalised as reaction time, number of omissions and number of translation errors) under varying conditions. With regard to cognate status, her results lead her to conclude that “in addition to being translations, cognates have an extra reason to be linked in lexical memory. This could be reflected in relatively strong T1 links” – T1 links being links between lexical nodes at the level of lexical memory, without resorting to conceptual memory. Translation between cognates, then, would be favoured by strong connectivity of a special kind; and there is no reason to suppose that this cannot hold true for *false* cognates too, since links between lexical nodes cannot be expected to reflect expert etymological knowledge.

## 2 Testing the Gravitational Pull Hypothesis on verbs in Catalan

The GPH is not incompatible with other models of the translation process. Carl et al. (2019) present a model based on the concept of *entropy*, borrowed from the fields of physics and information theory. Entropy describes “the amount of disorder in a system” (Carl et al. 2019: 217). In a context of translation, the more possible translations are activated in word and phrase translation systems, the higher the entropy. When entropy is high, the translator needs to invest a great deal of effort to find a solution. When a translator finds a complex word or structure for the first time, the information available is low and the degree of entropy at its highest, so much cognitive energy is required. Finding a satisfactory solution creates internal structure and reduces the degree of entropy, so less cognitive energy will need to be spent when the same word or structure recurs a second or third time. The process of entropy reduction over time is captured by the concept of *entropic gravity*. Entropy may arise from variability both in lexical and syntactic choices. In the model presented by these authors, activation of translation solutions in a system is non-selective for language, as elements are activated in the system on the basis of phonological and semantic associations in both languages. This initial stage is followed by a task-dependent decision process in which elements activated solely on the basis of phonological similarity, or belonging to the source language, are discarded and a satisfactory translation solution reached.

Carl et al. (2019: 226) claim that their model “relates to Halverson’s (2003) *gravitational pull hypothesis*”. However, they think it “unfortunate” (Carl et al. 2019: 227) that Halverson should have split her initial concept of gravitational pull into the three causes of translational effects mentioned above, among other reasons because “each of Halverson’s salience, link and connectivity effects might be more simply and coherently described in terms of entropic gravity, which assumes similar underlying mechanisms for producing the various translational effects” (Carl et al. 2019: 227). They further claim that there may be more than just three causes of translational effects (Carl et al. 2019: 227), although they do not mention any. Pending specification of such causes, it may be safe to stick to the three posited by Halverson. However, *factors* may be introduced that favour the activation of these causes. The research reported on in this paper suggests that syntactic isomorphism and phonological/graphological similarity strengthen connectivity – or, alternatively, entropic gravity by reducing the degree of entropy. This should not be seen as an attempt to alter the GPH in any fundamental way, but to refine it by introducing the notion of factor. Only further research will determine whether the attempt is worth pursuing or not.

## Acknowledgements

This work was supported by Universitat Jaume I [UJI-B2017-58] and by the Spanish Ministry of Science and Innovation [PID2019-103953GB-I00].

## References

- Altenberg, Bengt. 1999. Adverbial connectors in English and Swedish: Semantic and lexical correspondences. In Hilde Hasselgård & Signe Oksefjell (eds.), *Out of corpora: Studies in honour of Stig Johansson*, 249–268. Amsterdam: Rodopi.
- Badia i Margarit, Antoni. 1994. *Gramàtica de la llengua catalana: Descriptiva, normativa, diatòpica, diastràtica*. Barcelona: Enciclopèdia Catalana.
- Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*, 233–250. Amsterdam: John Benjamins.
- Carl, Michael, Andrew Tonge & Isabel Lacruz. 2019. A systems theory perspective on the translation process. *Translation, cognition and behavior* 2(2). 211–232.
- De Groot, Annette M. B. 1992a. Bilingual lexical representation: A closer look at conceptual representations. In Ram Frost & Leonard Katz (eds.), *Orthography, phonology, morphology, and meaning*, 389–412. Amsterdam: North Holland.
- De Groot, Annette M. B. 1992b. Determinants of word translation. *Journal of experimental psychology: learning, memory, and cognition* 18(5). 1001–1018.
- Dyvik, Helge. 2002. Translations as semantic mirrors: From parallel corpus to wordnet. In Karin Aijmer & Bengt Altenberg (eds.), *Advances in corpus linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23)*, 311–326. Amsterdam: Rodopi.
- Gavarró, Anna & Brenda Laca. 2002. Les perífrasis temporals, aspectuals i modals. In Joan Solà, Maria Rosa Lloret, Joan Mascaró & Manuel Pérez-Saldanya (eds.), *Gramàtica del català contemporani*, vol. 3, 2663–2726. Barcelona: Empúries.
- Halverson, Sandra. 2003. The cognitive basis of translation universals. *Target* 15(2). 197–241.
- Halverson, Sandra. 2010. Cognitive translation studies: Developments in theory and method. In Gregory M. Shreve & Erik Angelone (eds.), *Translation and cognition*, 349–70. Amsterdam: John Benjamins.
- Halverson, Sandra. 2017. Developing a cognitive semantic model: Magnetism, gravitational pull, and questions of data and method. In Gert de Sutter, Marie-Aude Lefer & Isabelle Delaere (eds.), *Empirical translation studies: New methods and theoretical traditions*, 9–45. Berlin: Mouton de Gruyter.

## 2 Testing the Gravitational Pull Hypothesis on verbs in Catalan

- Hareide, Lidun. 2017. 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 Meng Ji, Michael Oakes, Li Defeng & Lidun Hareide (eds.), *Corpus methodologies explained: An empirical approach to translation studies*, 188–231. London: Routledge.
- Labbé, Cyril & Dominique Labbé. 2013. La modalité verbale en français contemporain: Les hommes politiques et les autres. In David Banks (ed.), *La modalité, le mode et le texte spécialisé*, 33–61. Paris: L'Harmattan.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. Vol. 1: Theoretical prerequisites. Stanford: Stanford University Press.
- Langacker, Ronald W. 1991. *Foundations of cognitive grammar*. Vol. 2: Descriptive Applications. Stanford: Stanford University Press.
- Lewis, Diana. 2015. A comparable-corpus based approach to the expression of obligation across English and French. *Nordic Journal of English Studies* 14(1). 152–173.
- Malmkjær, Kirsten. 2008. Norms and nature in translation studies. In Gunilla Anderman & Margaret Rogers (eds.), *Incorporating corpora: The linguist and the translator*, 49–59. Clevedon: Multilingual Matters.
- Payrató, Lluís. 2002. L'enunciació i la modalitat oracional. In Joan Solà, Maria Rosa Lloret, Joan Mascaró & Manuel Pérez-Saldanya (eds.), *Gramàtica del català contemporani*, vol. 2, 1149–1220. Barcelona: Empúries.
- Radden, Günter & René Dirven. 2007. *Cognitive English grammar*. Amsterdam: John Benjamins.
- Rigau, Gemma. 1999. Los predicados impersonales relativos en las lenguas románicas. *Revista española de lingüística* 29(2). 317–355.
- Rigau, Gemma. 2005. Estudi microsintàctic del verb *caldre* en el català antic i en l'actual. *Caplletra* 38. 241–258.
- Schmid, Hans-Jörg. 2010. Does frequency in text really instantiate entrenchment in the cognitive system? In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 101–133. Berlin: Walter de Gruyter.
- Tirkkonen-Condit, Sonja. 2002. Translationese—a myth or an empirical fact? *Target* 14(2). 207–220.
- Tirkkonen-Condit, Sonja. 2004. Unique items—over- or underrepresented in translated language? In Anna Mauranen & Pekka Kujamäki (eds.), *Translation universals: Do they exist?*, 177–184. Amsterdam: John Benjamins.
- Toury, Gideon. 1995. *Descriptive translation studies and beyond*. Amsterdam: John Benjamins.

*Josep Marco*

Vandevorde, Lore. 2020. *Semantic differences in translation: Exploring the field of inchoativity*. Berlin: Language Science Press.