

Chapter 5

Using the Gravitational Pull Hypothesis to explain patterns in interpreting and translation: The case of concatenated nouns in mediated European Parliament discourse

Marie-Aude Lefer^a & Gert De Sutter^b

^aUniversité catholique de Louvain ^bGhent University

In this chapter, we present a corpus study of the French rendition of English concatenated nouns, such as climate change, comparing two modes of interlingual mediation at the European Parliament, namely simultaneous interpreting and written translation. Using parallel corpus data extracted from the European Parliament Translation and Interpreting Corpus, we examine how frequently English concatenated nouns are rendered with semantically equivalent items in the two mediation modes, and which factors stimulate the use of these equivalent (vs non-equivalent) renditions. Alongside the complexity and lexicalization of English concatenated nouns, we consider several frequency-related variables inspired by Halverson2017's (Halverson2017) cognitive linguistic model of translation, the gravitational pull hypothesis. The model posits three cognitive sources of translation effects: gravitational pull (source salience), connectivity (cross-linguistic link strength) and magnetism (target salience). The results show that there are far fewer semantically equivalent renditions in interpreting than in translation. In addition, the regression analysis provides strong evidence that connectivity and magnetism play a crucial role in the selection of semantically equivalent vs non-equivalent renditions in interpretations and translations, alongside the length of source concatenated nouns, with stronger effects in interpreting. By contrast, source-language variables related to gravitational pull and lexicalization do not seem to influence renditions in French. The study brings to the fore key commonalities between translation and interpreting and shows that the three cognitive sources in Halverson's gravitational pull model can be successfully disentangled in a multifactorial research design.

Using the Gravitational Pull Hypothesis to explain patterns in interpreting and translation: The case of concatenated nouns in mediated European Parliament discourse. To appear in: Marta Kajzer-Wietrzny, Adriano Ferraresi, Silvia Bernardini and Ilmira Ivaska. Empirical investigations into interpreting Berlin: Language Science Press. [preliminary page numbering]

1 Introduction

The last 20 years have seen the application of a wide array of corpus-based and corpus-driven techniques to increasingly large amounts of translated text, in many languages. Corpus-based translation studies (CBTS) has produced numerous descriptions of translation-related phenomena, ranging from translation procedures for specific linguistic items and structures (e.g. culture-specific lexis) to typical features of translated text (e.g. increased explicitness). In recent years, in the wake of Shlesinger's pioneering work in corpus-based interpreting studies (CIS) (Shlesinger1998), CBTS has progressively branched out to include intermodal studies, where different mediated language varieties are compared (typically, written translation and simultaneous interpreting; BernardiniEtAl2016). This type of intermodal research has been further promoted by Kotze2020's (Kotze2020) constrained-language framework, which aims to identify the commonalities between language varieties where constraints of different kinds play an above-average role (see e.g. Kajzer-WietrznyIvaska2020). The key constraint dimension along which translation and interpreting differ is the 'register/modal-ity' dimension, as translation and interpreting represent written and spoken language production respectively. What they have in common is that they both rely on a preexisting text (the source text or speech) and involve bilingual language processing, where two languages are simultaneously activated, one as the source, the other as the target.

The present study adds to the growing body of intermodal corpus research by examining the French renditions of English noun concatenations (i.e. sequences of at least two nouns, such as *food prices*) in two modes of interlingual mediation commonly practiced at the European Parliament (EP), namely simultaneous interpreting of the speeches delivered during EP plenary sessions and written translation of the official verbatim reports of these speeches. The reason for examining concatenated nouns is that they have been described as difficult, error-prone items in both modalities. Intermodal research on this topic, however, is still scarce. In the present study, we aim to assess the impact of a large set of frequency-, complexity- and lexicalization-related variables on the use of semantically equivalent (vs non-equivalent) renditions in translation and interpreting, drawing both on insights from previous empirical research on noun sequences and on Halverson's (Halverson2003; Halverson2007; Halverson2010; Halverson2017) cognitive linguistic model of translation, the gravitational pull hypothesis. The model posits three cognitive sources of translational effects: source language salience (*gravitational pull*), target language salience (*magnetism*)

citation
not in
bib

and cross-linguistic link strength (*connectivity*), where salience is operationalized as, among other things, frequency of use. To date, the model has been tested on a handful of linguistic items, such as morphemes and individual lexemes (Hareide2016; Vandevoorde2020; Marco2021), but it has rarely been used to study items above the word level. However, we believe that the model holds great potential for the study of structures such as concatenated nouns, since psycholinguistic research has shown the crucial role played by frequency in processing and producing these items (BaayenEtAl2010). In addition, to the best of our knowledge the model has not been applied to interpreting, nor has it been used in robust multifactorial research designs such as the one we propose here.

citation
not in
bib

The chapter is structured as follows. §2 presents the phenomenon under scrutiny here, English concatenated nouns and their French equivalents, introduces Halverson's gravitational pull model and shows how it can be used to inform the inter-modal study of concatenated nouns. In §3, we describe the corpus data used, the data extraction and coding procedures adopted and the multivariate statistics applied to the dataset at hand. The results of the analysis are presented and discussed in §4. The chapter ends with concluding remarks and suggestions for future research.

2 Background

2.1 Concatenated nouns in English-French translation and interpreting

The notion of 'concatenated noun' is a blanket term for two main types of noun sequence: (1) established (i.e. institutionalized and lexicalized) compounds and multiword terms (e.g. *car insurance*, *food chain*) and (2) non-institutionalized, nonce formations, which are created ad hoc (e.g. *kinship child*, *poultry and pig establishment*) (Bauer1983; Hohenhaus2005; note that there is no watertight borderline between the two categories, see Bauer1998). English noun concatenations encompass several structures (also called *patterns* or *schemas*), including N+N (e.g. *pork products*), [N+N]+N (e.g. *trade defence instruments*), [A+N]+N (e.g. *national unity government*) and A+[N+N] (e.g. *small market share*).

Three main aspects of concatenated nouns in English and French are worth considering in contexts of bilingual language production: the complex semantics of English concatenated nouns, and English-French cross-linguistic differences in pattern productivity and word order. First, English N+N sequences are semantically versatile, i.e. they convey a large variety of semantic relations. Classic taxonomies typically range from circa 10 to 50 semantic relations, thereby displaying

various degrees of granularity (Fernandez-Dominguez2020). For instance, Levi (Levi1978: 75–118; quoted in Fernandez-Dominguez2020) lists nine semantic relations found in English N+N sequences: CAUSE, HAVE, MAKE, USE, BE, IN, FOR, FROM and ABOUT (*chocolate éclair* ‘an éclair which *has* chocolate’, for example, illustrates the HAVE relation). In addition, some N+N concatenations display semantic indeterminacy, i.e. they cannot easily be disambiguated or interpreted, even when their co-text is taken into consideration. In a corpus-based study of more than 500 N+N compounds, Fernandez-Dominguez2020 finds that a third of the items under scrutiny can be attributed a second reading on top of the most obvious, primary reading (e.g. *army plan*: ABOUT ‘a plan which is *about* the army’ vs IN ‘a plan which is prepared/implemented *in* the army’). Second, the core N+N pattern, which is attested in both English and French (e.g. *coin cuisine* lit. ‘corner kitchen’), is much more productive in English (Paillard2000; ArnaudRenner2014). As a result, many English N+N sequences need to be rendered in French by means of other patterns, such as N+A (e.g. *trade agreement* > *accord commercial*) or N+prep+N (e.g. *security wall* > *mur de sécurité*), with cases where the two patterns are found to alternate (e.g. *fishing stocks*: *stocks halieutiques*_{N+A} vs *stocks de poissons*_{N+PREP+N}). Finally, as regards constituent order, English N+N sequences are typically right-headed (e.g. *timber products*), while their French equivalents, whatever the pattern, are mostly left-headed (e.g. *produits du bois*). This aspect of concatenated nouns has been examined in compound acquisition research, where English-French bilingual children have been shown to produce N+N novel compounds in reversed order (i.e. left-headed in English and right-headed in French), under the influence of crosslinguistic transfer (Nicoladis2002; see also DeCatEtAl2015).

The contrastive literature on English-to-French translation mentions two major types of translation difficulty (see e.g. ChuquetPaillard1987). The first is that, in addition to obvious shifts in word order, English concatenated nouns often require explication of the semantic relation that holds between the constituents of the sequence (which, in English, is not overtly expressed), for instance through the insertion of prepositions (e.g. *adoption law* > *loi sur l’adoption*, *foreign policy objectives* > *objectifs en matière de politique étrangère*). Explication of the semantic relation between head and modifier(s) in the concatenation is no easy task, in view of the above-mentioned semantic versatility of the English N+N pattern. The second difficulty frequently mentioned in the contrastive literature is that the underlying structure of some of the longer English sequences is potentially ambiguous and hence difficult to interpret and translate. This is often the case when an adjective or a noun premodifies an N+N sequence (e.g. *modern history section*: [*modern history*] *section* vs *modern* [*history section*]). This causes acute

problems in learner translation. In their error analysis of English-to-French student specialized translations, **KueblerEtAl2022** find numerous translation errors triggered by English noun phrases whose structure is ambiguous (e.g. *stable solution complexation* > **complexation stable de solution* instead of *complexation en solution stable*).

Similar difficulties are also discussed in the field of interpreting, where it is stressed that the interpretation of English concatenated nouns is effortful because they are informationally very dense and require major syntactic changes (i.e. reordering) in French and other Romance languages (see e.g. **Gile1995** on proper name compounds). Noun concatenations have been investigated empirically in CIS. Relying on the *European Parliament Interpreting Corpus*, **Ghiselli2018** analyzes Italian interpretations of English complex noun phrases (phrases where nominal heads are premodified by several items, be they nouns, adjectives, numbers or participles). She finds that only 55% of English complex noun phrases are rendered successfully in Italian. In her dataset, incomplete or wrong renditions are particularly prominent when source speech delivery is fast (180+ words per minute), pointing to the effect of time constraints on interpreters' renditions of complex noun phrases. In their French-Dutch study based on the *European Parliament Interpreting Corpus Ghent* and the parliamentary debate sub-corpus of the *Corpus Gesproken Nederlands*, **DefrancqPlevoets2018** investigate intra-word filled pauses, including intra-compound pauses, in Dutch interpreted from French and in non-mediated (original) Dutch. The authors provide tentative evidence that "compounds are an important factor in the increase of cognitive load during interpretation" (ibid. 57). As acknowledged by the authors themselves, however, the data sample analyzed is small (the analysis is based on 18 occurrences of intra-compound filled pauses in interpreted Dutch).

In a pilot study based on the *European Parliament Translation and Interpreting Corpus* (**FerraresiBernardini2019**), **LeferDeClerck2021** find that English concatenated nouns are interpreted with French semantically equivalent renditions in only half of the cases, the other half being made up of incomplete and wrong renditions, as well as omissions. Although based on a small dataset, their qualitative analysis of the disfluencies typically found in the vicinity of incomplete or wrong renditions suggests that three types of N+N sequence are particularly vulnerable in interpreting: ad hoc (i.e. non-lexicalized) sequences, long sequences (made up of 3 constituents or more) and rare (i.e. infrequent) sequences. These preliminary findings point to the potential role of the lexicalization, length and frequency of English concatenated nouns in shaping the use of semantically (non-)equivalent renditions in French. Although admittedly very tentative, this ties in with the ample evidence provided by psycholinguistic studies on compound processing,

where it is shown that compound and constituent length and frequency all play a decisive role in lexical access (see e.g. BaayenEtAl2010). To date, however, length and frequency have not been examined concomitantly in robust multifactorial research designs in corpus-based translation and interpreting studies devoted to concatenated nouns (or compounds in general). This is what we intend to do in the present study, relying on Halverson's gravitational pull model to inform cognitively motivated frequency analyses.

2.2 Applying Halverson's gravitational pull model to the study of concatenated nouns

Combining insights from cognitive grammar, psycholinguistic approaches to bilingualism and second language acquisition research, Halverson2017 posits three cognitive sources of translational effects (patterns of under- and overrepresentation, source-language interference, normalization, etc.): (1) source language salience (*gravitational pull*), (2) target language salience (*magnetism*) and (3) cross-linguistic link strength (*connectivity*), where salience is operationalized as frequency of use and ease of recall. Gravitational pull is described as "a cognitive force that makes it difficult for the translator to escape the cognitive pull of highly salient representational elements in the source language" (ibid. 14). This force can cause interference in translation. Magnetism is a force that affects the cognitive search for a target language item, whereby "the translator is more likely to be drawn to a target language item with high salience/frequency" (ibid.). Connectivity is "the nature and strength of links between elements in a bilingual's two languages" (ibid.). Halverson's hypothesis is that "the more established (entrenched) a link is, the more likely it will be activated and used in translation, and vice versa" (ibid. 15).

In her 2017 study, Halverson takes as a test case the English polysemous verb *get* and two of its Norwegian equivalents, triangulating monolingual and parallel corpus data, elicitation data and keystroke logs. While her results provide initial support for the posited cognitive forces, some of the predicted overrepresentation patterns are not found in the corpus and keystroke data examined. Vandevoorde2020 uses the model as a post-hoc interpretative framework in her corpus-based study of the Dutch inchoative verbs *beginnen* 'begin' and *starten* 'start' in Dutch translated from English and French and in non-translated Dutch. Vandevoorde shows that the model can be used to explain some of the patterns observed in her data, but she acknowledges that in some specific cases, several cognitive forces overlap (e.g. gravitational pull and magnetism), making it difficult to disentangle their cumulative effects.

In addition to polysemous verbs, the gravitational pull model has also been tested on unique items, i.e. linguistic items that “lack straightforward linguistic counterparts in other languages” (Tirkkonen-Condit2004). As pointed out by Tirkkonen-Condit, unique items are not necessarily untranslatable, rather, “they are simply not similarly manifested (e.g. lexicalized) in other languages” (ibid.). Typically, in this context, translated texts from two source languages are compared: one source language where a given phenomenon is not attested, the other where it is. For example, Hareide2016 has examined the Spanish gerund in texts translated from English (a language with progressive and non-finite adverbial phrases) and Norwegian (a language that has no gerund), providing strong support for the gravitational pull model. A similar approach is taken in MarcoOster2018, which deals with diminutive suffixes in Catalan translated from German (which has productive diminutive suffixes) and English (which has no productive diminutive suffix), and in Marco2021, devoted to modal verbs expressing obligation and necessity in Catalan translated from English and from French.

Remarkably, few empirical studies have examined items and structures above the morpheme or word level (Halverson2017) or used multifactorial statistical testing to account for the relative strengths of the three cognitive forces at play. Also notable is the fact that the model has attracted little attention in corpus-based interpreting research to date. In this chapter, we set out to go some way towards remedying these gaps and further exploiting the full potential of Halverson’s cognitive model by applying it to a structure situated above the word level (concatenated nouns), in two types of interlingual mediation (written translation and simultaneous interpreting), using multivariate statistics (regression analysis). In doing so, we also aim to extend previous translation and interpreting research on complex noun phrases and nominal compounds by relying on several corpus frequency counts that function as operationalizations of the three cognitive forces included in Halverson’s model, namely *gravitational pull* (frequency of the concatenation and its constituents in the source language, here English), *connectivity* (cross-linguistic correspondence of the English source concatenation and its French translation or interpretation) and *magnetism* (frequency of the rendition in the target language, here French). The frequency variables drawn from the gravitational pull model will be considered alongside other factors that have been shown to influence compound processing, namely length and lexicalization, with a view to singling out the factors that condition the use of semantically equivalent (vs non-equivalent) renditions in French. We expect similar trends to emerge in the two modalities, namely that lexicalization, short length and high frequency (reflecting strong gravitational pull, strong connectivity and/or

strong magnetism) will go hand in hand with semantically equivalent renditions. However, we expect the effects of these variables to be more visible in interpreting, in view of the fact that “[b]ecause interpreting affords only limited opportunity for restatement or corrections, it can be seen as the practitioner’s default version, with written translation representing a more polished rendition” (ShlesingerMalkiel2005). Contrary to interpreting, written translation is an offline activity, often involving the use of resource tools, self-revision and editorial intervention. In other words, our prediction is that ad hoc concatenations, long concatenations, and concatenations that display low gravitational pull, low connectivity and/or whose equivalents display low magnetism will trigger incomplete renditions, wrong renditions and omissions more frequently in interpreting than in translation.

3 Data and methodology

3.1 Corpus data used

In this study, we made use of corpus data extracted from the *European Parliament Translation and Interpreting Corpus* (EPTIC; BernardiniEtAl2016; FerraresiBernardini2019)¹. EPTIC is a multilingual intermodal corpus developed at the University of Bologna in collaboration with other European universities, among them UCLouvain in the case of the English-French language pair. The corpus comprises four components, two spoken, two written: transcripts of speeches delivered at the EP and transcripts of their simultaneous interpretations; verbatim reports of the same speeches and their official written translations. Transcriptions are performed on the basis of the videos of the plenary sessions made available online by the EP, adhering to detailed transcription conventions specifically developed for EPTIC. Verbatim reports and their official translations are derived from the EP website, where EP proceedings are archived and available to the public. One of the unique features of the corpus is that the source speeches (spoken component) and the source verbatim reports (written component) are almost identical, which makes it possible to study the interpretations and translations of practically the same input. The corpus, whose compilation is still ongoing at the time of writing, is made available to the research community through the NoSketch Engine platform (Rychly2007), the open source version of Sketch Engine (KilgarriffEtAl2014). It is sentence-aligned and POS-tagged.

¹<https://corpora.dipintra.it/eptic/>

5 Gravitational Pull Hypothesis explains interpreting and translation patterns

In this study, we relied on 106 speeches delivered in English by Members of the European Parliament, commissioners and guests, and their French simultaneous interpretations by highly skilled professionals who are all native speakers of French. We also used the verbatim reports of these speeches and their French translations. No information is available on the translators who produced the translations included in EPTIC, but it can be assumed that they are also highly skilled professionals translating into their native language. The spoken and written components of the subcorpus used in the study each total ca 60,000 tokens (see Table 1).²

Table 1: Size of the English-to-French EPTIC subcorpus used in the study (in tokens)

	English sources	French targets	Total
Spoken component	29,457	28,317	57,774
Written component	28,068	31,897	59,965

To code the EPTIC dataset with reference frequencies in English, French and English-French translation (see §3.2), we used the Europarl corpus as a reference corpus (Koehn2005). Europarl is a multilingual parallel corpus that comprises the EP verbatim reports produced between 1996 and 2011 (the year that translation of the reports was discontinued at the EP). Europarl is here taken to be representative of EP discourse as a whole, monolingually (EP discourse in English and EP discourse in French) and bilingually (EP discourse in the English-French pair). We used version 7 (Europarl7), available on Sketch Engine. The English version of Europarl7 totals 53+ million tokens. It is a mix of verbatim reports of speeches originally delivered in English (by native or non-native speakers of English) and speeches originally delivered in other languages and subsequently translated into English. The French version of Europarl7 contains 59+ million tokens. Like the English version of the corpus, it is comprised of verbatim reports of speeches originally delivered in French (by native speakers of French, with few exceptions) alongside speeches delivered in other languages and translated into French (in some cases with English as a pivot language). The English and French versions of the Europarl corpus, which are sentence-aligned, was also used in the

²The source speeches included in the English-to-French EPTIC subcorpus used in the study were read-out (44% of the subcorpus), impromptu (34%) or mixed (22%), with an average speed of delivery of 161 words per minute. They were given by both native and non-native speakers of English (corresponding to 60% and 40% of the subcorpus, respectively).

present study as a parallel corpus. It is important to stress, however, that it was used here as a *non-directional* parallel corpus, in the sense that we disregarded translation directions and use of English as a pivot (Lefer2020). In other words, the full English-French parallel Europarl used as a reference corpus in the study includes texts in original English translated into French, texts in original French translated into English and texts produced in other languages and translated into both English and French.

3.2 Data extraction and coding

English concatenated nouns used in EPTIC source speeches and verbatim reports were automatically extracted on the basis of a CQL query aimed at identifying all sequences of at least two common nouns. Irrelevant occurrences were then manually removed, i.e. POS-tagging errors (e.g. *the consequences of printing money* too cheaply), contiguous nouns that are not concatenated (e.g. *all the remarks people* have made) and strings containing titles (e.g. *madam chairman*).

The resulting dataset contains 853 occurrences, equally distributed among the spoken and written components of the subcorpus used³, which were then manually matched with their renditions in interpreted and translated outputs, relying on EPTIC sentence alignment. All occurrences were coded for the response variable ‘semantically equivalent rendition’ vs ‘semantically non-equivalent rendition’. Transfer of meaning is central to this distinction. The ‘semantically equivalent rendition’ category was attributed to outputs where the propositional content conveyed by the source concatenation was also found in the interpretation or translation, as in *euro crisis* > *crise de l’euro* and *tax evasion* > *évasion fiscale*. This category also includes (rare) cases where the semantic relation that holds between the constituents of the source concatenation is explicitated in the output (e.g. *rural development policy* > *politique dans le domaine du développement rural*) (cf. Wadensjoe1998 on expanded renditions). The ‘semantically non-equivalent rendition’ category, by contrast, subsumes three types of rendition: (1) incomplete renditions, where part of the propositional content found expressed in the source concatenation is left out in the output (e.g. *adoption process* > *processus* ‘process’, *European fishing industry* > *industrie européenne* ‘European industry’); (2) wrong renditions, where the propositional content of the rendition is not semantically equivalent to that of the original (cases of misinterpretation, incorrect meaning, etc.) (e.g. *export figures* > *importations* ‘import’, *partner country* > *pays d’origine* ‘country of origin’) (AmatoMack2011); and (3) omissions, when

³All but nine occurrences occur in both source speeches and corresponding verbatim reports.

source concatenations are entirely omitted in the output.

The data were also coded for the following explanatory variables: speech-text id (unique id attributed to pairs of source speeches and corresponding verbatim reports), modality (translation or interpreting) and ten gravitational-pull-, magnetism-, connectivity-, complexity- and lexicalization-related variables, which are all described and illustrated below.

3.2.1 Gravitational-pull-related variables

The gravitational pull of source concatenated nouns (i.e. their salience in English EP discourse) was operationalized by means of two corpus frequency variables: (i) their overall frequency and (ii) the average frequency of their individual constituents. Relative frequencies per million words were computed on the basis of the full English version of Europarl7 (53+ million tokens). The reason for operationalizing gravitational pull both as concatenation frequency and constituent frequency is that, as mentioned in §2.2, psycholinguistic research has shown that nominal compounds are accessed both as wholes and via their component parts (BaayenEtAl2010; Gagne2011). Examples of noun concatenations with particularly strong gravitational pull in EP discourse (at the level of the whole concatenation) include *labour market*, *action plan* and *climate change*, while weak gravitational pull items are, for instance, *birth country*, *legality assurance system* and *tuna processing facilities*. Some of the items that display a weak concatenation-based gravitational pull exert a rather strong pull at the level of their individual constituents, such as *information measures*, *development needs*, *security situation* and *market construction*, which shows the usefulness of including different frequency operationalizations of gravitational pull when dealing with items above the word level.

3.2.2 Magnetism-related variable

The magnetism (i.e. salience) of the French renditions found in interpreted and translated outputs was operationalized as their overall frequency (normalized per million words) in the French version of Europarl7 (59+ million tokens). Renditions with strong magnetism in French EP discourse include, for instance, *sécurité alimentaire*, *proposition de resolution* and *états membres*. Examples of weak-magnetism renditions are *accords en matière de pêche*, *droit familial* and *coût de l'énergie*.

3.2.3 Connectivity-related variables

The strength of the cross-linguistic link between a given source noun concatenation and its rendition in translation or interpreting (i.e. connectivity) was operationalized on the basis of both bilingual lexicographic/terminographic and parallel corpus data. First, we coded whether the source noun concatenation and its rendition were recorded as equivalents in English-French bilingual entries (i) in the Interactive Terminology for Europe (IATE) database and (ii) in the subscription-based Oxford English-French bilingual dictionary. We chose to rely on the Oxford bilingual dictionary because contrary to other online English-French dictionaries, the two sides of the dictionary can be queried simultaneously, with direct access to main entries and subentries. Three values were used to code these two connectivity-related lexicographic variables: YES (when the concatenation and its rendition were listed as equivalents in IATE or the Oxford bilingual dictionary, e.g. *interest rate* > *taux d'intérêt*), NO (when they were not, e.g. *dioxin scare* > *alerte à la dioxine*) and PARTIAL (for longer concatenations, when part of the source concatenation and part of its rendition were recorded as equivalents in a bilingual entry in IATE or Oxford, e.g. *draconian maternity leave* > *congé de maternité draconien*). In addition, we relied on a frequency-based variable, taking advantage of the parallel nature of the Europarl corpus. For each pair of source concatenation and corresponding rendition in either interpreting or translation, we computed a Pointwise Mutual Information (PMI) score on the basis of (i) the frequency of the source noun concatenation in Europarl7-English, (ii) the frequency of its rendition in Europarl7-French and (iii) the frequency of their cross-linguistic correspondence in the English-French parallel version of Europarl7. Specifically, we used the following formula, where p = probability, s = source (English noun concatenation), t = target (French rendition), $s-t$ = source-target correspondence: $\log (p(s-t)^3 / p(s) * p(t))$ (RoleNadif2011). The higher the PMI³ score, the stronger the connectivity. For example, the pair *health services-services de santé* displays a stronger cross-linguistic link in EP discourse (PMI³ = 2.77) than the pair *health services-services en matière de santé* (PMI³ = -6.93). The main advantage of PMI³, compared with other corpus-based measures of correspondence (such as Altenberg1999's (Altenberg1999) mutual translatability), is that it does not give excessive scores to pairs that involve low-frequency items (RoleNadif2011). These low-frequency pairs are in fact quite numerous in the dataset at hand (some English-French pairs from our EPTIC dataset occur only one or twice in the whole Europarl corpus).

3.2.4 Complexity-related variables

We coded the length of the source concatenations in terms of the number of constituents they contain, distinguishing between concatenations made up of two words and those made up of three or more words. To account for potentially complex (and hence cognitively demanding) co-text, we also coded whether the noun concatenation under scrutiny was embedded in a larger noun phrase, as in *the carbon footprint of Brazilian beef* or *part of our contribution to global food security*. The main reason for including these two complexity-related variables is that Halverson2017's (Halverson2017) model in its current form, being primarily aimed at translated text, does not cater for some of the cognitive constraints inherent in online tasks such as simultaneous interpreting (e.g. time constraints, memory load). In the case of concatenated nouns, we expect long noun sequences and sequences embedded in larger noun phrases to be responsible for increases in cognitive load in interpreting (cf. DefrancqPlevoets2018), and hence to be potential triggers for non-equivalent renditions.

3.2.5 Lexicalization-related variables

Finally, in line with LeferDeClerck2021's (LeferDeClerck2021) observations, and because we examined concatenated nouns irrespective of their status as syntactic constructions (noun phrases) or lexical units (nominal compounds), we also coded lexicalization. Practically speaking, it was operationalized as attestedness in the Oxford English Dictionary (OED) and in IATE, whether as a main entry or subentry (Hilpert2019). For each lexicographic variable, we distinguished between lexicalized concatenations (e.g. *energy efficiency*, *food chain*, *free trade zone*, *road map*, listed in OED and IATE), partially lexicalized concatenations (e.g. *excessive price volatility*, with *price volatility* listed in IATE) and ad hoc concatenations (e.g. *transportation corridor*, *pork product*, which are not recorded in these two resources).

Table 2 provides an overview of the explanatory variables used in the present study.

3.3 Statistical testing

Preliminary tests on the frequency variables discussed in §3.2, which are all numerical, showed that their distribution was skewed. They were therefore log-transformed. We measured the simultaneous effect of the explanatory variables on our response variable, namely the use of a semantically equivalent vs non-equivalent rendition, by means of a generalized linear mixed-effects model (glmm),

Table 2: Overview of the explanatory variables used in the study

	Variable	Description
modality	modality	Simultaneous interpreting vs written translation
gravitational pull	freq_concat	Relative frequency of the source noun concatenation in English EP discourse (per million words)
	freq_constit	Average relative frequency of the individual constituents of the source noun concatenation in English EP discourse (per million words); calculated by adding up the lempos frequencies of constituents and dividing the sum by the total number of constituents in the concatenation
magnetism	freq_rendition	Relative frequency of the rendition of the source noun concatenation in French EP discourse (per million words)
connectivity	connect_PMI	Pointwise mutual information (PMI ³) of the source noun concatenation and its rendition in English-French EP discourse (per million words)
	connect_IATE	Cross-linguistic link between the source noun concatenation and its rendition as recorded in an entry or subentry of the Interactive Terminology for Europe (IATE) database: yes, partial, no
	connect_bil_dic	Cross-linguistic link between the source noun concatenation and its rendition as recorded in an entry or subentry in the online Oxford English-French bilingual dictionary: yes, partial, no
complexity	source_length	Length of the source noun concatenation, measured as the number of words it contains: 2 words vs 3 or more words
	NP_embedding	Embeddedness of the source noun concatenation in a larger noun phrase (whether as head or as postmodifier): yes, no
lexicalization	lex_OED	Attestedness of the source noun concatenation in the online Oxford English Dictionary: yes, partial, no
	lex_IATE	Attestedness of the source noun concatenation in the Interactive Terminology for Europe (IATE) database (irrespective of a potential cross-linguistic link with a French equivalent): yes, partial, no

using RStudio 1.1.383 (RCoreTeam2018). The regression model we used makes it possible to determine whether the test variables have a statistically significant effect on the response variable, what the effect of each variable is and what the overall performance of the model is in terms of descriptive and predictive adequacy.

4 Results and discussion

Figure 1 shows that semantically non-equivalent renditions account for 26% of the EPTIC dataset ($n=224/853$), while the remaining 74% are equivalent renditions ($n=629/853$). As shown in Figure 2, however, the distribution of the two types of rendition is markedly different in the two mediation modes: while translators produce semantically equivalent renditions in an overwhelming 96% of cases ($n=407/422$), the proportion drops to a mere 52% in simultaneous interpreting ($n=222/431$). This intermodal difference is statistically significant ($\chi^2(1)=220.04$, $p < 2.2e-16$). This figure is very similar to the proportion of successful renditions reported in Ghiselli2018's (Ghiselli2018) analysis of Italian interpretations of English complex noun phrases (55%) and provides additional evidence that English concatenated nouns are vulnerable in simultaneous interpreting into Romance languages (Gile1995), leading as they do to substantial numbers of incomplete and wrong renditions.

Figure 1: General distribution of semantically equivalent vs non-equivalent renditions of English concatenated nouns ($n=853$).

Figure 2: Association of modality (interpreting vs translation) and use of semantically equivalent vs non-equivalent renditions ($n=853$).

In the remainder of this section, we take a multifactorial approach to the EPTIC dataset at hand with a view to assessing how the variables under scrutiny simultaneously condition the use of semantically equivalent renditions (vs non-equivalent renditions) in the two mediation modes. We ran a glmm-model, using (non-)equivalent rendition as response variable. Modality and the ten above-mentioned gravitational-pull-, magnetism-, connectivity-, complexity- and lexicalization-related variables were used as fixed effects, with speech/text id as random effect to accommodate variation across individual speeches. We adopted a stepwise procedure, starting from a null model containing only the random intercepts and then incrementally adding fixed effects which significantly reduced the Akaike Information Criterion (AIC) value of the model. Next to the main effect

of each of the fixed factors, we also checked whether a model with two-way interactions containing modality significantly reduced the AIC value. We avoided overfitting by adopting the rule of thumb that the number of regressors multiplied by 20 should not be higher than the least frequent level of the response variable (Harrell2015).

The significant fixed effects emerging from the glmm are shown in Figure 3 (the full model is given in Appendix A). This model, which contains two main effects and two interaction effects, outperforms an intercept-only model significantly ($\chi^2(8) = 483.1$, $p < 2.2e-16$). The marginal R^2 value is 0.69, the conditional R^2 value is 0.72 and the c-score is 0.94. These indicate that the model performs very well in explaining and predicting the variation at hand.

Figure 3: Effect plots of a generalized linear mixed-effects model with semantically equivalent vs non-equivalent rendition as response variable, connectivity_PMI, connectivity_IATE, freq_rendition, and source_length as fixed effects and speech/text id as random effect (n=853).

Three main trends emerge from Figure 3. First, we see that the probability of using a semantically equivalent rendition increases with connectivity (as measured by two variables: the corpus frequency-based PMI³ score and inclusion of a given source-target pair in an IATE bilingual entry). Chances of using a semantically equivalent rendition almost reach 100% when PMI³ scores are at their highest (strong cross-linguistic link of the source-target pair in parallel Europarl) and when the source concatenation and its rendition are recorded as cross-linguistic equivalents in IATE. This is in line with one of the basic tenets of Halverson's cognitive model that "the more established (entrenched) a link is, the more likely it will be activated and used in translation" (Halverson2017: 15) and, we should add, in interpreting too. Second, we find that the probability of using a semantically equivalent rendition decreases as magnetism increases (as measured here by the frequency of the rendition in French EP discourse). This shows that in the case of concatenated nouns, when translators and, even more, interpreters are "drawn to a target language item with high salience/frequency" (ibid. 14), they are actually drawn to renditions that are not equivalent to the source concatenated nouns (incomplete renditions or wrong renditions). Examples of such cases from interpreting include *partner country* > *pays d'origine* 'country of origin', *young university graduate* > *étudiant* 'student', *dioxin contamination* > *pollution* 'pollution', *league table* > *liste* 'list'. Third, the glmm shows that the probability of using a semantically equivalent rendition decreases when source noun concatenations are longer (three or more constituents). Importantly, the effects of the latter two

variables (magnetism and noun concatenation length) are significantly stronger in interpreting than in translation, as indicated by the two-way interactions with modality in the glmm.

In view of the above observations, we can say that our initial predictions are only partly borne out. Only two of the three cognitive forces from Halverson2017's (Halverson2017) model are found to shape translators' and interpreters' use of semantically equivalent vs non-equivalent renditions, i.e. connectivity (i.e. cross-linguistic link strength, in two of its guises, one based on parallel corpus frequencies, the other on terminographic data) and magnetism (i.e. target language salience, here operationalized as frequency of use in EP discourse). While strong connectivity goes hand in hand with the use of semantically equivalent renditions, strong magnetism pulls in the opposite direction, as it draws translators, and more strikingly, interpreters, to the use of non-equivalent renditions (be they incomplete or wrong). Gravitational pull (i.e. source language salience), contrary to our expectations, does not seem to impact on the use of semantically equivalent vs non-equivalent renditions in our data. In addition, the results of the regression analysis confirm the crucial role played by the length of source concatenated nouns, a complexity-related variable. This shows that source-language-related variables have an effect on content transfer (or lack thereof) in the target language, though not an effect that is directly related to frequency or salience in the source language. Overall, we find that the same factors condition the use of semantically equivalent vs non-equivalent renditions similarly in simultaneous interpreting and written translation, with two predictors (magnetism and source concatenation length) having a significantly stronger effect in interpreting. This is in line with our initial expectation that the same factors condition renditions across mediation modes, but that their effect should be more visible in interpreting, given its very specific constraints (time, memory load, etc.). Note, finally, that lexicalization in the source language does not appear to be a driving force here.

5 Conclusion

This chapter represents the first attempt at addressing the full complexity of Halverson2017's (Halverson2017) gravitational pull model to account for the rendition of a linguistic phenomenon above the word level, concatenated nouns, across two modes of interlingual mediation, simultaneous interpreting and written translation, using robust multifactorial statistics that make it possible to assess simultaneously the effect of the three cognitive forces in play. A key finding

emerging from the regression analysis in that regard is that both connectivity and magnetism exert a strong influence on translators' and interpreters' use of semantically equivalent renditions of English concatenated nouns. While highly entrenched cross-linguistic links draw translators and interpreters alike to semantically equivalent renditions, the opposite force is observed in the case of magnetism (target language salience), with strong magnetism leading translators and, more particularly, interpreters to the use of semantically non-equivalent renditions, such as incomplete renditions and wrong renditions. We have proposed an elaborate research design to operationalize gravitational pull, magnetism and connectivity in both translation and interpreting, relying on a large reference corpus and on the pointwise mutual information score to derive cognitively motivated corpus frequency variables. We have also complemented the predictors inspired by Halverson's gravitational pull hypothesis with complexity- and lexicalization-related predictors, so as to better account for the specific features of the linguistic phenomenon at hand, concatenated nouns, and, more generally, interpreting.

One striking result of the present investigation is the lack of a source-language-induced pull effect, which raises the following question: is the rendition of more complex linguistic structures such as noun concatenations not affected by such a mechanism (an outcome which should be interpreted along cognitive-linguistic lines) or do the research topic and research design we have adopted simply prevent a pull effect from emerging (a methodological reason)? It must be acknowledged that our research topic and our research design are both quite different from those adopted in previous GPH research (e.g. Halverson2017, Marco2021), and that this may have impacted the results we obtained. The linguistic phenomenon under scrutiny here has no direct formal equivalent in the target language: as mentioned above, English noun concatenations are right-headed, whereas their French equivalents are left-headed, and very often also require the insertion of a preposition or the transposition of the modifier noun into an adjective. This makes a *formal pull effect*, such as would cause the structure of the source construction to shine through in the target text, highly unlikely (at least in professional translation and interpreting). In addition, even if translators and interpreters occasionally rendered this English structure in an (ungrammatical) word-for-word fashion, our current research design would not capture it, since the central variable in the study focuses on semantic equivalence, and not on the formal features of the renditions. With the benefit of hindsight, these two aspects may explain why a gravitational pull effect was not very likely to occur in the present study.

Admittedly, gravitational pull effects do not occur only through specific (unconventional or ungrammatical) constructions that are formally similar to source-text constructions, but can also emerge at a more aggregate level, namely when a certain linguistic phenomenon in translated language is over- or underused in comparison with non-translated language as a result of a higher or lower frequency of the equivalent representation in the source language (Halverson2017, for instance, has studied this type of gravitational pull at a semantic level). Coming back to the present study, it is possible, for example, that noun concatenations in translated or interpreted French are used significantly more often in comparison with original French under the influence of the high frequency of noun concatenations in the English source language. Once again, however, our research design does not make it possible to detect that type of pull effect, since we adopted a *parallel-corpus* design (not a *comparable-corpus* design, as in previous studies) in which the translation of *individual* source-language items was analyzed (and not just aggregate patterns of over- and underrepresentation).

In other words, the question remains whether the topic and research design we adopted in this study were suited to picking up on gravitational pull effects. One could argue, of course, that highly salient noun concatenations can affect the semantically equivalent rendering in French positively, but it is hard to think of such an effect that works independently of a connectivity or a magnetism effect: for translators and interpreters, highly salient noun concatenations in the source language will unavoidably also have a high connectivity effect, i.e. the more frequent a construction in the source language, the likelier a translator or interpreter is to have encountered this construction before and hence the likelier she is to have a routinized translation solution at her disposal.

These considerations raise important questions as to how the theoretical model developed by Halverson can be tested in a variety of empirical research designs. It is important to stress, however, that the gravitational pull model, which aims to be a comprehensive cognitive-linguistic model that can be used to explain and predict translational choices, should not be restricted to studies on over- and underuse of particular linguistic phenomena based on comparable corpora (even though the model originated from that type of research), but that it can also be used to account for local translation choices, above the word level, such as the ones studied in this chapter (Halverson, personal communication).

Although we believe that the present study has gone some way towards showing how the gravitational pull model can be tested empirically in all its complexity, thereby paving the way for further elaboration of the model, it can be complemented in several ways. First, the operationalizations of the three cognitive forces included in the gravitational pull model can be refined. For gravitational

pull (source language salience), another variable worth considering is the productivity of nouns in the semantic relations in which they are frequently used, whether as heads or as modifiers (KrottEtAl2009; Fernandez-Dominguez2020). In our dataset, for instance, we noticed that some nouns are particularly productive in EP concatenated nouns, either as heads (e.g. *cattle products*, *construction products*, *pork products*, *timber products*) or premodifiers (e.g. *trade agreement*, *trade benefit*, *trade flow*, *trade partner*). Corpus-derived operationalizations of magnetism (target language salience) could be refined along the same lines, also taking into consideration the magnetism exerted by competing equivalents in the target language. Connectivity (cross-linguistic link strength) also deserves closer attention. In particular, variables indexing the connectivity of individual constituents also need to be taken into consideration, ideally distinguishing between senses for polysemous nouns (e.g. *plant* in *plant species* vs *tuna processing plant*; see SchaeferBell2020) and between cognate vs non-cognate equivalents (ShlesingerMalkiel2005).

The length of source concatenated nouns (taken as a proxy for complexity) also emerges as a driving force behind translators' and – more crucially – interpreters' renditions. Care should therefore be taken to examine other length-related variables (e.g. constituent length, in characters) together with variables related to the temporal and cognitive constraints inherent in simultaneous interpreting, relying on cognitive-load-related parameters often investigated in CIS. These include, at the level of the speech, delivery rate, use of numbers, lexical density, syntactic complexity and formulaicity (see e.g. PlevoetsDefrancq2018cog). From the perspective of the constrained-language framework (Kotze2020), it would also be interesting to consider the native vs non-native status of the speakers ('proficiency' constraint dimension), together with directionality (translation/interpreting into the native vs non-native language; 'language activation' constraint). Likewise, the sociocultural and technological factors that typically constrain written translation should, ideally, also be taken on board to better account for intermodal commonalities and differences. Finally, it should be borne in mind that corpus data ultimately need to be complemented with other data types, such as elicitation data, as "corpus data gives us only indirect evidence of cognitive linguistic structure" (Halverson2017).

We hope that the intermodal research design proposed in this study, together with the avenues for future work we have outlined above, will lead to more systematic and more refined explorations of the gravitational pull model in empirical translation and interpreting studies and, in the longer run, to a better understanding of the commonalities and differences that typify mediated language varieties.

Acknowledgements

We wish to thank the two anonymous reviewers for their insightful feedback on an earlier version of the chapter. Their comments and suggestions helped us to broaden the scope of our initial analyses and improve our research design in many ways. We also wish to thank Sandra Halverson for fascinating discussions on the GPH and how it can be applied to the parallel corpus study of concatenated nouns. Her insights proved an invaluable help in better interpreting the findings of the present study and their implications for the GPH. Any remaining shortcomings are of course our own. Thanks are also due to the EPTIC project directors at the University of Bologna for their coordination and constant support, and to the UCLouvain students who contributed to the collection, transcription and alignment of the EPTIC subcorpus used here: Salomé Debève, Athina Danhier, Marie De Clerck, Émilie Degueldre, Tiffany Scohy, Sophie Steil, Fiona Thewissen, Florent Thirion, Antoine Van Gompel and Coraline Zizi. We also thank Salomé Debève for extracting the corpus data and coding part of it.

Appendix A Generalized linear mixed model

A.1 Random effects

Groups	Name	Variance	Std. Dev.
text_id	(Intercept)	0.4187	0.6471

Number of obs: 853, groups: text_id, 89

A.2 Fixed effects

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.862675	0.260998	3.305	0.000949 ***
modalitytranslation	3.069382	0.539696	5.687	1.29e-08 ***
connectivity_PMI	0.108200	0.016236	6.664	2.66e-11 ***
log(freq_rendition+1e-06)	-0.173665	0.036375	-4.774	1.80e-06 ***
connect_IATEP	0.009811	0.341616	0.029	0.977089
connect_IATEY	3.290843	0.642208	5.124	2.99e-07 ***
modalitytranslation:				
log(freq_rendition+1e-06)	-0.249602	0.101689	-2.455	0.014106 *
modalityinterpreting:				
source_length3.or.more.words	0.283585	0.349198	0.812	0.416731
modalitytranslation:				
source_length3.or.more.words	-1.405105	0.693004	-2.028	0.042606 *