Is linguistic decision-making constrained by the same cognitive factors in student and in professional translation?

Evidence from subject placement in French-to-Dutch news translation[*]

Gert De Sutter, Marie-Aude Lefer and Bram Vanroy

Ghent University / Université catholique de Louvain / Ghent University

This article analyses the extent to which four well-known general cognitive constraints – syntactic priming, cognitive routinisation, markedness of coding and structural integration – impact the linguistic output of translation students and professional translators similarly. It takes subject placement variation in Dutch as a test case to gauge the effect of the four constraints, and relies on a controlled corpus of student and professional French-to-Dutch L1 news translations, from which all declarative main clauses with either a preverbal or a postverbal subject were extracted. All corpus instances were annotated for four random variables, the fixed variable *expertise* and ten other fixed variables, which were considered good proxies for the cognitive constraints. A mixed-effects regression analysis reveals that by and large the cognitive constraints have an identical effect on student and professional translators' output, with priming and structural integration having the strongest impact on subject placement. However, students diverge from professionals when translating French clauses with a left-dislocated adjunct into Dutch, which is interpreted as an indication of a difference in automatisation when dealing with specific French-Dutch cross-linguistic differences.

[*] This research was carried out within the INTERACT network, funded by FWO – Research Foundation Flanders (grant number W002220N). We are very grateful to Emmelie Vermeersch for helping us code part of the data, and to Jóhanna Barðdal and Robert Hartsuiker for helpful suggestions with respect to the four cognitive constraints under investigation. We also wish to express our gratitude to the three reviewers and Sylviane Granger for their critical and constructive feedback, which helped us improve the paper greatly.

Keywords: translation expertise; cognitive routinisation; priming; markedness of coding; structural integration

1. Introduction

In usage-based linguistics, it is commonly assumed that language use involves a decision-making process whereby language users choose between a wide variety of linguistic means in order to encode their intended message. Language users choose, for instance, between (near-)synonymous lexemes to denote a specific concept (e.g. *to start* vs *to begin*) or between interchangeable syntactic structures (e.g. active vs passive) (Barlow and Kemmer, 2000; Bybee and Hopper, 2001; Diessel, 2017). Within this framework, linguists agree that these choices are motivated by several (partly automatised) cognitive and social constraints, such as priming (Bock, 1986; Pickering and Ferreira, 2008) and prestige (Labov, 1965). The present study aims to find out how four well-known domain-general cognitive constraints, viz. priming, markedness of coding, cognitive routinisation and structural integration, influence translational behaviour, and to what extent these constraints operate similarly in translation students as compared with professional translators, while keeping social parameters as stable as possible.

Previous corpus research on the role of expertise in translation has focused mainly on the distributional patterns of shallow lexico-grammatical, lexical and syntactic features in corpora containing student and expert (professional) translations. This body of research shows that linguistic characteristics in student translations deviate at least to some extent from those produced by professionals, for instance in terms of lexical diversity, lexical

density and clause length (De Sutter, Cappelle, De Clercq, Loock & Plevoets, 2017; Kunilovskaya, Morgoun & Pariy, 2018; Lapshinova-Koltunski, Popović & Koponen, 2022; Lapshinova-Koltunski, 2022), complexity (Corpas Pastor, Mitkov, Afzal & Pekar, 2008), conventionality (Redelinghuys and Kruger, 2015; Kunilovskaya and Lapshinova-Koltunski, 2022), and connector use (Kajzer-Wietrzny, 2019). In view of these differences between students and professionals, it might be hypothesised that the cognitive constraints underlying translational behaviour in younger and less experienced translators (here, students) may be significantly different from those underlying that of experienced professional translators. Although it is still unclear how translators' cognitive architecture changes with increasing expertise, previous research in cognitive science has shown relevant differences in the functioning of the bilingual brain in proficient vs less proficient bilinguals. Findings to date suggest that with repeated practice of specific cross-language tasks, there are changes in the nature and operation of the lexico-semantic system and other cognitive structures, such as the inhibitory control mechanism that suppresses L2 influence when producing in L1 (Diamond and Shreve, 2010: 291)[1].

In order to test the relationship between expertise and the functioning of cognitive constraints, a corpus-based approach has been adopted, using a controlled self-compiled corpus. We examine a case of variation in Dutch, viz. preverbal (non-inverted) vs postverbal (inverted) subject placement in declarative main clauses, in order to tease apart the four cognitive constraints under scrutiny and to ascertain the extent to which they operate similarly in student vs expert translations.

In Section 2, we discuss subject placement variation in Dutch, focusing both on the general characteristics of the variation and on the factors that shape it. We also show how subject placement variation can be related to expertise and the four cognitive constraints

[1] Diamond and Shreve (2010) provide an excellent overview of research in cognitive bilingualism studies, while explicitly linking the findings to hypotheses for translation and interpreting studies.

under scrutiny. Section 3 is devoted to corpus data and methodology. We describe the corpus compiled for the study and the way it was processed linguistically. We also present data selection, hypotheses and the operationalisation of the four cognitive constraints, alongside the statistical approach adopted. Sections 4 and 5 present and discuss the results of our analysis, and Section 6 concludes the article with a look towards future research.

2. Subject placement variation in Dutch: general characteristics, determinants, constraints and expertise

Like German, Dutch is regarded as a mixed SVO, SOV and VSO language. In declarative main clauses, finite verbs are placed in the second syntactic slot and are preceded by a single constituent (often the subject of the clause, hence SVO). Non-finite verbs, if any, cluster together at the end or towards the end of the clause. In most subordinate clauses, both finite and non-finite verbs occur at the end or towards the end of the clause (SOV), and in interrogative and exclamative clauses, the finite verb precedes all other constituents (VSO; see Haeseryn, Romijn, Geerts, Rooij & van den Toorn, 2019; Zwart, 2011).

Although Dutch subjects often precede the finite verb in typical SVO contexts such as declarative main clauses (Broekhuis (2020) refers to this as the unmarked order, Vogels and Van Bergen (2017) call it the Subject First Preference; cf. Example 1), subjects can easily be inverted, thus following the finite verb and leaving the first syntactic slot either empty (Example 2), which is typically the case in interrogative, exclamative and

adhortative clauses, or to be taken by another constituent (e.g. a temporal adverbial, as in Example 3). In the examples, subjects are underlined and finite verbs are in bold.[2]

(1) Wij **hebben** "ok" geantwoord. (P3-11-24)
We have "ok" answered.
'We answered "ok".'

(2) **Is** de Alliantie tegen de corrida erin geslaagd haar stem te laten horen? (P4-12-7)
Is the alliance against the corrida succeeded her voice to let hear?
'Did the alliance against the corrida succeed in making its voice heard?'

(3) In enkele jaren tijd **is** haar leven een levensechte psychologische thriller geworden. (S50-11-10)
In a couple of years' time is her life a real-life psychological thriller become.
'In a couple of years her life became a real-life psychological thriller.'

An important determinant of constituent ordering in general (cf. Arnold, Kaiser, Kahn & Kim, 2013) and subject placement in Dutch in particular is information structure. The Dutch reference grammar *Algemene Nederlandse Spraakkunst* (Haeseryn et al., 2019: 21.3.1.1, 21.1.2.1), for instance, argues that a constituent's position is determined not only by its syntactic function, but also by its discourse function. Dutch clauses start out with constituents representing 'given', 'accessible' or context information (e.g. temporal or local adverbials; see Example 3), followed by new, unintroduced, inaccessible, more prominent information in the right-hand clausal context (there are some exceptions, however, such as contrastive clauses and news clauses, which start with new, prominent

[2] All examples in this article are taken from the corpus that was used for the study (see Section 2). After each example, translator status (P = professional, S = student), translator-id, text-id and segment-id are given (in that order).

information; cf. Haeseryn et al., 2019: 21.3.1.3). As a consequence, any constituent which optimally represents 'given' information in a certain discourse context can appear in the first syntactic slot, preceding the finite verb, without needing to be the grammatical subject. As such, in deciding on the relative order of the subject and other syntagmatically variable informational elements in the clause, language users mark the information structure of the clause, broadly dividing the clause into given and new information and thereby signalling "what the speaker is attending to, what the speaker wishes the addressee(s) to focus on, what is assumed to be already known, what is considered most important, or what is treated as background information" (Arnold et al., 2013: 2; cf. also Diessel, 2019: 26, who refers to this as a strategy to establish cognitive common ground between speaker/writer and hearer/reader).

In a multivariate corpus study on spoken Dutch plural nominal subjects in main clauses, Vogels and Van Bergen (2017) investigate the relationship between what they call the Subject First Preference (preverbal subjects are the unmarked option) and information structure. More particularly, they fit a mixed-effects model with subject placement as response variable, and information structure (operationalised as definiteness of the subject), animacy and length of the subject, tense, aspect and voice of the verb, and type and length of the prepositional phrase (PP) as fixed predictor variables. Their results show that preverbal and postverbal subjects are equally likely to occur in spoken Dutch (contrary to the expectation that there is a preference for subjects to occur preverbally). In addition, the probability of subjects occurring in preverbal (vs postverbal) position is significantly greater when the subject is definite (i.e. presents 'given' information) than when it is indefinite. Animacy only plays a role in an interaction effect with definiteness: in clauses with definite subjects, animacy does not affect subject placement significantly. However, it affects the position of indefinite subjects: indefinite animate subjects are

more likely to occur preverbally than indefinite inanimate subjects. These results show the importance of information structure (definiteness) for subject placement in the first place, and that animacy can tip the balance in favour of preverbal or postverbal position when subjects are indefinite (Vogels and Van Bergen, 2017: 391). Finally, the results also show that the length of the subject is negatively associated with preverbal position (the longer the subject, the less likely it is to occur in preverbal position), and that spatio-temporal PPs (vs abstract PPs) increase the likelihood of subjects in postverbal position. The verb-related predictors TENSE, ASPECT and VOICE did not yield any significant effects. Similar results about the effect of grammatical function, definiteness and length are obtained in Bouma (2008), who uses similar spoken corpus data from Dutch.

Grondelaers and Speelman (2007) investigate the relative order of locative adjuncts and postverbal subjects in Dutch presentative clauses introduced by *er* ('there'). They also show that a discursive mechanism, referred to as 'informational prominence', plays a major role: "the constituent which makes the greatest informational contribution to the clause – be that the subject or the adjunct – tends to be sentence-final, whereas less prominent materials are 'backgrounded' to penultimate position" (ibid, 182). Jansen and Wijnands (2004), using a small ad hoc corpus of journalistic texts, observe that the subject in main clauses most often coincides with given information, independently of its position in the clause. The strong association between the subject in Dutch and discourse givenness is also confirmed by a diachronic multi-register corpus study by Coussé (2009), who investigates the incidence of postverbal subjects and objects in 13th to 18th century Dutch. Although her study does not focus on the alternation between preverbal and postverbal subjects in main clauses as such, it can be deduced from her results that the variation already existed in these older sources, and that subjects were placed in postverbal position if they were discursively prominent (ibid, 11-14).

The conclusion that information structure and animacy play a major role in subject placement in Dutch is hardly surprising, as these determinants have also figured prominently in understanding other cases of syntactic variation in various languages (e.g. Hawkins, 1994; Rosenbach, 2005; Wasow, 2002). Using insights from cognitive linguistics, effects of information structure and animacy can be linked to an underlying so-called domain-general cognitive constraint, viz. markedness of coding. Domain-general constraints are general principles that govern all aspects of cognitive processing, not just language production or comprehension, and are formulated in accordance with what is known about the functioning of the brain from other disciplines, such as cognitive psychology and neuroscience (this is the so-called 'cognitive commitment' in cognitive linguistics; Lakoff, 1990).

The markedness-of-coding constraint means that a linguistic element has a higher probability of being selected if it best matches the idea (conceptualisation) a language user wants to encode. Applied to subjects, this means that linguistic elements that are conceptually highly prominent, i.e. are most likely to attract the speaker's attention, are prime candidates to occupy the subject position, thus leading to an optimal correspondence between conceptualisation (prominent referent) and form (grammatical function) (Claes, 2017: 3-4). Since cognitive research has shown that animates are perceptually more salient than inanimates (Calvillo and Jackson, 2014), and definite referents discursively more salient than indefinites (Falk, 2014), animate and definite referents have a higher likelihood of being selected for the subject position. Additionally, given the unmarked SVO word order in Dutch declarative main clauses (Bouma, 2008; Broekhuis, 2020), we can add that this optimal correspondence between conceptualisation and grammatical function is best realised when subjects are placed in preverbal position, thus adding a syntactic component to the definition of this constraint. As a corollary, the

effect of subject length can also be indirectly related to this cognitive constraint: given the general preference in Dutch for placing shorter constituents before longer ones – the so-called 'Principle of End Weight' (Wasow, 2002; Haeseryn et al., 2019: 21.1.3) – short constituents are (syntactically) more suited to functioning as subject in Dutch than long constituents.

Next to markedness of coding, at least three other domain-general cognitive constraints are potentially relevant in explaining subject placement in Dutch. These constraints have received a great deal of attention in variational-linguistic research on grammatical variability in many languages, but not yet with respect to subject placement in Dutch. These constraints are structural priming, cognitive routinisation and structural integration.

Structural priming is undoubtedly the best-known constraint, as it figures prominently in psycholinguistics (Bock, 1986) and usage-based linguistics (e.g. Gries, 2005; Szmrecsanyi, 2003; De Sutter, Colleman and Ghyselen, 2021. It is regarded as a residual activation effect, which increases the likelihood of a given morphosyntactic phenomenon (e.g. a specific word order configuration) that was activated recently (either in production or perception) being recycled (see Gries and Kootstra, 2017)[3]. With respect to subject placement in Dutch, this would mean that structural priming predicts that a preverbal subject will be selected if a preverbal subject was activated in preceding clauses (and likewise, *mutatis mutandis*, in the case of postverbal subjects). Priming effects have been attested between languages as well as within one language (Hartsuiker, Beerts, Loncke, Desmet and Bernolet, 2016; Maier, Pickering and Hartsuiker, 2017; Pickering and Ferreira, 2008), which has important implications for empirical translation studies: in a translated text, subject placement can be influenced by a prime in the source text

[3] An alternative account considers priming to be an implicit learning effect (see Khoe, Tsoukala, Kootstra & Frank, 2021).

(intertextual, cross-linguistic priming) as well as by a prime in the preceding context of the translated text itself (intratextual priming; cf. Szmrecsanyi, 2005; Gries, 2005)[4]. In Example (4b), the position of the subject in the translated Dutch text may be influenced by the position of the subject in the French source text (4a), i.e. intertextual priming from source to target text. In Example (5c), by contrast, the position of the subject is more likely to be influenced by the position of the subject in the preceding clause in the translated text (5b) than by the position of the subject in the source text (5a), i.e. intratextual priming.

(4a) Mais Banksy n'**avait** peut-être pas imaginé […]. [source text, preverbal subject]

(4b) Maar Banksy **had** zich wellicht niet kunnen inbeelden […]. (P1-1-21) [target text, preverbal subject]

But Banksy had probably not can imagine […].

'But Banksy could probably not have imagined […].'

(5a) Alors que la campagne électorale bat son plein, on **peut** d'ailleurs s'interroger sur les véritables motivations du plaignant. [source text, preverbal subject]

(5b) In deze dichtbevolkte badplaats met heel wat gepensioneerden met bescheiden middelen, **is** hij kandidaat om herverkozen te worden bij tussentijdse verkiezingen [target text, sentence preceding (5c), postverbal subject]

(5c) Terwijl de verkiezingscampagne in volle gang is, **kunnen** we ons trouwens vragen stellen bij de echte redenen van de klager. [target text, postverbal subject]

[4] One of the reviewers suggested that it is not very likely to see an effect of intratextual priming in news reports, the genre investigated in this study, given the often explicit stylistic recommendations in journalistic contexts to vary sentence structure regularly. Although this might be the case, it should not prevent us from at least controlling for a potential effect of intratextual priming in this study, since it has been shown to be a relevant factor in other types of morphosyntactic variation.

While the election campaign is in full swing, may we indeed questions ask about the real reasons of the plaintiff.

'While the election campaign is in full swing, we may indeed wonder about the plaintiff's real reasons.'

A third constraint, cognitive routinisation or automatisation, refers to the strengthening of linguistic elements in memory through repetition (Diessel, 2019: 34-35; see also Bybee, 2007). This is what is often referred to as entrenchment in cognitive linguistics: morphemes, lexemes, morphosyntactic structures, etc. that are produced and perceived more frequently than others are better entrenched in memory, giving them an advantage over competing elements to be selected for production or comprehension. Repetition (frequency) leads to routinisation, which in turn leads to ease of production and comprehension. The articulatory gestures needed to produce common speech sounds or the rapid comprehension of frequent lexemes, which often happen within milliseconds, are two cases in point. Routinisation also applies to frequent combinations of linguistic elements (e.g. *sort of*, *by the way*), which are assumed to be stored as units in cognitive organisation (Bybee, 2007: 324). As a consequence, linguistic elements that frequently co-occur have a higher probability of being selected for encoding than (more schematic) linguistic elements that can be syntagmatically assembled in a clause but do not co-occur as frequently. This also applies to specific linguistic elements that frequently occur in certain schematic positions of a construction (Schmid, 2015), suggesting that linguistic elements that are used as a subject more often (compared to their use as non-subject) will have a higher probability of being selected for the subject position in subsequent language production than linguistic elements that are less frequently used as subject.

The final constraint is structural integration, which is one of the crucial aspects of Gibson's (2000) so-called 'dependence locality theory' in cognitive science and psycholinguistics (cf. also Hawkins, 2001). This constraint is based on the general assumption that words that are syntactically connected (e.g. a head and its modifier) tend to be placed in adjacent positions. Building on the results of empirical work in corpus linguistics and psycholinguistics, structural integration becomes cognitively costlier if two words that are syntactically connected are separated by other words, and the size of the processing cost increases with each intervening word.[5]

Although there is a large body of empirical research in linguistics investigating the effect of one or more of these constraints (in some form) on morphosyntactic variational phenomena, empirical studies have not yet examined whether these effects are contingent on expertise (cf. Tiselius and Hild, 2017, for an overview of research into translation expertise or competence). One could easily imagine, however, that the effect of cognitive routinisation in students' language production is weaker compared with professionals, simply because automatisation in students has not progressed as far as in professionals (cf. also Diamond and Shreve, 2010). This is especially true of cross-linguistic cognitive routines, i.e. the automatisation of certain translation solutions given a particular source-language stimulus. On the other hand, in a translational context, the effect of structural priming from the source text may be stronger in students' output since students might not have developed (automatised) meta-linguistic strategies to assess the suitability of a similar structure in the target text (cf. Castagnoli, 2016, for a case of information-structural transfer or priming in student translations from L2 English into L1 Italian).

The central research question in this article is thus how the four cognitive constraints affect subject placement in L1 Dutch translations from L2 French, and to what

[5] It must be acknowledged that Gibson's theory is far more specific than laid out here, but we will rely on this broad characterisation in the present study.

extent these constraints operate differently in translation students compared with professional translators.

3. Data and methodology

3.1 Corpus design and processing

A new corpus was compiled for this study: eight professional translators who were all members of the Belgian Chamber of Translators and Interpreters (CBTI-BKVT) were commissioned in 2021 to each translate eight journalistic source texts in French (varying between 313 and 599 words) into (Belgian) Dutch. All the translators were Belgian, grew up and lived in (Dutch-speaking) Belgium and were native speakers of (Belgian) Dutch, with French as one of their main working languages. Seven translators had obtained a master's degree in translation, and one translator had a master's degree in modern languages. The average amount of working experience as a translator was 13.1 years (min. 5, max. 22). Most of the translators were female (6 out of 8). The translators were instructed to produce a translation with an informative goal, directed towards a general, non-specialised audience in Dutch-speaking Belgium and the Netherlands. The translators were not allowed to use machine translation or translation memories, or to ask another translator to revise their work. They were free to use other (digital and non-digital) resources (e.g. dictionaries, glossaries, terminological databases). They were asked to report on the time they took to complete each translation. They spent on average 54.6 minutes per text, with large inter-translator differences (min. 14 minutes, max. 120 minutes). All the professional translators were paid their usual fees.

The same source texts were also translated by 59 students, who were undergraduates at Ghent University[6]. They all grew up and lived in (Dutch-speaking) Belgium and were native speakers of (Belgian) Dutch. There were 9 male and 50 female students, none of them with working experience in the translation industry. The students produced the translations as part of a translation course in 2017, and received no feedback before submitting their translations. Unlike the professional translators, who were all commissioned to translate the eight source texts, the students translated only one or two source texts into Dutch. They were given the same translation brief as the professional translators.

All the translations were automatically sentence-aligned with their respective source texts, with systematic post hoc correction (using Alignfactory Light[7]). The texts were also POS-tagged and lemmatised using Stanza (Qi, Zhang, Zhang, Bolton and Manning, 2020).

3.2 Data extraction and annotation

We extracted all Dutch declarative main clauses with a nominal or pronominal grammatical subject. Subordinate clauses, second clauses of coordinated clauses, titles and subtitles were left out. This yielded 2,282 relevant instances: 63.9% of these instances contained a preverbal subject in Dutch (n = 1,459) and 26.1% contained a postverbal subject in Dutch (n = 823).

Each instance was manually annotated for the response variable SUBJECT PLACEMENT (preverbal vs postverbal) and a series of random and fixed predictor variables. We regard subject placement in declarative main clauses in Dutch as an

[6] Part of the corpus was previously used in De Sutter et al.'s (2017) study on translation quality.
[7] https://terminotix.com/

alternation, in the sense that language users must choose between placing the subject preverbally or postverbally, and this choice is likely to be constrained by a number of factors (cf. Vogels and Van Bergen, 2017 for a similar line of reasoning, and Pijpops, 2020 for a more general overview on what counts as an alternation).

The random variables include TRANSLATOR-ID, TEXT-ID, SUBJECT LEMMA and VERB LEMMA. The fixed variables were chosen with respect to expertise (professionals vs students) and the four central cognitive constraints mentioned in Section 2.

3.3 Operationalisations of the constraints (fixed variables)

This section presents the operationalisations of the four cognitive constraints under investigation (Table 1).

Table 1. Overview of the operationalisations (fixed variables) for each of the cognitive constraints

Constraint	Operationalisations (fixed variables)
Markedness of coding	1. Subject animacy 2. Subject discourse status 3. Subject length 4. Subject concreteness
Cognitive routinisation	5. Statistical attraction between (pro)nouns and the subject category 6. Statistical attraction between the verb and pre- or postverbal subjects
Structural priming	7. Intertextual priming 8. Intratextual priming 9. Indirect intertextual priming
Structural integration	10. Complexity of the VP

3.3.1 Markedness of coding

Given the general definition of this constraint, as laid out in Section 2, the following operationalisations were considered good proxies for markedness of coding: SUBJECT ANIMACY (animate referents are conceptually more salient than inanimates, giving them a higher likelihood of being selected as subject; Calvillo and Jackson, 2014), SUBJECT DISCOURSE STATUS (referents that relate to entities which were already introduced in the preceding discourse are discursively more salient than unintroduced, new entities, giving them a higher likelihood of being selected as subject; Falk, 2014) and SUBJECT LENGTH (given the general preference of Dutch for placing subjects early in the clause and given the Principle of End Weight, shorter constituents are syntactically better candidates for the subject function than longer subjects). We also add SUBJECT CONCRETENESS to this

list (concrete referents are conceptually more salient than abstract ones, giving them a higher likelihood of being selected as subject), following psycholinguistic research which found evidence that concrete words have a cognitively more profound semantic representation than abstract words, as a consequence of which they are better retained in memory and more easily learned (see Kaushanskaya and Rechtzigel, 2012, and references therein); we assume from this that concrete referents are conceptually more salient.

- SUBJECT ANIMACY – the Cornetto database[8] was used to determine the animacy of the subject's head noun. Four levels or categories were distinguished: *animate*, *inanimate*, *institution* and *no noun*. The category *institution* was included because it can be considered partly animate, partly inanimate (e.g. *The West Australian Opera*).
- SUBJECT CONCRETENESS – for determining the concreteness of the subject's head, the independent concreteness ratings list collected by Brysbaert, Stevens, De Deyne, Voorspoels & Storms (2014) was used. If the subject head was a compound, only the head of the compound was considered, and proper names were automatically assigned the highest concreteness score. Nouns that were not in this list were assigned 'NA'. Values ranged between 0 (highly abstract) and 5 (highly concrete); median value = 2.8, mean = 2.9.
- SUBJECT DISCOURSE STATUS – the discourse status of the subject's head was determined manually by verifying to what extent the subject's referent had been previously introduced in the preceding clauses. Four different categories or levels were distinguished: *new referent* (not previously introduced in the discourse), *old referent* (previously introduced in the discourse), *discourse-new referent* (new in

[8] http://www.cltl.nl/projects/previous-projects/cornetto/

the textual discourse, but retrievable from broader contextual knowledge; typically a well-known person or institution, such as the *European Parliament* or a generic indefinite pronoun, such as *one*) and *related referent* (a new referent which is clearly related to a given referent, e.g. *cigarette manufacturer* when the preceding context mentioned *cigarette packs*).

- SUBJECT LENGTH – measured as number of characters, excluding spaces and punctuation. Values ranged from 2 to 181; median value = 10.0, mean = 17.3.

3.3.2 Cognitive routinisation

Linguistic elements that occur more often as subject have a higher probability of being selected for the subject position than linguistic elements that are less frequently used as subject. Hence, (STATISTICAL) ATTRACTION BETWEEN (PRO)NOUNS AND THE SUBJECT CATEGORY (how often is a given (pro)noun used as a subject vs as a non-subject?) and STATISTICAL ATTRACTION BETWEEN THE VERB AND PRE- VS POSTVERBAL SUBJECTS (how often does a given verb coincide with a preverbal vs a postverbal subject?) seem to be reliable operationalisations of this constraint.

- ATTRACTION TO SUBJECT CATEGORY – using a subset of the *Sonar* reference corpus of Dutch (viz. the subset containing written-to-be-read and published texts; Oostdijk, Reynaert, Hoste & Schuurman, 2013), it was computed how often the subject's head in our dataset occurred as a subject in *Sonar* compared with its frequency as a non-subject. This yielded a proportion for each head. To do this, the data was automatically parsed with a state-of-the-art linguistic parser, namely spaCy's nl_udv25_dutchalpino_trf model.[9] The values range from 0 (never used

[9] https://github.com/bramvanroy/sv-order-2021

as subject in *Sonar*) to 0.99 (always used as subject in *Sonar*); median value = 0.3091, mean = 0.4085.

- VERB ATTRACTION TO PREVERBAL VS POSTVERBAL SUBJECTS – using the same *Sonar* subcorpus, a delta-p score was computed for each verb lemma, expressing the probability of observing this verb lemma with a preverbal subject minus the probability of observing it with a postverbal subject (see Ellis, 2006). Values range from -0.65 (verb is highly attracted to postverbal subjects) to 0.35 (verb is highly attracted to preverbal subjects); median value = 0.045, mean = 0.036.

3.3.3 Structural priming

Given the ubiquitous effect of structural priming in previous syntactic alternation research, it can be reasonably expected (despite the lack of direct evidence) that preverbal vs postverbal positioning of subjects in Dutch declarative main clauses is contingent on a preceding prime: a preverbal subject in preceding clauses will prime a preverbal subject in the subsequent clause; a postverbal subject will prime a postverbal subject. Priming may refer to the subject position in the source text (INTERTEXTUAL PRIMING) or the subject position in the preceding clause of the translated text itself (INTRATEXTUAL PRIMING) (for an extensive literature review of the effects of different types of structural priming, see Pickering and Ferreira, 2008).

- INTERTEXTUAL PRIMING – the position of the subject in the French source text: *preverbal* vs *postverbal*.
- INTRATEXTUAL PRIMING – the position of the subject in the preceding Dutch clause: *preverbal* vs *postverbal*.

Given a cross-linguistic difference between French and Dutch, a third type of priming will be considered as well, which we term here INDIRECT INTERTEXTUAL PRIMING: in French declarative main clauses, it is not unusual that an adjunct is left-dislocated, then followed by the subject and the finite verb (in that order; see Example 8a). In Dutch, on the other hand, a left-dislocated adjunct immediately followed by the subject is ungrammatical or – at most – very uncommon in writing (cf. Haeseryn et al., 2019: 21.8.2.5). Adjuncts occur in clause-initial position in Dutch, but not as a left-dislocation, and when they occur, they always cause inversion, i.e. forcing the subject into postverbal position (Example 8b). In Example (8), the clause-initial adjunct is italicised, subject underlined and finite verb in bold.

(8a) *Friand d'actualité*, Banksy **a** déjà peint des espions sur un parapet […].

(8b) *Verzot als hij is op de actualiteit*, **heeft** Banksy al eens spionnen geschilderd op een borstwering […]. (P5-5-18)

Fond as he is of current affairs, has Banksy previously spies painted on a parapet […].

'Fond as he is of current affairs, Banksy has previously painted spies on a parapet […].'

Hence, if the clause-initial adjunct in the French source text (*Friand d'actualité* in 8a), which is the first constituent to be translated, primes a clause-initial adjunct in Dutch (*Verzot als hij is op de actualiteit* in 8b), this automatically results in placing the subject postverbally in Dutch. Placing the subject preverbally would result in an ungrammatical clause. In such circumstances, the position of the subject is not directly primed by a preceding postverbal subject, but indirectly by the priming of the immediately preceding constituent.

3.3.4 Syntactic integration

In a clause, the subject and the verb stand in a close syntactic (and semantic) relationship to each other. As such, they will tend to be as adjacent as possible, thereby minimising processing costs. In the present study, this is operationalised indirectly by looking at the complexity of the VP in the Dutch clause:

- COMPLEXITY OF THE VP – There are two levels: *complex VP* (if the finite verb in the second position of the clause is complemented by one or more non-finite verbs at the end of the clause) and *simple VP* (if the finite verb is the only verb in the clause). In the latter case, the finite verb is the main verb in the clause and is immediately adjacent to the subject (either in preverbal or postverbal position; Example 9). In the former case, one of the non-finite verbs at the end of the clause is the main verb, as a consequence of which the subject and the main verb are not immediately adjacent (Example 10).

(9) Op zes oktober [haalt]$_{Vfin/Vmain}$ [haar foto]$_{subject}$ de voorpagina van de Daily Telegraph. (S65-162-4)

On six October reaches her photograph the front page of the Daily Telegraph.

'On the sixth of October, her photograph made it to the front page of the Daily Telegraph.'

(10) [De pakjes]$_{subject}$ [zouden]$_{Vfin}$ een olijfgroene achtergrond [hebben]$_{Vmain}$ […]. (P4-20-6)

The packs would an olive green background have […].

'The packs would have an olive green background […].'

3.4 Hypotheses

3.4.1 Expertise

In view of the general trends with respect to the effect of expertise reviewed in Sections 1 and 2 – student translations diverge linguistically from professional translations (corpus-based studies), and cognitive processing in less proficient bilinguals is different from that in proficient bilinguals (cognitive studies), we formulate the following hypothesis:

Hypothesis 1 – There is a significant interaction effect between expertise and the effect of one or more of the cognitive constraints, thus providing evidence for a different cognitive organisation in students compared with professionals. Given the lack of research in this area, more specific hypotheses cannot be formulated for the time being.

3.4.2 Markedness of coding

Hypothesis 2 – Given that animate, concrete, given and short referents are more salient than inanimate, abstract, unintroduced and long referents and therefore more likely to function as subject, it is predicted that prime candidates for subjecthood have a higher probability of occurring in a syntactically somewhat more marked position, viz. postverbal position (vs preverbal position), compared with less salient subjects (i.e. inanimate, abstract, indefinite, long referents), which are more in 'need' of the preverbal position to mark their subjecthood. Hence, it is predicted that animate, concrete, given and short subjects (vs inanimate, abstract, unintroduced, long subjects) have a higher probability of occurring in postverbal position (vs preverbal position).

3.4.3 Cognitive routinisation

Hypothesis 3 – (Pro)nouns that occur more frequently as subject (compared to their use as non-subject) have a higher probability of being selected for the subject position, and because of that status they can be more easily placed in marked positions, viz. postverbal position (vs preverbal position) (we follow the same line of argumentation here as in Hypothesis 2). Hence, the prediction is that there will be a positive association between high attraction to the subject category and the postverbal position. Moreover, verbs that co-occur more frequently with preverbal subjects than with postverbal subjects will have a higher likelihood of occurring with a subject in preverbal position (in the present corpus). The same applies to postverbal subjects.

3.4.4 Structural priming

Hypothesis 4a – Subjects in the translated text that correspond to a French source clause with a preverbal subject have a higher likelihood of occurring preverbally as well (vs postverbal position). The same applies to postverbal subjects.

Hypothesis 4b – Subjects in the translated text that are preceded by a clause with a preverbal subject in the translation have a higher likelihood of occurring preverbally as well (vs postverbal position). The same applies to postverbal subjects.

Hypothesis 4c – Clauses in the French source text that start with an adjunct increase the likelihood of subjects in the Dutch translated text occurring postverbally (vs preverbal position).

3.4.5 Structural integration

Hypothesis 5 – Subjects in a clause with a complex VP will have a higher probability of being placed postverbally (compared with subjects in clauses with a simple VP), so as to

minimise the distance between the subject and the main verb of the clause (and hence minimise structure integration processing costs).

3.5 Statistical analysis

After annotating each of the 2,282 observations in the dataset for each of the above-mentioned predictors, a multi-level generalised linear mixed-effect model was fitted (glmm, using *lme4* in RStudio; Bates, Mächler, Bolker & Walker, 2014; R Core Team, 2016) in order to find out which of the predictors significantly influence the choice between preverbal and postverbal subjects in Dutch translated texts produced by professionals and students. The numerical variable SUBJECT LENGTH was logarithmically transformed, since preliminary exploratory analyses revealed that its distribution was very skewed (R-package *e1071*; Meyer, Dimitriadou, Hornik, Weingessel & Leisch, 2015). The distributions of all other numerical variables (SUBJECT CONCRETENESS, ATTRACTION TO THE SUBJECT SLOT and VERB ATTRACTION TO PREVERBAL VS POSTVERBAL SUBJECTS) were not skewed. We started out from an intercept-only model containing only random intercepts for TRANSLATOR-ID, TEXT-ID, SUBJECT LEMMA and VERB LEMMA and a specification of the nested structure of our dataset (TEXT-ID was nested in TRANSLATOR-ID), thereby following Gries's (2015) recommendations. Fixed variables were then added one at a time, and at each step the effect of the fixed variable on the overall model quality was evaluated: if an ANOVA test revealed that the inclusion of a fixed variable significantly reduced the Aikake Information Criterion (AIC) value of the model (compared to a model without this variable), it was retained, otherwise it was removed. The same procedure was adopted to test the two-way interactions and the random slopes of each of the retained fixed variables. When assessing the effect of the random slopes, it

sometimes occurred that the model ran into convergence or singular fit errors; in that case, the random slope was removed from the final model. The maximal model consisted of the four random variables (TRANSLATOR-ID, TEXT-ID, SUBJECT LEMMA and VERB LEMMA), and the eleven fixed variables mentioned above, including their two-way interactions and their random slopes. We avoided overfitting by following 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 (cf. Harrell, 2015: 72). Additionally, potential overfitting was evaluated by computing confidence intervals based on 1,000 bootstrap repetitions.

4. Results

The resulting glmm model contains random intercepts for SUBJECT LEMMA and VERB LEMMA, and separate slopes for SUBJECT LENGTH per VERB LEMMA and per SUBJECT LEMMA, as well as five significant fixed main effects and five significant two-way interactions. This model outperforms an intercept-only model significantly ($\chi^2(21) = 600.06$, $p < 2.^{2e\text{-}16}$). The c-score of 0.99 indicates that the model fits the data very well and correctly predicts 96% of the observations, compared to a baseline of 64%. Finally, the marginal R^2 value of 0.45 and the conditional R^2 value of 0.96 indicate that both the fixed variables and the random variables have a high explanatory value. In this section, we discuss the random effects (Section 4.1) and each of the fixed effects (Section 4.2).

4.1 Random effects

As can be seen from Table 2, the random intercepts for SUBJECT LEMMA (283 types) and VERB LEMMA (389 types) indicate that the choice between a preverbal and a postverbal

subject in Dutch news translations from French is contingent on the specific subject and verb used in the clause. In other words, some subject lemmas have a preference for occurring either preverbally or postverbally, whereas some verb lemmas are more often combined with a preverbal or a postverbal subject than others. Closer inspection reveals, for instance, that the pronouns *wat* 'what'*, we* 'we' and *ze* 'they' (unstressed) in subject function have the highest probability of occurring preverbally, compared with all other subjects in our dataset, whereas *maatregel* 'measure' has the highest probability of occurring postverbally. The verbs *maken* 'make' and *horen* 'hear' have the highest probability of combining with a postverbal subject, whereas *schudden* 'shake'*, aankomen* 'arrive' and *bezingen* 'sing' are the three verbs with the highest probability of combining with preverbal subjects. By adding random intercepts to the model, we are able to account for this idiosyncratic behaviour of specific verbs and subjects. From Table 2, it can be deduced that most variance is covered by the random intercept for VERB LEMMA (36.5), followed by the random intercept for SUBJECT LEMMA (15.57). Furthermore, the results for the retained random slopes show that the effect of SUBJECT LENGTH on the choice between pre- and postverbal subjects varies considerably per VERB LEMMA and SUBJECT LEMMA (the general effect of SUBJECT LENGTH is discussed below). Finally, it should be noted that the random effects for TRANSLATOR-ID and TEXT-ID did not contribute to the model, which means that there are no major between-translator and between-text differences in subject placement in our dataset.

Table 2. Random effects in the multi-level generalised linear mixed-effects model of subject placement in French-to-Dutch news translation

	Variance	**Standard deviation**
Random intercept: VERB LEMMA	36.5	6.05

Random slope: SUBJECT LENGTH | VERB LEMMA	7.50	2.74
Random intercept: SUBJECT LEMMA	15.57	3.95
Random slope: SUBJECT LENGTH | SUBJECT LEMMA	6.05	2.45

4.2 Fixed effects

Table 3 presents the effects of the fixed predictors. The following predictors were added to the model and did not turn out to affect subject placement significantly: INTRATEXTUAL STRUCTURAL PRIMING, SUBJECT CONCRETENESS, SUBJECT ANIMACY, ATTRACTION TO THE SUBJECT SLOT and VERB ATTRACTION TO PREVERBAL VS POSTVERBAL SUBJECTS (thereby refuting hypotheses 3, 4b and – partially – 2). This means that we were able to capture subject placement variation in French-to-Dutch professional and student news translation on the basis of the six remaining predictors, either as a main effect or an interaction effect. Their effects will be discussed below (together with the non-significant effects).

When relating the significance or non-significance of the selected predictors to the four major cognitive constraints they embody, the conclusion is that all constraints but one affect subject placement in Dutch: structural priming (from the French source text: intertextual priming and indirect intertextual priming, but not from the preceding Dutch clause: intratextual priming), structural integration (COMPLEXITY OF THE VP) and markedness of coding (SUBJECT LENGTH and SUBJECT DISCOURSE STATUS). Cognitive routinisation (ATTRACTION TO THE SUBJECT SLOT, VERB ATTRACTION TO PREVERBAL VS POSTVERBAL SUBJECTS), however, has no influence on subject placement.

In view of the main goal of this study, i.e. to verify to what extent the operation of the four cognitive constraints, operationalised by ten predictors, is contingent on

expertise, it can be concluded that there is no major difference in subject placement between student and professional translators (EXPERTISE does not show up as a main effect in the model), although there is a significant interaction effect between EXPERTISE and INDIRECT INTERTEXTUAL PRIMING (which will be discussed in more detail below; this confirms hypothesis 1). We can thus conclude that the cognitive constraints underlying linguistic decision-making in student vs professional translators function equally to a very large extent, with the exception of (a subtype of) structural priming.

Table 3. Significant fixed effects in the multi-level generalised linear mixed-effects model of subject placement in French-to-Dutch news translation, using reference coding (the reference value of each categorical fixed predictor is given in brackets); c = 0.99, $\chi^2(17) = 637.19$, $p < 2.2^{e-16}$ ***

Predictor	**Coefficient**	**S.E.**	**z value**	**p value**
(Intercept)	-3.01	1.37	-2.19	0.03 *
INTERTEXTUAL PRIMING				
(ref: preverbal subject in source text)				
Postverbal subject	7.67	2.07	3.71	0.0003 ***
INDIRECT INTERTEXTUAL PRIMING				
(ref.: no adjunct in 1st position in source text)				
Adjunct in 1st position in source text	9.30	1.15	8.06	7.85^{e-16} ***
COMPLEXITY OF THE VP				
(ref: simple VP)				
Complex VP	4.30	1.20	3.59	0.0004 ***
SUBJECT DISCOURSE STATUS				
(ref: new referent)				
Discourse-new ref.	1.12	2.98	0.38	0.71

	Associated ref.	8.50	2.39	3.56	0.0004 ***
	Given ref.	3.84	1.43	2.68	0.008 **
SUBJECT LENGTH (logged)		-1.07	0.61	-1.76	0.08 .
INTERTEXTUAL PRIMING * SUBJECT LENGTH (logged)					
(ref: preverbal subject in source text)					
	Postverbal	2.69	0.88	3.03	0.002 ***
INTERTEXTUAL PRIMING * COMPLEXITY OF THE VP					
(ref: preverbal subject in source text) (ref: simple VP)					
	Postverbal subject in ST, complex VP	-5.82	1.83	-3.18	0.002 **
COMPLEXITY OF THE VP * DISCOURSE STATUS					
(ref: simple VP) (ref: new referent)					
	Complex VP, discourse-new ref.	1.31	2.20	0.60	0.55
	Complex VP, associated ref.	-5.05	1.56	-3.22	0.002 **
	Complex VP, given ref.	-3.68	1.25	-2.94	0.004 **
SUBJECT LENGTH (logged) * DISCOURSE STATUS					
	(ref: new referent)				
	discourse-new ref.	-2.16	1.76	-1.23	0.22
	associated ref.	-2.78	0.85	-3.26	0.002 **
	given ref.	-1.49	0.61	-2.43	0.02 *
INDIRECT INTERTEXTUAL PRIMING * EXPERTISE					
(ref. no adjunct in 1st position in source text)					
(ref. professional)					
	Adjunct in 1st position, student	-1.55	0.60	-2.57	0.01 *

We will now explore the effects of the significant fixed predictors in more depth (*effects* package in R; Fox and Hong, 2009). Since all fixed predictors occur in interaction effects, their separate main effects will not be discussed (as these are not stable across different contexts). We start with the predictors related to structural priming.

4.2.1 The interaction effect of intertextual priming and subject length

As can be seen in Figure 1, when comparing the position and slope of the lines in the left and right panes, it is clear that subjects that are positioned preverbally in the French source text (left-hand side of Figure 1) are strongly associated with preverbal subjects in the Dutch translated text (since the probability of occurring in postverbal position in Dutch is overall lower than 0, with 0 = 'no effect', < 1: preference for preverbal position, > 1: preference for postverbal position; cf. y-axis), whereas postverbal subjects in the French source text (right-hand side of Figure 1) are associated with postverbal subjects in the Dutch translated text (probability is much higher than 0). Since regression coefficients with INTERTEXTUAL PRIMING turn out to be the highest of all coefficients (cf. Table 3), it can be concluded that subject placement in Dutch translations is chiefly impacted by structural priming from the source text (confirming hypothesis 4a).

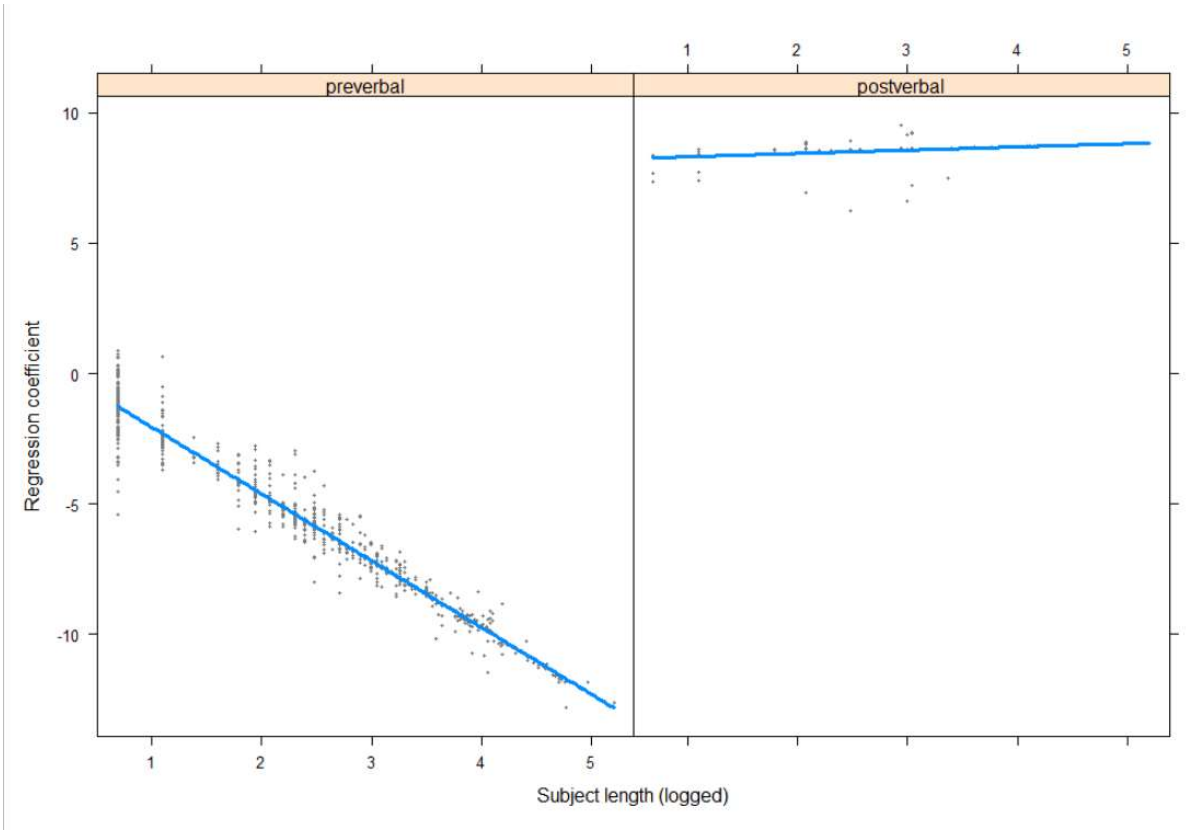

Figure 1. Effect plot of the interaction between SUBJECT LENGTH (in characters, on a log-scale) and INTERTEXTUAL PRIMING (preverbal = preverbal subject in French source text, postverbal = postverbal subject in French source text)

Second, when taking subject length into account, it was observed that subject length does not really play a great role in contexts where the subject is primed by a postverbal subject in the French source text (the line in the right-hand pane does not increase or decrease by much), whereas subject length does play a major role in contexts with a preverbal French source-text prime. The longer subjects get, the stronger the structural priming effect: long subjects in the Dutch translation that are primed by a preverbal subject in the French source texts have a much higher likelihood of being placed in preverbal position than shorter Dutch subjects (partially confirming hypothesis 2).

Third, given the absence of a significant interaction effect between EXPERTISE on the one hand and INTERTEXTUAL PRIMING and SUBJECT LENGTH on the other, it can be concluded that both predictors function equally in students and professionals.

4.2.2 The interaction effect of intertextual priming and complexity of the VP

The interaction effect between INTERTEXTUAL PRIMING and the COMPLEXITY OF THE VP, which operationalises the structural integration constraint, is represented in Figure 2. The comparison of the two panes gives an insight into the effect of INTERTEXTUAL PRIMING, with the same conclusion as above: preverbal subjects in the source text trigger the use of preverbal subject in the Dutch translations, and postverbal subjects in the French source text trigger postverbal subjects in the Dutch translations. More interestingly, the effect of COMPLEXITY OF THE VP appears to be larger in contexts with a postverbal subject prime in the French source text compared to contexts with a preverbal subject prime in the

source text (the distance between the lines in the right pane is much greater than in the left pane). In contexts with a preverbal prime in the source text, complex VPs slightly increase the probability of having a postverbal subject compared to simple VPs. This is in line with our expectations (hypothesis 5 is confirmed), in that it shows that translators indeed try to minimise structure integration processing costs by placing the subject and the main verb of the clause, which is located at the end of the clause in cases of complex VPs, as adjacent as possible (i.e. postverbally). In clauses with postverbal French subjects (right pane), the likelihood of being placed postverbally is much higher with simple VPs than with complex VPs (nevertheless, in both contexts postverbal subjects are in any case preferred), which runs counter to our expectations. We will return to this in the next section. Finally, since EXPERTISE does not participate in an interaction effect with complexity of the VP, we can conclude that the effect of the cognitive constraint of structural integration functions similarly in students and professionals.

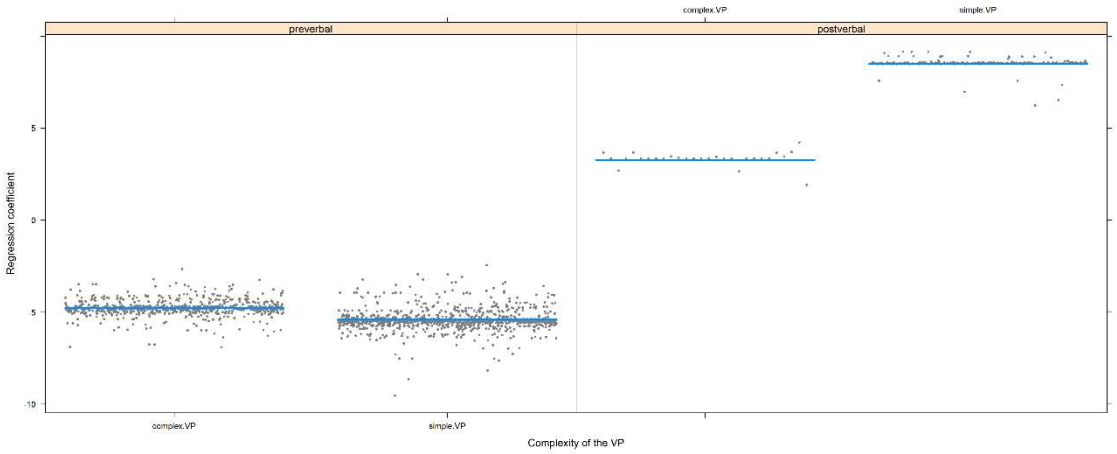

Figure 2. Effect plot of the interaction between COMPLEXITY OF THE VP and INTERTEXTUAL PRIMING (preverbal = preverbal subject in French source text, postverbal = postverbal subject in French source text)

4.2.3 The interaction effect of indirect intertextual priming and expertise

Figure 3 shows the interaction effect between EXPERTISE and INDIRECT INTERTEXTUAL PRIMING: in contexts without an adjunct in sentence-initial position in the French source text, both students and professionals have a similar tendency to predominantly place subjects preverbally (the regression coefficients are well under 0; left pane). However, in cases where a clause-initial adjunct is present in the French source text, it often primes an adjunct in sentence-initial position in the Dutch translation, thereby causing inversion and forcing the subject in Dutch into postverbal position (right pane). This is visible in both student and professional translations (confirming hypothesis 4c), although to a lesser extent in student translations, which more frequently resort to preverbal subject position than professionals.

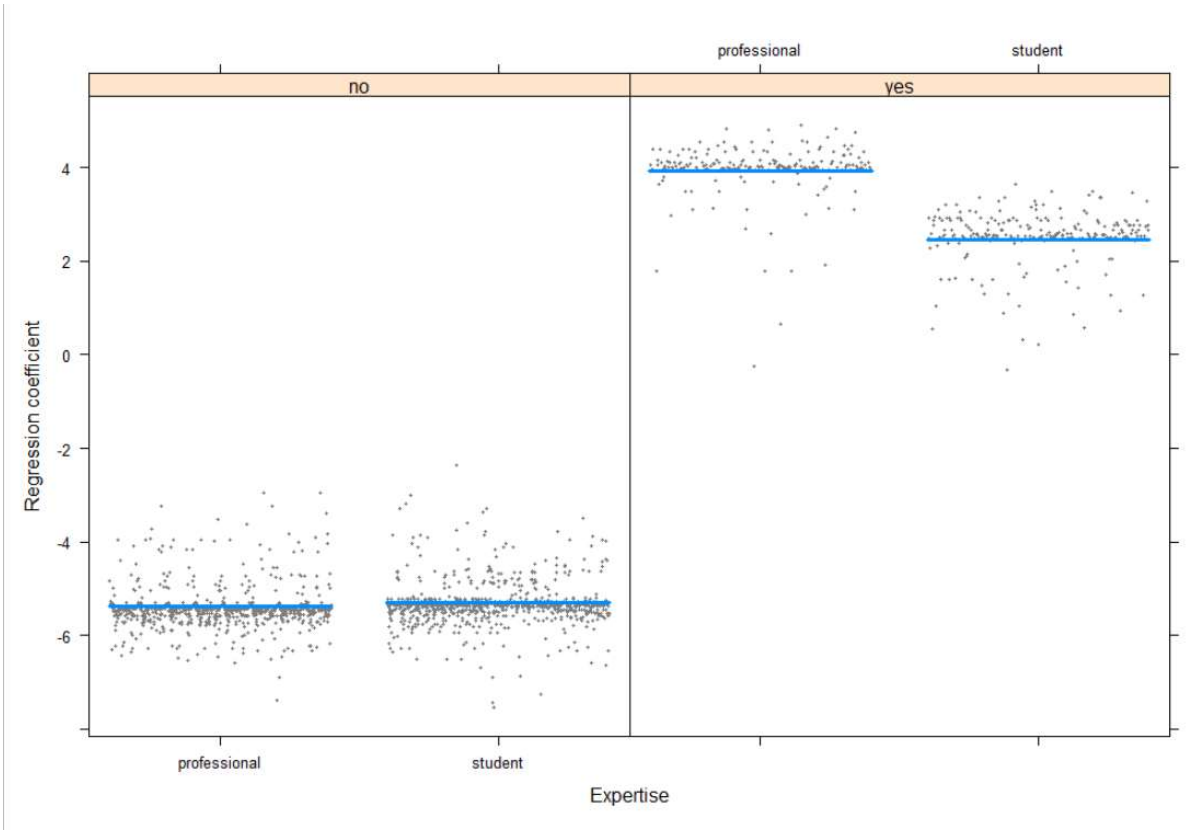

Figure 3. Effect plot of the interaction between EXPERTISE and INDIRECT INTERTEXTUAL PRIMING (no = no adjunct in sentence-initial position in the source text, yes = adjunct in sentence-initial position in the source text)

In order to understand this observation, compare Examples (11), (11a) and (11b), which are representative of these cases in professional and student translation. The source segment (Example 11) contains a left-dislocated adjunct (in italics), followed by the subject (underlined) and the finite verb (in bold). The professional translator (11a) started the Dutch translation with an equivalent adjunct, followed by an inverted verb-subject order (thus shifting the preverbal subject in the French source text to postverbal position in the Dutch translation), whereas the student translator (11b) postponed the adjunct. In this example, the student split the original sentence into two separate sentences, one with the translation of the main clause itself, followed by another sentence containing the translation of the French adjunct (which thus becomes an independent clause in the Dutch translation). Other students connected the postponed adjunct in the same sentence with vague connectors (such as *en* 'and') or embedded it as a relative clause or an apposition. All in all, students do not seem to behave in such a uniform way as professional translators when confronted with this type of left-dislocated adjuncts in French.

(11) *Dans cette station balnéaire densément peuplée de retraités modestes*, il **devrait** être réélu lors d'une législative partielle […]. [source segment]

'In this seaside town densely populated with modest pensioners, he should be re-elected in a by-election […].'

(11a) *In deze badplaats die dichtbevolkt is met bescheiden gepensioneerden,* **zou** hij naar verwachting in een tussentijdse verkiezing moeten worden herkozen […]. [professional translation: P2-2-11]

'In this seaside town densely populated with modest pensioners, he is expected to be re-elected in a by-election […].'

(11b) Hij **moet herkozen worden** in de voorverkiezing voor de partij die de angst voor vreemdelingen misbruikt. *In de badplaats Clacton-on-Sea wonen veel gepensioneerden met een bescheiden inkomen*. [student translation: S-33-62]

'He should be re-elected in the by-election for the party that exploits the fear of foreigners. The seaside town of Clacton-on-Sea is home to many pensioners on modest incomes.'

In conclusion, this interaction effect confirms the effect of yet another intertextual structural priming predictor while at the same time showing that the way that professionals deal with the specific situation of a left-dislocated adjunct in the French source text is much more automatised (viz. keep the adjunct in the translation in sentence-initial position and invert the subject-verb order) than the way students deal with it (different postponing strategies). This aspect of the priming constraint thus seems to function differently in students compared with professionals (confirming hypothesis 1).

4.2.4 The interaction effect of subject discourse status and complexity of the VP

The final constraint that affects subject placement in Dutch translations – alongside structural priming and structural integration – is markedness of coding, for which SUBJECT LENGTH, which was already discussed in a previous interaction, and SUBJECT DISCOURSE STATUS were considered good proxies (the other proxies did not yield significant effects).

Figures 4 shows the effect of SUBJECT DISCOURSE STATUS in interaction with VP COMPLEXITY.

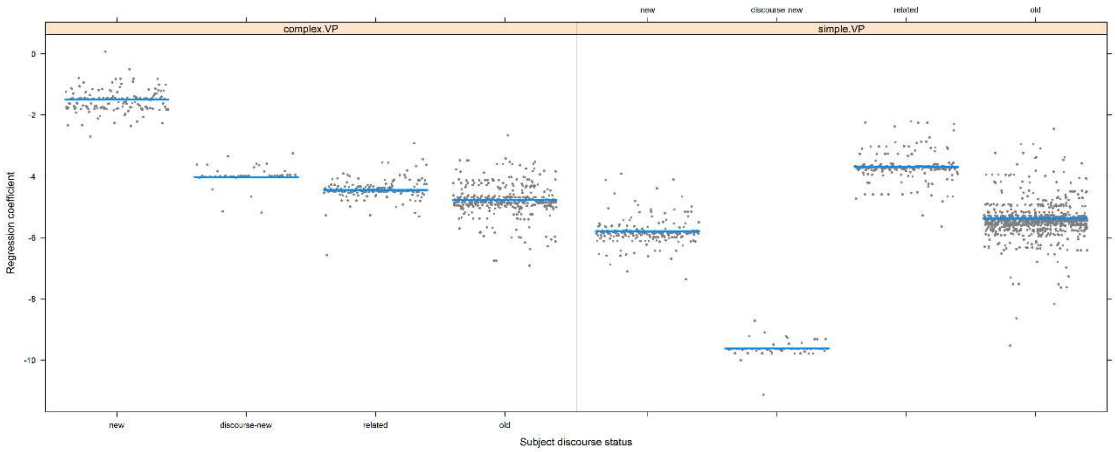

Figure 4. Effect plot of the interaction between SUBJECT DISCOURSE STATUS and COMPLEXITY OF THE VP

It emerges that, overall, clauses with a complex VP (left pane) have a higher likelihood of containing a postverbal subject (vs a preverbal subject) than clauses containing a simple VP (right pane), which is evident from the generally lower position of the lines in the left pane compared with those in the right pane. This is what was already observed in Figure 2. More importantly, the effect of the predictor SUBJECT DISCOURSE STATUS plays out quite differently in clauses with and without a complex VP. In clauses with a simple VP (i.e. consisting of just one finite verb, which is the main verb of the clause; right pane), new and discourse-new subjects are more likely to be placed preverbally, just as hypothesis 2 predicted: these types of referent were held to be less salient and thus in need of the most unmarked position for subjects in Dutch, i.e. the preverbal position. This is especially true of discourse-new referents, and less so of (entirely) new referents. It is not clear where this difference comes from, although it can be observed that discourse-new subjects in clauses with a simple VP take the form of a proper noun proportionally much

more often than new subjects, which more often take the form of a common noun. When turning to clauses with complex VPs (left pane), the position of new and discourse-new subjects changes completely, as these two categories have a much higher probability of occurring postverbally, contrary to what hypothesis 2 predicted. The position of the two types of subject containing (more or less) given information does not change drastically. One possible (and admittedly very speculative) reason for the preference of new and discourse-new subjects to occur postverbally in complex VP clauses might be that these subjects are more 'in need' of being as close as possible to the main verb, which is located at the end of the clause. If that should indeed turn out to be the case, this type of clause constitutes an area of conflict between the markedness of coding constraint and the syntactic integration constraint, with the latter constraint overruling the former.

4.2.5 The interaction effect of subject discourse status and subject length

Figure 5 shows that increasing subject length decreases the likelihood of a postverbal subject, although the steepness of the slope differs significantly depending on the discourse status of the subject. This is in line with our initial prediction (hypothesis 2): longer subjects are less salient subjects, hence have a lower probability of being placed in a more marked position, i.e. the postverbal position.

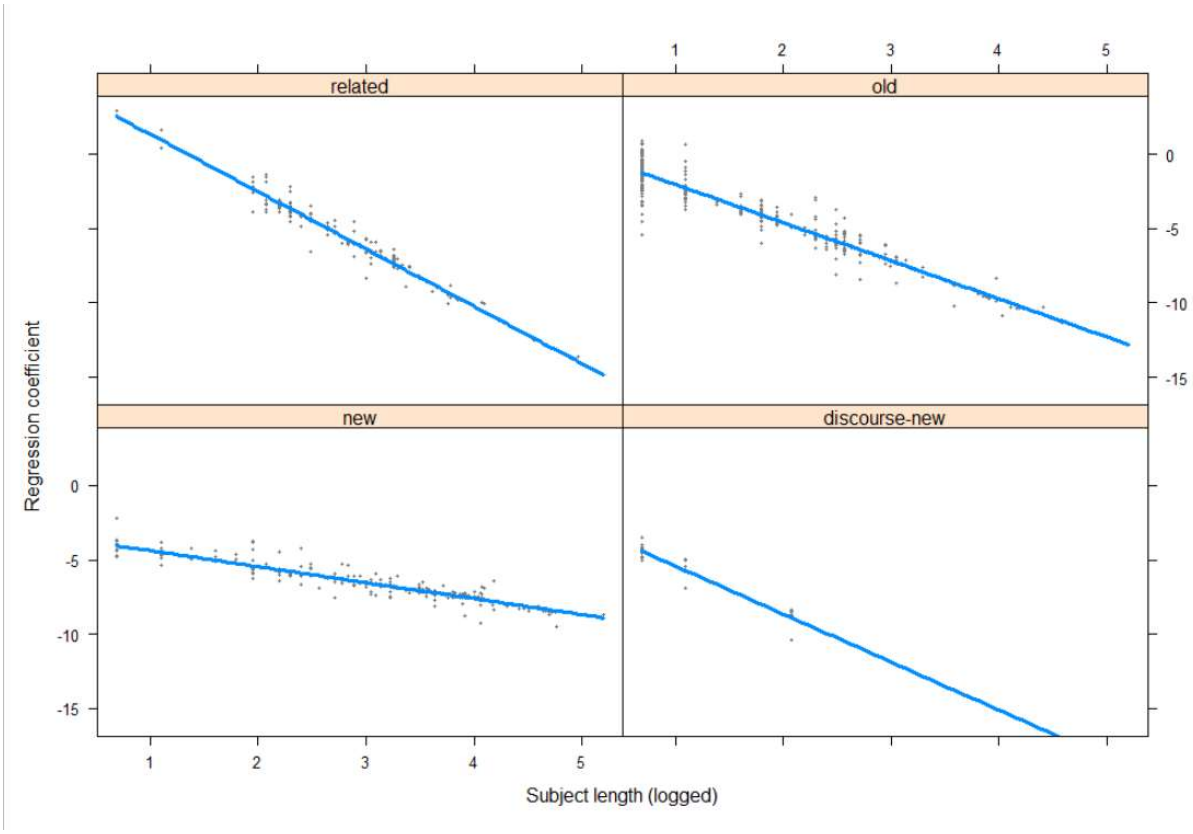

Figure 5. Effect plot of the interaction between SUBJECT DISCOURSE STATUS and SUBJECT LENGTH

5. Discussion

The main finding that emerges from our study is that the position of the subject in Dutch translations is shaped by the same cognitive constraints in both student and professional translations: viz. structural priming, markedness of coding and structural integration. The only difference between students and professionals is the way indirect intertextual priming functions, which was tentatively explained as a difference between automatised translation strategies in professionals vs ad hoc strategies in students. The only constraint that does not have a significant effect at all is cognitive routinisation, operationalised as the statistical attraction between the verb and a preverbal vs postverbal subject, and the

statistical attraction of a (pro)nominal lemma to the preverbal vs postverbal position. In this section, these empirical results are placed in a broader perspective.

We will start with the observation that subject placement in student translations is by and large influenced by the same cognitive constraints as in professional translations. This finding may be remarkable, given the available empirical evidence of differences between students' and professionals' linguistic output in previous studies, but there are reasons why such a difference is less likely in subject placement variation. First, subject placement is a ubiquitous phenomenon. When producing language, any language user, irrespective of their expertise, needs to decide (consciously or unconsciously) at least once per sentence where to place the subject. Additionally, subject-verb combinations in Dutch are acquired early by children (from 2.5 years onwards; cf. Blom and de Korte, 2008). Hence, given the frequency of occurrence of the construction and the early age of acquisition, one could assume that the part of the probabilistic grammar governing subject placement variation is no longer subject to significant fluctuations by the time students start their higher education, especially when taking into consideration that both students and professionals are native speakers of the target language in the corpus used here. Variable linguistic phenomena which are somewhat less frequent might show more fluctuations and consequently divergences across expertise levels.

Second, one could argue that potential differences in subject placement between the two groups, which would emerge in other types of writing where there is no source text to start from, are simply attenuated in translation given the dominance of (intertextual) structural priming from French source texts. This dominance is especially striking in cases of postverbal French subjects, which are almost always translated with postverbal subjects in Dutch (96.6%) (see Table 4).

Table 4. Frequency of pre- and postverbal subjects in French across pre- and postverbal subjects in Dutch (row percentages)

	Preverbal subjects in Dutch	**Postverbal subjects in Dutch**
Preverbal subjects in French	71.8% (n = 1450)	28.2% (n = 570)
Postverbal subjects in French	3.4% (n = 9)	96.6% (n = 253)

Almost all of these postverbal subjects in French are located in clauses with initial direct speech, as in Example 12 (cf. Riegel, Pellat and Rioul, 2009: 256).

(12) ‘Je suis obligée de fonctionner ainsi pour préserver ma sécurité’, [explique]$_{Vfin}$ t'[elle]$_{subject}$ […]. (P1-33-8)
‘I am obliged to work like that to keep my security’, explains she […].
‘’I have to work like that for my own security’, she explains […].’

In clauses with preverbal subjects in French, translators do shift the subject much more often to postverbal position (28.2%; compared with clauses with postverbal French subjects, which are shifted to the other position in Dutch in only 3.4% of the cases), but, as we have seen, the impact of intertextual structural priming remains very significant.

The structural priming constraint, operationalised as three different predictors, each representing a specific type of priming – INTERTEXTUAL, INTRATEXTUAL and INDIRECT PRIMING – mainly had the expected effect, in line with previous corpus-linguistic and psycholinguistic research: preverbal subjects in the source text prime

preverbal subjects in the target text, and the same applies to postverbal subjects. Of all the constraints investigated in this study, (intertextual) structural priming has the strongest effect on subject placement in Dutch translations, thereby confirming the importance of cross-linguistic transfer of syntagmatic configurations that are shared by the source and target languages. Although our research design does not reflect translation processes as such, our results tie in nicely with Schaeffer and Carl's (2013) recursive model of translation, which predicts that initial phases in the translation process, which they call 'default, horizontal translation', are automatic, in the sense that they are largely based on mental representations which are shared between source and target language items and constructions. This is the part of the translation process where priming from the source text occurs (cf. 'transcoding' in Christoffels, De Groot and Waldorp, 2003). So-called 'vertical translation processes' monitor this output and intervene as soon as the translation is unacceptable given target-language norms or contextual considerations. This model thus explains the high impact of structural priming (horizontal translation), as well as the systematic shift from preverbal position in French to postverbal position in Dutch in clauses with fronted adjuncts (indirect intertextual priming).

Intratextual priming, i.e. priming from a preceding sentence in the Dutch translation to the next sentence in that translation, did not have any effect. This is in line with previous research by De Sutter et al. (2021), who reach a similar conclusion in a corpus-based study on optional *that* in translated and non-translated English. It is not quite clear yet why intratextual priming does not play a role in translation, although it might be hypothesised that translators primarily switch back and forth between source text reading and target text production, with most source-text fixations around four to six words to the right of the word being translated (Carl, Dragsted & Jakobsen, 2011). This switching behaviour does not leave much room for intratextual priming to play an important role.

Later phases in the translation process obviously also involve (monolingual) target-text reading and revision, but since this is mostly not a text production phase any more, it seems that intratextual priming is no longer able to affect the text produced.

Next to priming, the syntactic integration constraint had the second-largest effect on subject placement in Dutch translations, confirming the importance of adjacency of subjects and the main verb of the clause as structurally and conceptually closely related clausal entities. Also, the hypothesised effect of markedness of coding in terms of SUBJECT LENGTH and (partially) SUBJECT DISCOURSE STATUS was corroborated, showing that constituents that are somewhat less salient to occur as subject in a clause (long subjects, new referents) occur preferentially in the unmarked preverbal position when they do take up the role of subject. Other operationalisations of the markedness of coding constraint such as ANIMACY and CONCRETENESS did not turn out to be significant, which is in line with previous research on subject placement in non-translated Dutch (Vogels and Van Bergen, 2017: 385-389). Given the effect of the other predictors, animacy does not play a major role in subject placement.

Finally, it was observed that cognitive routinisation, which plays a central role in usage-based linguistic theories, does not affect subject placement at all. What we can conclude, for the time being at least, is that co-occurrence frequency between specific lemmas and abstract syntagmatic slots (i.e. the preverbal and the postverbal slot) is not stored in memory, and neither is co-occurrence frequency of a specific verb lemma and an abstract syntagmatic slot. Even if this information is stored somehow, it apparently does not affect subject placement in translation in a significant way, given the effect of the other cognitive constraints that are significant.

6. Conclusion

This article set out to investigate the effect of four well-known cognitive constraints on subject placement in French-to-Dutch student vs professional news translation, using a corpus-based approach. Our results have revealed that subject placement in both student and professional translated declarative main clauses is influenced by structural priming from the source text, markedness of coding and structural integration, but not by cognitive routinisation. The way these constraints operate in student and professional translators is remarkably similar. We found only one context in which structural priming in student translations operated differently than in professional translations: when the French source clause starts with a left-dislocated adjunct, professional translators (as compared to students) are more often primed to place the adjunct at the start of the Dutch translation, thus causing inversion (i.e. subject in postverbal position), which was considered a case of indirect priming (the position of the subject is a consequence of a primed constituent earlier in the sentence). Students, by contrast, hardly choose this option: they tend to postpone (the contents of) the adjunct.

We hope to have shown that it is possible, at least to some extent, to conduct theoretically relevant, cognitively inspired research using an advanced, multifactorial corpus methodology. In our opinion, much is to be gained in the field of corpus-based translation studies and learner corpus research by considering general cognitive constraints which have been used to explain other language patterns and, more generally, other patterns that involve cognitive processing, thereby gaining firmer theoretical ground and offering opportunities for more comparison and cooperation across disciplinary borders.

Obviously, there is still much that remains to be done. For one thing, the effect of subject discourse status in clauses with complex VPs was puzzling, with new and discourse-new subjects preferring a postverbal position where a preverbal position was expected. To what extent the proposed explanation – structural integration overrides markedness of coding – makes sense will need to be examined more closely, for instance using an experimental design.

From a contrastive perspective, a question that also needs to be answered in future research, using a reference corpus of non-translated Dutch journalistic texts, is to what extent the structural priming effect causes deviances in the frequency with which pre- and postverbal subjects are used in Dutch translated texts and, as a corollary, to what extent information structure in translations diverges from that in non-translations. French has a more pronounced preference for preverbal subjects (cf. Riegel et al., 2009: 212, 243-244) and pronominal subjects (Lambrecht, 2010: 77). It is also more tolerant towards placing more than one constituent in the preverbal position (Fries, 1981), and resorts much more frequently to structures which are more marked in Dutch, such as clefting, pseudoclefting and dislocation. Conversely, Dutch, like English, is more tolerant towards using subjects in focus positions and placing focus expressions in preverbal position. In (spoken) French, there is "a near one-to-one mapping between focus structure and phrase structure: Topic expressions occur overwhelmingly in preverbal position and in pronominal form, while focus expressions occur postverbally. To avoid violating this near one-to-one mapping constraint, spoken French makes abundant use of grammatical realignment constructions, especially clefts" (Lambrecht 2010: 77). These cross-linguistic differences between French and Dutch are likely to have an effect on information structure in translations vs non-translations.

References

Arnold, J. E., Kaiser, E., Kahn, J. M., & Kim, L. K. (2013). Information structure: linguistic, cognitive, and processing approaches. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(4), 403–413.

Barlow, M., & Kemmer, S. (Eds.). (2000). *Usage-based models of language*. Center for the Study of Language and Information (CSLI).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Blom, E., & de Korte, S. (2008). De verwerving van het Nederlands: dummies en Verb Second. *Nederlandse Taalkunde*, *13*, 133–159.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Bouma, G. (2008). *Starting a sentence in Dutch. A corpus study of subject- and object-fronting.* Groningen Dissertations in Linguistics, *66*.

Broekhuis, H. (2020). 8.1.2. Noun phrases in clause-initial position. *Taalportaal*. https://www.taalportaal.org/taalportaal/topic/link/syntax____Dutch__np__n8__nouns8_Syntactic.8.1.2.xml.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, *150*, 80–84.

Bybee, J. L., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure* (Vol. 45). John Benjamins.

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford University Press.

Calvillo, D. P., & Jackson, R. E. (2014). Animacy, perceptual load, and inattentional blindness. *Psychonomic Bulletin & Review*, *21*(3), 670–675.

Carl, M., Dragsted, B., & Jakobsen, A. L. (2011). A taxonomy of human translation styles. *Translation journal*, *16*(2), 155–168.

Castagnoli, S. (2016). Investigating trainee translators' contrastive pragmalinguistic competence: a corpus-based analysis of interclausal linkage in learner translations. *The Interpreter and Translator Trainer*, *10*(3), 343–363.

Christoffels, I. K., De Groot, A. M., & Waldorp, L. J. (2003). Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism: Language and Cognition*, *6*(3), 201–211.

Claes, J. (2017). Probabilistic grammar: The view from Cognitive Sociolinguistics. *Glossa: A Journal of General Linguistics*, *2*(1), Art. 62.

Corpas Pastor, G., Mitkov, R., Afzal, N., & Pekar, V. (2008). Translation universals: do they exist? A corpus-based NLP study of convergence and simplification. In: *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas: Research Papers*, 75–81.

Coussé, E. (2009). Focus, complexiteit en extrapositie. Over de veranderende woordvolgorde in het Nederlands. *Neerlandistiek.nl*, *9*(4). https://dspace.library.uu.nl/handle/1874/35192.

De Sutter, G., Cappelle, B., De Clercq, O., Loock, R., & Plevoets, K. (2017). Towards a corpus-based, statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translations. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, *16*, 25–39.

De Sutter, G., Colleman, T., & Ghyselen, A.-S. (2021). Intra- and inter-textual syntactic priming in original and translated English. In G. Kristiansen, K. Franco, S. De

Pascale, L. Rosseel, & W. Zhang (Eds.), *Cognitive sociolinguistics revisited* (Applications of Cognitive Linguistics 48) (pp. 264–276). Mouton de Gruyter.

Diamond, B. J., & Shreve, G. M. (2010). Neural and physiological correlates of translation and interpreting in the bilingual brain. *Translation and cognition*, *15*, 289–321.

Diessel, H. (2017). Usage-based linguistics. In M Aronoff, *Oxford research encyclopedia of linguistics*. Oxford University Press. https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-363

Diessel, H. (2019). *The grammar network*. Cambridge University Press.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1–24.

Falk, S. (2014). On the notion of salience in spoken discourse-prominence cues shaping discourse structure and comprehension. *Travaux interdisciplinaires sur la parole et le langage*, *30*. http://journals.openedition.org/tipa/1303 ; DOI : https://doi.org/10.4000/tipa.1303

Fox, J., & Hong, J. (2009). Effect Displays in R for Multinomial and Proportional-Odds Logit Models: Extensions to the effects Package. *Journal of Statistical Software*, *32*(1), 1–24.

Fries, P. H. (1981). On the status of Theme in English: arguments from discourse. *Forum Linguisticum*, *6*, 1–38.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, *2000*, 95–126.

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of psycholinguistic research*, *34*(4), 365–399.

Gries, S. Th. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, *10*(1), 95–125.

Gries, S. T., & Kootstra, G. J. (2017). Structural priming within and across languages: A corpus-based perspective. *Bilingualism: Language and Cognition*, *20*(2), 235–250.

Grondelaers, S., & Speelman, D. (2007). A variationist account of constituent ordering in presentative sentences in Belgian Dutch. *Corpus Linguistics and Linguistic Theory*, *3*(2), 161–193.

Haeseryn, W., Romijn, K., Geerts, G., Rooij, J., & Toorn, M. van den (2019). *Algemene Nederlandse Spraakkunst*. https://e-ans.ivdnt.org/.

Harrell, F. E. (2015). *Regression modeling strategies. With applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

Hartsuiker, R. J., Beerts, S., Loncke, M., Desmet, T., & Bernolet, S. (2016). Cross-linguistic structural priming in multilinguals: Further evidence for shared syntax. *Journal of Memory and Language*, *90*, 14–30.

Hawkins, J. A. (1994). *A performance theory of order and constituency* (No. 73). Cambridge University Press.

Hawkins, J. A. (2001). Why are categories adjacent? *Journal of Linguistics*, *37*(1), 1–34.

Jansen, F. & Wijnands, R. (2004). Doorkruisingen van het links-rechtsprincipe. *Neerlandistiek.nl* 04.01. https://dspace.library.uu.nl/handle/1874/28494.

Kajzer-Wietrzny, M. (2019). Linking words in inter- and intralingual translation. Combining corpus linguistics and key-logging data. In L. Vandevoorde, J. Daems & B. Defrancq (Eds.), *New Empirical Perspectives on Translation and Interpreting* (pp. 114–138). Routledge.

Kaushanskaya, M., Rechtzigel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychonomic Bulletin & Review*, *19*, 935–941.

Khoe, Y. H., Tsoukala, C., Kootstra, G. J., & Frank, S. L. (2021). Is structural priming between different languages a learning effect? Modelling priming as error-driven implicit learning. *Language, Cognition and Neuroscience*, 1–21.

Kunilovskaya, M., & Lapshinova-Koltunski, E. (2020). Lexicogrammatic translationese across two targets and competence levels. *Proceedings of the 12th Language Resources and Evaluation Conference*, 4102–4112.

Kunilovskaya, M., Morgoun, N., & Pariy, A. (2018). Learner vs professional translations into Russian: Lexical profiles. *Translation and Interpreting*, *10*(1), 33–52.

Labov, W. (1965). On the mechanism of linguistic change. *Georgetown monographs on language and linguistics*, *18*, 91-114.

Lakoff, G. (1990). The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics*, *1*(1), 39–74.

Lambrecht, K. (2010). Constraints on subject-focus mapping in French and English: A contrastive analysis. In C. Breul & E. Göbbel (Eds.), *Comparative and Contrastive Studies of Information Structure* (pp. 77–100). John Benjamins.

Lapshinova-Koltunski, E. (2022). Detecting normalisation and shining-through in novice and professional translations. In S. Granger & M.-A. Lefer (Eds.), *Extending the Scope of Corpus-Based Translation Studies* (pp. 182–206). Bloomsbury.

Lapshinova-Koltunski, E., Popović, M. & Koponen, M. (2022). DiHuTra: a Parallel Corpus to Analyse Differences between Human Translations. *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, 335–336.

Maier, R. M., Pickering, M. J., & Hartsuiker, R. J. (2017). Does translation involve structural priming? *Quarterly Journal of Experimental Psychology*, *70*(8), 1575–1589.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2015). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group*. R package version 1.6-7. https://CRAN.R-project.org/package=e1071

Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013). The construction of a 500-million-word reference corpus of contemporary written Dutch. In P. Spyns & J. Odijk (Eds.), *Essential Speech and Language Technology for Dutch. Theory and Applications of Natural Language Processing* (pp. 219–247). Springer.

Pickering, M. J., & Ferreira V. S. (2008). Structural priming: a critical review. *Psychological bulletin*, *134*(3), 427.

Pijpops, D. (2020). What is an alternation?: Six answers. *Belgian Journal of Linguistics*, *34*(1), 283–294.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.

R Core Team (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.R-project.org/

Redelinghuys, K., & Kruger, H. (2015). Using the features of translated language to investigate translation expertise: A corpus-based study. *International Journal of Corpus Linguistics*, *20*(3), 293–325.

Riegel, M., Pellat, J.-C., & Rioul, R. (2009). *Grammaire méthodique du français*. Presses Universitaires de France.

Rosenbach, A. (2005). Animacy versus weight as determinants of grammatical variation in English. *Language*, 613–644.

Schaeffer, M., & Carl, M. (2013). Shared representations and the translation process: A recursive model. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, *8*(2), 169–190.

Schmid, H. J. (2015). A blueprint of the entrenchment-and-conventionalization model. *Yearbook of the German Cognitive Linguistics Association*, *3*(1), 3–26.

Szmrecsanyi, B. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, *1*(1), 113–150.

Tiselius, E., & Hild, A. (2017). Expertise and competence in translation and interpreting. In J. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 423–444). Wiley.

Vogels, J., & Van Bergen, G. (2017). Where to place inaccessible subjects in Dutch: The role of definiteness and animacy. *Corpus linguistics and linguistic theory*, *13*(2), 369-398.

Wasow, T. (2002). *Postverbal behavior*. CSLI.

Zwart, J.-W. (2011). *The syntax of Dutch*. Cambridge University Press.

Address for correspondence

Gert De Sutter

Ghent University

Abdisstraat 1, BE-9000 Gent

Belgium

gert.desutter@ugent.be