

Compounding in French-Dutch translation: Source-language priming and creative strategies – Preprint

How to cite: Hendrikx, I., Van Praet, W., & Van Goethem, K. (2026). Compounding in French-Dutch translation: Source-language priming and creative strategies. *Across Languages and Cultures: a multidisciplinary journal for translation and interpreting studies*, 16(1), 64-99.

Isa Hendrikx – Université de Liège (Corresponding author)
Isa.hendrikx@uliege.be

Wout Van Praet – University of Zürich
wout.vanpraet@es.uzh.ch

Kristel Van Goethem - F.R.S.-FNRS & Université Catholique de Louvain
Kristel.vangoethem@uclouvain.be

Abstract

This study examines how novice translators handle Dutch word formation, specifically compounding, when translating from French. While both languages permit multi-word expressions, Dutch more readily employs compounding, presenting challenges for French speakers. Building on previous research using the MULTINCo corpus – which showed that French-speaking learners of Dutch tend to overuse phrasal structures and produce creative or erroneous compounds – this study broadens the scope to include translations by both L1 Dutch and L1 French university students.

The findings confirm that translators tend to overgeneralize phrasal patterns as a result of source-language priming from French, irrespective of their native language. Unexpectedly, this effect is not more pronounced among L1 French speakers, suggesting that explicitation may be a broadly adopted strategy when translating French multi-word expressions into Dutch. Nonetheless, L1 French translators more frequently produce non-idiomatic multi-word expressions, particularly in terms of spelling, linking morphemes, and grammatical structure.

Regression analysis indicates that compound formation is influenced by lexical frequency, structural complexity, and L1 background. However, individual variation accounts for a substantial portion of the unexplained variance.

The results of the study underscore a continuum between creative translation strategies and linguistic errors, and reveal that even native speakers of the target language are not immune to translation challenges.

Keywords:

Dutch compounding; corpus-based translation studies; linguistic creativity; explicitation

1 Introduction

In both translation studies and second language acquisition research, the role of word formation and morphological awareness has garnered increasing attention as a key factor influencing language proficiency and translation quality (cf. Prinzie et al., 2022; Hendriks & Van Goethem, 2024; Spina, 2025). Dutch and French each offer a range of morphological and syntactic strategies for constructing multi-word expressions. For example, Dutch distinguishes between *wetenschapsbeleid* (‘science policy’) and *wetenschappelijk beleid* (‘scientific policy’) (Booij, 2019), while French uses forms such as *village(-)vacances* and *village de vacances* (‘holiday resort’). Notably, however, Dutch displays a stronger tendency toward compounding than French (see, among others, Van Goethem, 2009). These cross-linguistic differences, which will be further elaborated in Section 2.1., are challenging for French-speaking learners of Dutch. Building on previous research on the acquisition of Dutch compounds by French-speaking learners (Hendriks & Van Goethem, 2024), this study investigates how Dutch compounds are translated by two groups of novice¹ translators, native speakers of Dutch and French-speaking learners of Dutch.

Section 2.2 outlines how the present study builds on previous research into compound use among additional language learners. Since both learner language production and translation can be viewed as forms of “constrained communication”, this concept is introduced in Section 2.3. Section 2.4 then discusses how non-target-like or infrequent uses of compounds and phrasal constructions by novice translators can be positioned along a continuum ranging from creative to erroneous. The article proceeds with a presentation of the study’s precise aims and hypotheses in Section 3, followed by a description of the data and methodology in Section 4. The results of the corpus study are presented in Section 5. The paper concludes with a summary and discussion in Section 6.

2 State of the art

2.1 Cross-linguistic differences

Schlücker (2019) explores the similarities and differences between compounds and syntactically formed multi-word expressions across various European languages. Despite their formal differences, both structures can serve similar purposes and may even compete with each other. For instance, in Dutch, *atoomfysica* competes with *atomaire fysica* ‘atomic physics’ (Booij, 2019: 105).

¹ We use the term “novice translators” to refer to students studying applied linguistics (BA) or translation studies (MA) at one of five Belgian universities which participated in our study, see Section 4.1.

However, languages – even within the same genealogical family – can grammaticalize at different rates, leading to varying degrees of analyticity (e.g., van Haeringen, 1956, Hüning et al., 2006, Lamiroy, 2011). This difference is evident in compounding, where Dutch shows a stronger tendency towards compound formation than French (Van Goethem, 2009).

Noteworthy, the term “compound” is used quite inconsistently in French morphology literature (Van Goethem & Amiot, 2019), often referring to various types of complex lexical units regardless of their formation process. From a restrictive perspective (Fradin, 2009, Villoing, 2012), which we adopt as well, only a limited set of patterns should be considered true compounds in French, such as [Verb + Noun] (e.g., *brise-glace* ‘icebreaker’), [Adj + Adj] (e.g., *sinocoréen* ‘Sino-Korean’), [Noun + Noun] coordinates (e.g., *auteur-compositeur* ‘author-composer’), and [Noun + Noun] subordinates (e.g., *poisson-chat* ‘catfish’). In contrast, Dutch compounding is highly productive, creating complex nouns, verbs, adjectives, and prepositions. Dutch speakers can easily coin new compounds, which can be embedded into other compounds, leading to complex structures like *postzegelverzamelaarsbijeenkomst* ‘stamp collectors meeting’ (Don, 2009).

Besides a difference in productivity, we can also observe different word order patterns between both languages: Dutch compounds are typically right-headed, whereas French compounds and phrases tend to be left-headed (cf. *bloemkool* vs *chou-fleur* ‘cauliflower’). Finally, Dutch compounds are clearly distinguishable from phrases based on formal criteria such as spelling, prosody, and loss of adjectival inflection. In contrast, French compounds are not always easily distinguishable from syntactic expressions. For example, Dutch *zoetwaterzwembad* is clearly a compound, while its French equivalent *piscine d’eau douce* ‘freshwater pool’ should be considered a lexicalized phrase.

2.2 From learner data to translator data

Based on these differences between Dutch and French compounding, it can be expected that French-speaking learners of Dutch will find it difficult to use Dutch compounds correctly, and more specifically, that they will overuse phrasal structures. A previous study (Hendrikx & Van Goethem, 2024) focused on learner data from the MULTINCo corpus (Meunier et al., 2023), involving 195 learners in French-speaking Belgium. This study confirmed that compounds are more frequent in L1 Dutch than in L1 French and that French-speaking learners of Dutch tend to overuse phrasal structures in contexts where a compound should be used. For example, they produced *lessen van zwembad* ‘classes from swimming pool’ instead of *zwemlessen* ‘swimming classes’. Additionally, the study revealed that learners sometimes use compounds where a simplex word would be more appropriate, such as *kookman* ‘lit. cookman’ instead of *kok* ‘chef’, and that they may even code-switch within compounds, as seen in *verjaardagsgateau* ‘birthday cake’. The present paper provides a complementary study by focusing on novice translators’ use of Dutch compounds.

2.3 Translation as a form of “constrained communication”

The study adopts the corpus-based approach that Baker (1993) brought to translation studies. Her work emerged alongside a disciplinary shift from source-oriented models to target-system

perspectives, consistent with Toury's (1978) move from prescriptive "equivalence" to descriptive norms. Baker introduced the concept of translation universals, identifying distinctive features that commonly appear in translated texts and differ from those in original utterances. These universals, such as simplification, explicitation, and normalization, underscored the systematic tendencies in translation processes, shifting the focus away from analysing translations solely in relation to their source texts (Baker, 1993: 243).

However, over the last few decades, discourse within corpus-based translation research has provoked debates over the universality of translation phenomena. Scholars have increasingly viewed translation as a form of "constrained communication", since translators operate under multiple, often competing constraints, including sociocultural, technological and cognitive factors. From this perspective, translation choices are shaped less by universal linguistic tendencies and more by historically and culturally situated norms that govern what counts as an acceptable translation in a given community. Consequently, claims about universal translation tendencies (such as simplification or explicitation) are reinterpreted as probabilistic effects of contextual constraints rather than inherent properties of translation itself (Lanstyák & Heltai 2012; Kruger & Van Rooy 2016; De Sutter & Lefer 2020).

2.4 Source-language priming and creative translation strategies

This perspective is reinforced by cognitive studies, which also rehabilitate the focus on the source language, demonstrating widespread priming effects, occurring in both translation directions, though the magnitude of these effects tends to be larger from L1 to L2 (Wen & Van Heuven, 2017). Priming effects in translation can be understood through the lens of structural priming research in psycholinguistics, which demonstrates that prior exposure to syntactic patterns influences subsequent production (Bock, 1986; Pickering & Branigan, 1998; Ferreira & Bock, 2006). In our study, we hypothesize that French phrasal constructions will prime novice translators to favour multi-word expressions over Dutch compounds, even when the latter are more idiomatic. This aligns with previous findings that priming persists during translation tasks and that even experienced translators initially opt for more literal translations before refining them to more appropriate ones (Meyer & Schvaneveldt, 1971; Tirkkonen-Condit, 2002).

These insights are relevant to understanding the practices of novice translators, who often face the challenge of navigating linguistic limitations and knowledge gaps. These may include difficulties in vocabulary retrieval or uncertainties surrounding collocational norms. Overcoming such challenges frequently requires creative problem-solving. However, this raises the question: how can one distinguish between creative translation strategies and linguistic errors? Uhrig (2020) proposes *deliberateness* as a key criterion for differentiation: where creativity is marked by intentionality, errors typically arise unintentionally.

This distinction, however, is not always clear-cut and can depend on additional factors. For instance, frequency effects play a significant role in shaping language expectations and acceptability. The phenomenon of *preemption* or *negative entrenchment* demonstrates how frequent and conventionalized expressions can inhibit alternative forms. An example is the Spanish phrase *vino tinto* ('red wine'), which blocks the more literal but less idiomatic *vino*

rojo (Uhrig, 2020: 5). In this sense, even potentially plausible constructions may be perceived as erroneous if they violate entrenched patterns. Perceived creativity also hinges on the hearer's expectations and contextual interpretation (Uhrig, 2020: 8). Thus, when evaluating the linguistic output of novice translators or second-language users, it is essential to consider not only the form and frequency of constructions but also the broader communicative intentions and interpretive frameworks of the interlocutors.

Moreover, contemporary research advocates for a multifactorial analysis of translation decisions. Rather than isolating translation status as the sole determinant, scholars emphasize the need to consider factors like text type and the complexity and frequency of the translated element. This holistic approach aligns with the evolving understanding that linguistic choices in translation are influenced by a constellation of variables, necessitating nuanced analytical frameworks (Van Beveren, et al., 2018; De Sutter & Lefer, 2020).

3 Aims of the study and research hypotheses

The aim of this paper is fourfold. First, this study seeks to investigate language-specific tendencies towards the use of nominal compounds and their analytic equivalents. Second, it compares the use of Dutch compounds in French-to-Dutch translations by novice translators, whose L1 is Dutch or French. A third aim is to reveal the (creative) strategies novice translators use to tackle problems caused by cross-linguistic differences (e.g. paraphrases, explicitation). Finally, the study will identify the factors involved in the choice between the use of compounds vs. analytic equivalents.

In the light of these aims, we hypothesise that French-to-Dutch translations will overgeneralise the use of phrasal patterns as a priming effect of the French source text, both for L1 French and L1 Dutch translators: for instance, for the translation of French *un effet de surprise* (lit. ‘an effect of surprise’), we expect translators to resort more frequently to (less idiomatic) paraphrastic constructions (e.g., *een effect van verrassing*, lit. ‘an effect of surprise’) than the (more idiomatic) compound alternative (e.g., *een verrassingseffect*, lit. ‘a surprise.effect’). Second, we expect this priming effect to affect translations by L1 French translators more than those by L1 Dutch translators, because the former are also subject to priming from their native language. In line with this, we predict that L1 French translators are more likely to use non-idiomatic expressions. Finally, we conjecture that the complexity (cf. length) and frequency of a ‘potential’ Dutch compound will affect the translator’s choice to use a compound or alternative strategies.

4 Corpus data and methods

4.1 Data collection and corpus sample

The data in this study were collected through an (online) translation task. For the translation task, 30 French source texts were selected from the *Multilingual Traditional, Immersion and Native Corpus* (MulTINCo) (Meunier et al., 2023). The texts were produced by L1 French eighteen-year-olds, who were asked by the MulTINCo compilers to write a short email, in French, to a friend on one of two topics, either a recent vacation or a birthday party. The 30

texts collected for the current study were selected based on their having a high number of potential target compounds when translated to Dutch. In some cases, these texts were edited, namely shortened, for the translation task, so as to keep the timing of the task relatively consistent across the texts. The resulting French source texts consisted, on average, of 211.4 words per text.

The translation corpus for this study was collected in 2024 and consists of French-to-Dutch translations by novice translators. The participants in the translation task were all bachelor or master students in applied linguistics or translation studies at one of five Belgian universities: the University of Antwerp (UAntwerpen), the University of Ghent (UGent), the Free University of Brussels (VUB), the University of Liège (ULiège), and the University of Mons (UMons). Of these five universities, the first three fall under the education system of the Dutch-speaking community, where translation studies are only offered at MA level, following a BA programme in applied linguistics. The other two universities – ULiège and UMons – follow the education system of the French-speaking community, where translation studies are offered at both BA and MA level. Therefore, between participants from Dutch-speaking vs. French-speaking universities, there is a difference, at the BA level, between the specialization of their training as translators.² All participants, however, did have training in translation studies and knowledge of both Dutch and French, either as their L1 or as a non-native language that they are studying as part of their BA or MA programme.

In total, there were 71 participants, of whom 40 were L1 Dutch speakers (of which four were bilingual Dutch-French speakers) and 31 were L1 French speakers.³ Given the linguistic and political landscape in Belgium, the number of L1 Dutch and L1 French speakers need not coincide with the number of students from Dutch-speaking vs. French-speaking universities respectively. 37 participants were BA students, of which 36 were L1 Dutch speakers; 33 participants were MA students, of which 3 were L1 Dutch speakers and 30 L1 French speakers. 1 L1 Dutch participant was enrolled in a bridging programme, which prepares students who do not hold a BA degree in applied linguistics or translation studies for the master's programme in translation studies. The criteria for enrolling in the bridging programme do, however, state that the student should hold a degree in a related field, demonstrating their knowledge of the language(s) they wish to train in. Consequently, this particular L1 Dutch student will have had three years of higher education (partly) in French.

Participants performed the translation task in *LimeSurvey*, an online survey tool, which allocated one of the 30 French source texts randomly to the participants, who carried out the task individually either during class or in their free time. Before the start of the translation task, participants were instructed not to use any resources such as dictionaries, translation tools or spelling checks, with the explicit mention that they should draw only on their own acquired knowledge for the entirety of the task. To minimize the chances of participants using such resources, they were asked if they had read the instructions and agreed to the terms prior to the task (by ticking a box). After the task, they were also explicitly asked whether they had

² Students from BA programmes in applied linguistics might also choose to specialise in interpreting, multilingual communication, journalism, etc. Nonetheless, these programmes all train students in translation studies as well.

³ Some participants who indicated that their L1 was Dutch or French also reported having a second L1 (other than French), *viz.* Arabic (n = 1), Spanish (n = 1) and 'Algerian' (possibly Arabic or Berber) (n = 1).

nonetheless used such resources (by ticking either the ‘yes’ or ‘no’ box). If they answered ‘yes’ to the question, their task was not taken into account.

The task itself consisted of two parts, namely a questionnaire with questions probing for extralinguistic data (L1, region of residence, education level, etc.) and the actual translation task. The order of these parts was fixed, as participants could only start the translation task after filling out the questionnaire. While there was no imposed time limit for the task, participants took, on average, 137.86 seconds (or 2.30 minutes) to finish the questionnaire and 1,281.40 seconds (or 21.36 minutes) for the translation task itself. For the translation task, the French source text was given in full at the top of the page, and participants were asked to type their Dutch translations in the designated textbox below the source text. The whole French source text was available to participants while they were carrying out the translation task. Once submitted, participants could not go back to the translation task. At no point during the experiment were participants made aware of the goal of the study, which was only communicated in an optional debriefing after the data had been collected.

The resulting data are stored anonymously in a private repository by the principal investigator I. Hendriks at the Université de Liège, the anonymised data file utilised for the present analysis can be accessed via the following link: <https://doi.org/10.58119/ULG/VCPMPB>.

4.2 Data annotation

In the translated texts, we annotated all attested nominal compounds, as well as all non-compound expressions where a nominal compound was expected. The absence of such predicted compounds was coded based on the idiomaticity of the predicted compound vs. the actually attested expression (e.g., a paraphrase). As a benchmark for idiomaticity, the frequency of use of the two competing expressions was checked in the *Corpus Hedendaags Nederlands* (CHN): for a predicted compound to be judged as the ‘correct’ or more idiomatic option, we determined that it had to be at least ten times more frequent than its non-compound alternative. If that was not the case, both options were deemed idiomatic, and the attested non-compound expression was not coded as a ‘translation error’. The same threshold value was used in Hendriks & Van Goethem (2024), which facilitates the comparison of the results. To illustrate this with examples from the previous study, the French expression *mes camarades de classe* (‘my classmates’) could be translated as *de mensen in mijn klas* (‘the people in my class’) or as *mijn klasgenoten* (‘my classmates’), but since the former yields one hit in the CHN and the latter 123 hits, only the latter is considered idiomatic. By contrast, for the French *au centre de la ville* (‘in the centre of the city’), both *in het centrum van de stad* (‘in the centre of the city’) (n = 1,076) and *in het stadscentrum* (‘in the city centre’) (n = 956) are taken to be idiomatic, so that the former is not coded as incorrect.

All resulting relevant hits (i.e., attested compounds and non-idiomatic non-compound alternatives) were coded, firstly, for accuracy. Accuracy comprises both word-formation accuracy and accuracy of the compound. Word-formation accuracy concerns the idiomaticity of the translation, based on the same criterion explained in the previous paragraph (cf. frequency

in the CHN). For non-idiomatic ('inaccurate') translations, we further specified the kind of inaccuracy that was attested:

- (i) Use of a phrase instead of a compound (e.g., *'de mensen in mijn klas* 'the people in my class' > *mijn klasgenoten* 'my classmates')
- (ii) Use of a single-item lexeme instead of a compound (e.g., *kaars* 'candle' rather than *verjaardagskaars* 'birthday candle')
- (iii) Use of a compound when it was not predicted (e.g., *drie chocoladen cake* [sic] 'three chocolate cake' rather than *cake met drie soorten chocolade* 'cake with three kinds of chocolate')
- (iv) Omission of the target translation

Then, for all attested compounds – whether idiomatic or unidiomatic – the accuracy of the compound was evaluated, based on the following six parameters:

- (i) Spelling: e.g., **plateland* vs. *platteland* 'countryside'
- (ii) Linking morpheme: e.g., **verjaardagfeest* vs. *verjaardagsfeest* 'birthday party'
- (iii) Lexical choice: e.g., **kokentalent* vs. *kooktalent* 'cooking talent', **champagnetoren* 'champagne tower' vs. *champagneglas* 'champagne glass' for *flutte à champagne* '(long) champagne glass'
- (iv) Grammatical accuracy: e.g., **de verjaardagsfeest* vs. *het verjaardagsfeest* 'the birthday party', or **een heerlijk aardbeienmojito* vs. *een heerlijke aardbeienmojito* 'a delicious strawberry mojito'⁴
- (v) Codeswitching: e.g., **een fléchettes concours* vs. *een dartswedstrijd* 'a darts competition'; **firework* vs. *vuurwerk* 'firework'
- (vi) Word order: e.g., **vuurkamp* vs. *kampvuur* 'campfire'

In addition, all compounds were also analysed in terms of their formal properties, namely the compound length (i.e., the number of lexical components, for instance three in *openluchtbar* 'open air bar') and the compound structure (i.e., the combination of parts of speech of the compound components, e.g., [adjective+noun]_{noun} for *platteland* 'countryside', lit. 'flat.land', [noun+noun]_{noun} for *sportfan* 'sports fan', [verb+noun]_{noun} for *wachtrij* 'queue', lit. 'wait.row', [preposition/adverb+noun]_{noun} for *achtergrond* 'background', etc.).

5 Results

⁴ In Dutch, the inflection of adjectives depends on the definiteness, gender and number of the NP in which they occur. In general, adjectives used in plural NPs always take an *e*-suffix (e.g. *(de) mooie boeken* '(the) beautiful books'), as do adjectives used in singular definite NPs (e.g., *het mooie boek* 'the beautiful book' or *de mooie pen* 'the beautiful pen'). By contrast, adjectives used in indefinite NPs (with article *een*) only take an *e*-suffix if the head noun is gendered (male/female) (e.g., *een mooie pen* 'a beautiful pen' (grammatically female)) but not when the head noun is neuter (e.g., *een mooi boek* 'a beautiful book'). In (iv), the head noun *mojito* is grammatically male, requiring the inflected adjectival form *heerlijke*. The incorrect form **heerlijk* can thus be seen as a grammatical error, as it suggests the incorrect analysis of *aardbeienmojito* 'strawberry mojito' as grammatically neuter.

This section outlines the results for the study of compound use in the collected translation data. First, in Section 5.1, a brief overview is given of the average use of compounds by the two translator groups. In Section 5.2, we describe the general formal properties of the attested compounds, in terms of their length (Section 5.2.1) and internal structure (Section 5.2.2). Third, in Section 5.3, we zoom in on the accuracy of the translations, more specifically the word formation accuracy (Section 5.3.1) and the accuracy of the attested compounds themselves (Section 5.3.2).

5.1 Average compound use

In total, L1 Dutch translators ($n = 40$) produced 309 compounds, compared to 214 compounds produced by L1 French translators ($n = 31$). That means that there was no significant difference between the two groups regarding the average frequency of use of compounds (i.e., 7.73 compounds per L1 Dutch translator vs. 6.90 compounds per L1 French translator), as evidenced by the outcome of a Welch two-sample t -test ($p = .230$) (see also Figure 1). The frequency of compound usage varied significantly among translators, however. For example, ID-36 (a Dutch-speaking translator from Flanders) produced 15 compounds – 14 two-component compounds and one three-component compound. Some examples include: *le décalage horaire* – *een uurverschil* ‘lit. a hour.difference, a time difference’, *un traitement VIP* – *een VIP-behandeling* ‘a VIP treatment’, *une file* ‘a queue, a line’ – *een wachtrij* ‘lit. a waiting.line’.

In contrast, ID-62 (another Dutch-speaking translator living in Flanders) produced a phrase where a compound could have been used on seven occasions, for example: *2 derniers mois de vacances* ‘2 last months of holidays’ - *de laatste 2 maanden vakantie* instead of *de laatste 2 vakantiemaanden* ‘lit. the last 2 holidays.months’, *une équipe de football* ‘lit. a team of football’ - *een ploeg van Amerikaans football* instead of *een Amerikaans voetbalteam* ‘lit. an American football.team’, *sa ville voisine* ‘lit. her city neighbour’ - *de stad naast haar* instead of *de buurstad* ‘lit. the neighbour.city’.

In other words, L1 Dutch and L1 French translators are equally likely to use compounds in their translations from French to Dutch. However, as we will see in the remainder of this section and the following one, the complexity of the attested compounds, as well as the accuracy of the compound use, can be seen to differ between the two groups.

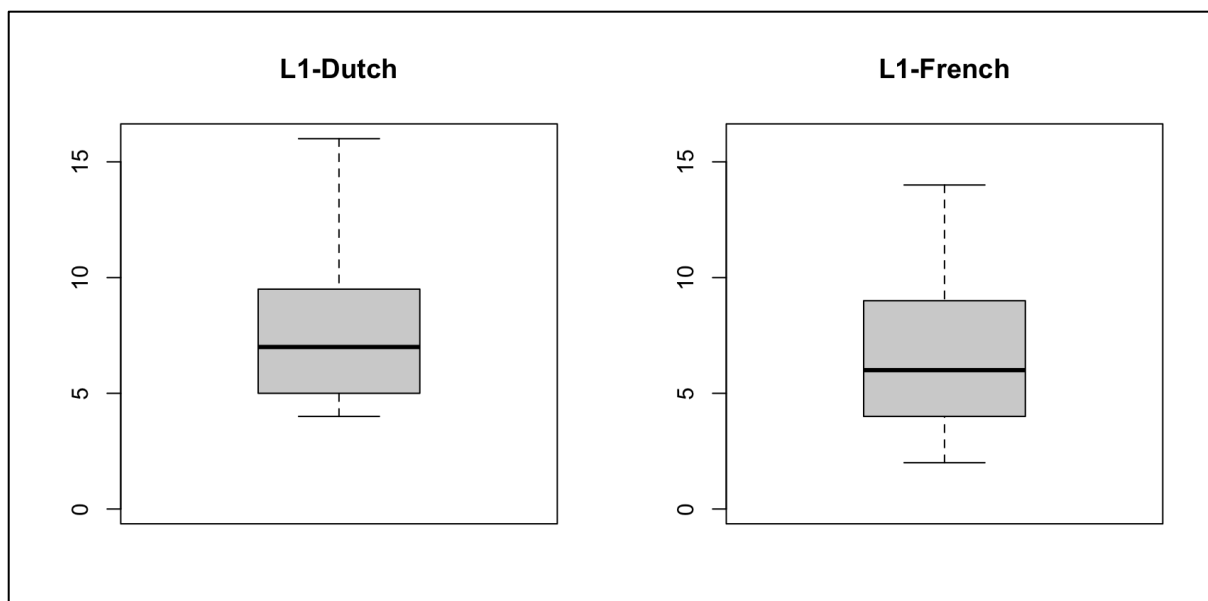


Figure 1. Boxplot illustrating the frequency of compound use for L1 Dutch vs. L1 French translators

5.2 Formal properties

5.2.1 Compound length

The first formal property, namely compound length, measures the number of components (i.e., lexical items) that make up the attested compounds (see Figure 2).

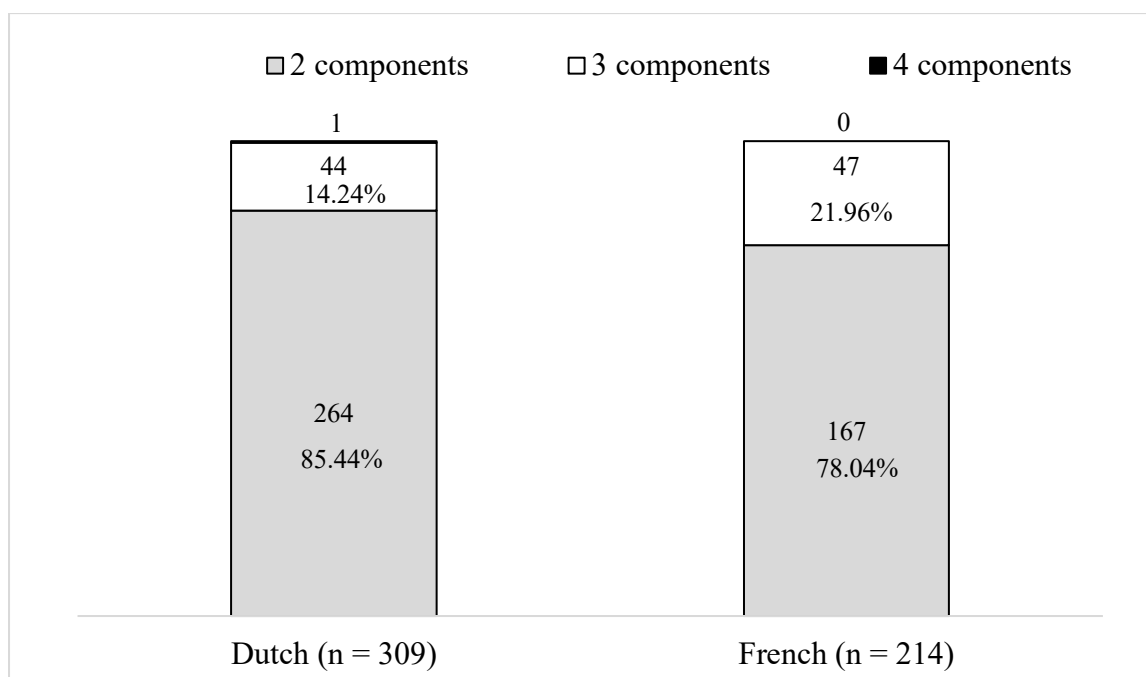


Figure 2. Compound length per translator group

Taken together both translator groups, most of the compounds consisted of two items only ($n = 431$, or 83%), e.g. (1), while compounds with three components ($n = 91$, or 17%) or four components ($n = 1$, or nearly 0%) were much less frequent, e.g., (2) and (3) respectively.

- (1) *Het was een echte **droomvakantie**.* (French *des vacances de rêve*)
‘It was a real **dream vacation** (lit. ‘dream.vacation’).’
- (2) *We besloten hem ... naar een **sportautosalon** te sturen.* (French *un salon de voiture de sport*)
‘We decided to send him ... to a **sportscar show** (lit. ‘sports.car.salon’).’
- (3) *Als een echte **zoogdierliefhebber**, kon hij eindelijk zijn kinderwens vervullen.* (French *fan de mammifères*)
‘As a true **mammal lover** (lit. ‘suckle.animal.love.haver’), he could finally fulfil his childhood wish.’

A significant difference was observed between the translator groups ($p = .0296$, Fisher Exact test), with L1 French translators being relatively more likely to use compounds with three components (i.e., $n = 47$, or 22%, for L1 French translators vs. $n = 44$, or 14% for L1 Dutch translators). Only one example of a compound with four components was observed, namely in the L1 Dutch group, see (3).

5.2.2 Compound structure

For the second formal property, we analysed the internal structure of the attested compounds, namely the combination of parts of speech of the individual components (see Figure 3).

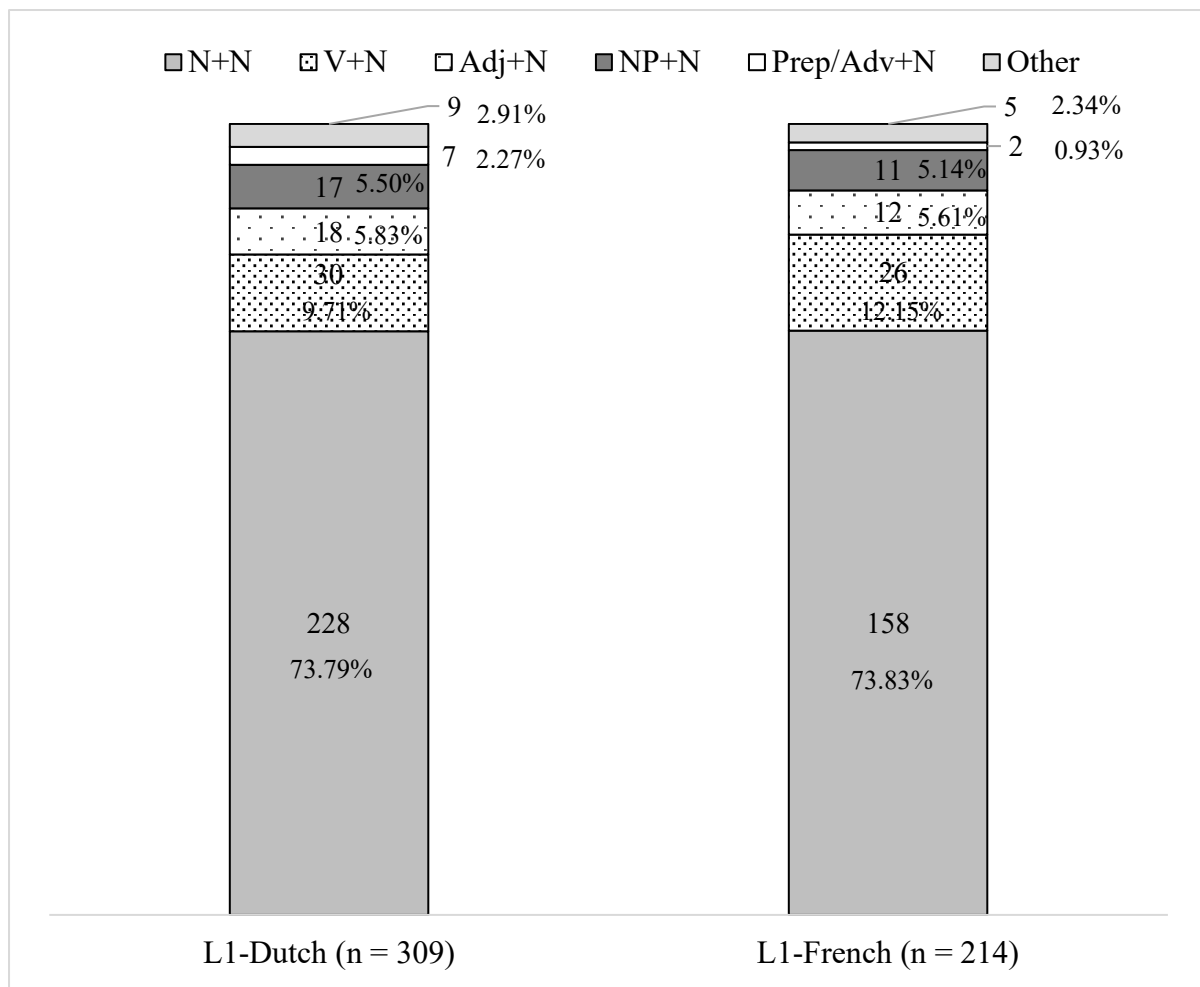


Figure 3. Compound structure per translator group

Between the translator groups, no significant differences were found ($p = .847$, Fisher Exact test). Most compounds, across translator groups, consisted of two nouns ($n = 386$, or 74%), e.g., *zomervakantie* ‘summer holidays’ in (4).

- (4) *Ik moet je dringend over mijn **zomervakantie** vertellen!* (French *mes vacances d’été*)
 ‘I urgently have to tell you about my **summer holidays**!’

Compounds with verb-noun combinations were the second most frequent option ($n = 56$, or 11%), e.g., *surflessen* ‘surf lessons’ in (5).

- (5) *Ik heb zelfs **surflessen** kunnen volgen!* (French *des cours de surf*)
 ‘I even got to take **surf lessons**!’

Adjective-noun and NP-noun combinations were roughly equally frequent (respectively $n = 30$, or 6%, and $n = 28$, or 5%), e.g., *hoogseizoen* ‘high season’ in (6) and *driesterrenhotel* ‘three-star hotel’ in (7) respectively.

- (6) *Het was natuurlijk niet het **hoogseizoen** voor de Amerikanen.* (French *la grande saison*)

‘Of course, it wasn’t the **high season** for the Americans.’

- (7) *Daarna stappen we richting het **driesterrenhotel**.* (French *l’hôtel trois étoiles*)
‘Afterwards, we walk towards the **three-star hotel** (lit. ‘three.stars.hotel’).’

Finally, combinations of prepositions or adverbs and nouns ($n = 9$, or 2%) are rare, e.g., *achtergrond* ‘background’ in (8), as are all other combinations ($n = 14$, or 3%), e.g., numeral-noun combinations such as *zesdejaars* ‘sixth-formers’ in (9).

- (8) *We hadden bijna altijd de zee op de **achtergrond** omdat Sicilië een eiland is.* (French *en arrière plan*)
‘We almost always had the sea in the **background** (lit. ‘behind.ground’) because Sicily is an island.’
- (9) *Deze vakantie [sic.] ben ik met een groep **zesdejaars** naar Kroatië geweest.* (French *une bande de dernière année secondaire*)
‘This holiday, I went to Croatia with a group of **sixth-formers** (lit. ‘sixth.years’).’

These findings suggest that the internal structure, or complexity, of the compound makes little to no difference for compound use between translator groups.

5.3 Accuracy

5.3.1 Word formation accuracy

For translation accuracy, we first analysed the accuracy of the word formation, i.e., the idiomaticity of translation choices for compounds vs. non-compound alternatives.

First, on a general level, L1 Dutch translators were, as expected, significantly less likely to make non-idiomatic choices ($n = 80$, or 21%) compared to the other L1 French ($n = 92$, or 31%) ($\chi^2(1) = 8.781$, $p = .003$; Cramer’s $V = 0.117$) (Figure 4).

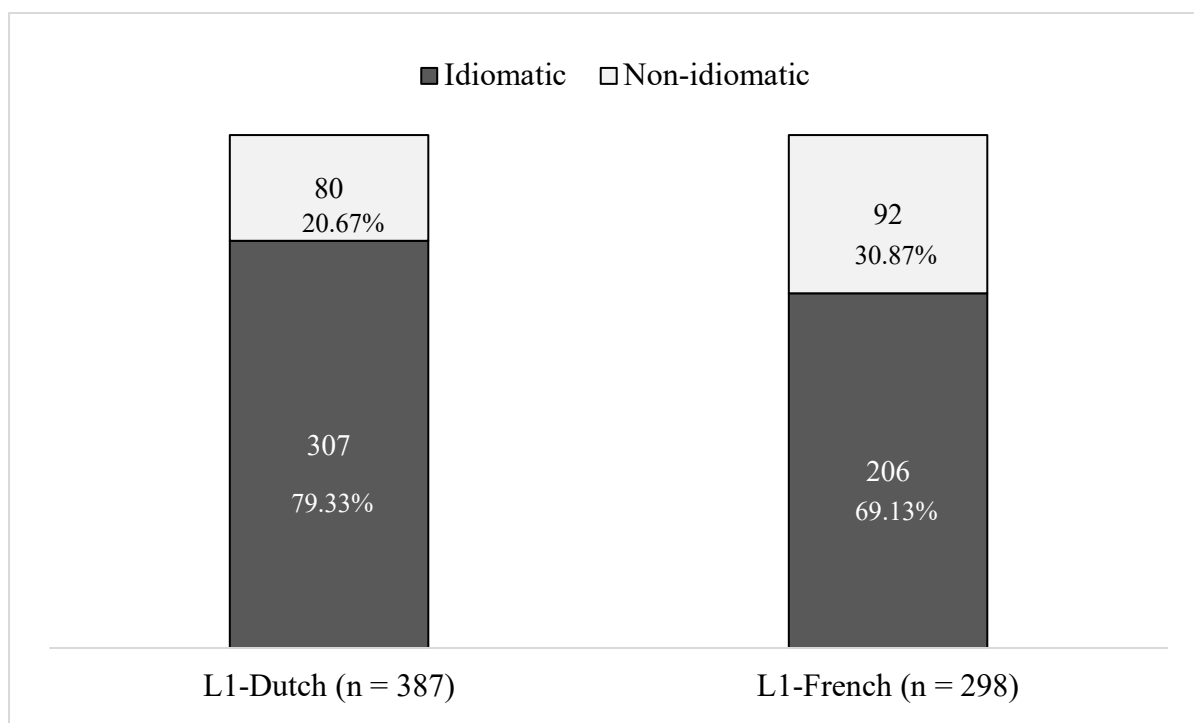


Figure 4. Word formation accuracy per translator group

Second, if we zoom in on the unidiomatic choices, there are no significant differences between L1 Dutch and L1 French translators ($p = .356$, Fisher Exact test) (Figure 5).

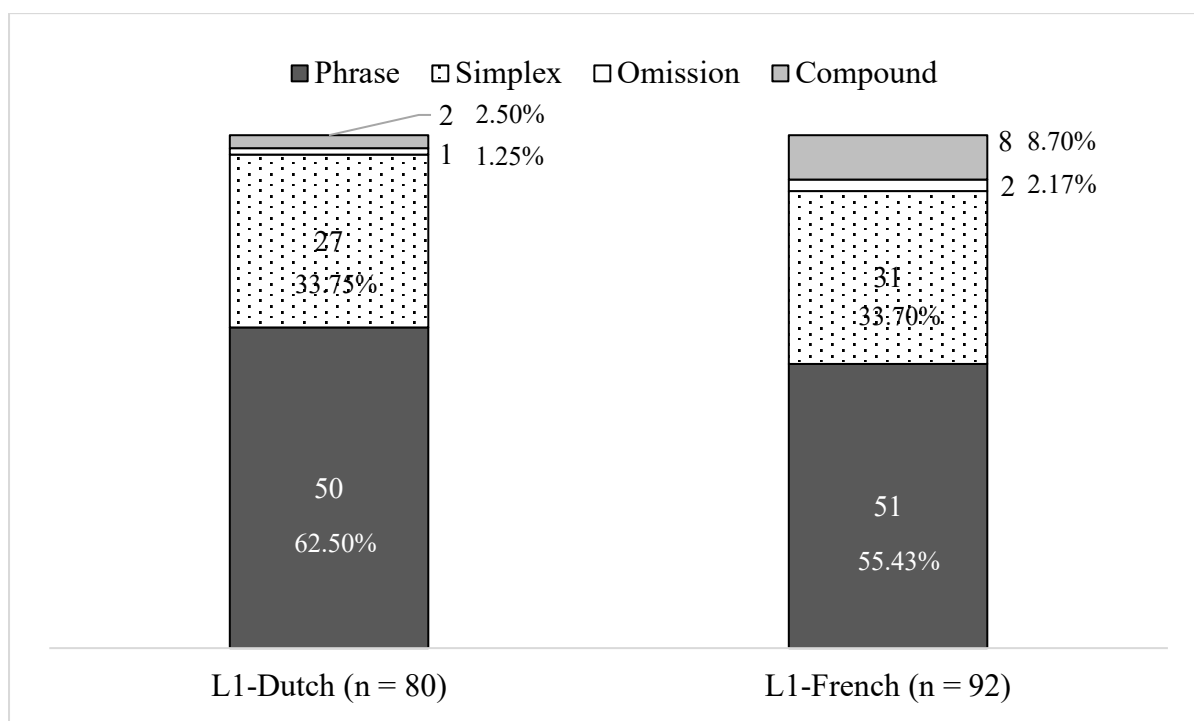


Figure 5. Distribution of specific unidiomatic word formation choices per translator group

In other words, while the overall frequency of unidiomatic choices differs between translator groups, the strategies they use when making unidiomatic choices are generally similar. That is, most unidiomatic choices result from translators using a phrase where a compound is expected ($n = 50$, or 62%, for L1 Dutch, and $n = 51$, or 55%, for L1 French), giving evidence of priming – e.g., *toestellen voor de landbouw* ‘machines for the agriculture’ rather than *landbouwtoestellen* ‘agricultural machines’ in (10) – or from them using a simplex word rather than a compound – e.g., *kaart* ‘card’ rather than *bankkaart* ‘bankcard’ in (11).

(10) *Maar hier in Chili zijn de toestellen voor de landbouw nog niet zo goed ontwikkeld.* (French *les machines agricoles*)

‘But here in Chile, the **machines for the agriculture** are not so well developed.’

(11) *Hij moest alle [sic] zijn documenten opnieuw maken en zijn kaart stopzetten.* (French *sa carte bancaire*)

‘He had to recreate all his documents and stop his **card**.’

Other unidiomatic choices – namely omissions and the use of compounds when they are not expected, e.g., *feestzaal* ‘party hall’ rather than *zaal* (‘hall, room’) in (12) – are rare.

(12) *Er was ook een zwembad naast de feestzaal.* (French *la salle*)

‘There was also a swimming pool next to the **party hall**.’

To understand what factors influence the likelihood of word formation errors, we used a logistic regression model. The best-fitting model included the following predictor variables: (i) the frequency of the target compound (i.e., the attested or expected compound), (ii) the length of the target compound (i.e., in number of lexical components), (iii) L1, (iv) education level, and (v) the individual translator. The model is summarized by means of an Anova in Table 1, which gives the effect of each individual predictor. The model has a Nagelkerke R^2 of 0.38. While not particularly high, the variation explained by this model is much greater than for a model that does not take the variation between individual translators into account, which had a Nagelkerke R^2 of only 0.14.

Table 1. Summary of the logistic regression model for word formation accuracy

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)	
NULL			684	772.04		
Frequency	1	35.554	683	736.49	2.481e-09	***
Length	3	15.951	630	720.54	0.001160	***
L1	1	8.526	679	712.01	0.003502	**
Education level	4	10.476	675	701.54	0.033126	**
Individual	66	131.319	609	570.22	3.141e-06	***
Significance codes: $p < .001$ ‘***’, $p < .01$ ‘**’, $p < .05$ ‘*’, $p < .01$ ‘.’						

Based on the model in Table 1, we can conclude that compound frequency ($p < .001$) and length ($p = .001$) are language-internal predictors of the likelihood of translators producing a compound when expected. In general, the higher the frequency of a target compound the more

likely it is to be used accurately, which seems to be due, to a large extent, to the highly lexicalized nature of these compounds (many of which correspond to a singular lexical item in French), e.g., *verjaardag* ‘birthday’ for French *anniversaire* in (13), and the fact that the L2-speakers have likely encountered and used this compound before, favouring the entrenchment of these highly frequent compounds (Ellis & Wulff, 2008). By reverse analogy, target compounds with a low frequency of use – often more creative compounds – are more likely to lead to formation errors. For instance, in (14), a more accurate translation for French *alcool de lichi* would have been the ad-hoc compound *lychee-alcohol* or *lycheelikeur* ‘lychee alcohol / liqueur’ rather than the use of the simplex *lichi* ‘lychee’. In (15), a better translation for French *l’heure de l’apéro* would have been the idiomatic compound *aperitief tijd* or *aperitief moment* ‘aperitif time’ rather than the less idiomatic phrase *tijd van de apéro* ‘time of the aperitif’.

- (13) *Gisteravond was het Sophie haar verjaardag.* (French *l’anniversaire*)
‘Yesterday, it was Sophie’s **birthday**.’
- (14) *Ik heb ook wat gefeest en veel lichi gedronken.* (French *alcool de lichi* ‘lychee alcohol’)
‘I also partied a bit and drank lots of **lychee**.’
- (15) *We zijn op tijd van de apéro aangekomen.* (French *l’heure de l’apéro*)
‘We arrived at (the) **time of the aperitif**.’

In addition, word formation errors are more likely to occur with longer target compounds than with shorter ones. In (16), for instance, the French phrase *musique d’ambiance* (lit. ‘music of ambience’) could be translated by means of a Dutch compound *sfeermuziek* (lit. ‘atmosphere.music’) (or potentially *achtergrondmuziek* ‘background.music’). Instead, the L1 French translator opted for the phrase *muziek in de achtergrond voor de sfeer* ‘music in the background for the atmosphere’, which allows them to translate both the literal meaning of French *ambiance* ‘ambience’ (via *voor de sfeer* ‘for the atmosphere’) and the implication that such music is typically background music (via *in de achtergrond* ‘in the background’). At the same time, the original French relative clause *qui sort des baffles* (‘that comes out of the speakers’) is omitted, reducing the complexity of the translated phrase.

- (16) *In de metro hebben we gebabbeld, met muziek in de achtergrond voor de sfeer.* (French *petite musique d’ambiance qui sort des baffles*, lit. ‘ambient music coming out of the speakers’)
‘In the metro, we chatted, with **music in the background for the atmosphere**.’

Furthermore, while translators’ L1 ($p = .004$) and education level ($p = .03$) also influence word formation accuracy as language-external variables, the variation among individual translators ($p < .001$) is greater than the variation between L1 groups or between education levels. Finally, the fact that even a model that takes individual variation between translators into account can only explain 38% of the total variation suggests that most of the variation in word formation accuracy is either random or attributable to some other factor not considered here.

5.3.2 Accuracy of the compounds

Finally, the attested compounds were analysed for accuracy, regarding their spelling, use of linking morphemes, lexical accuracy, grammar, codeswitching, and word order (Table 2 and Figure 6) (for examples see Section 4.2 above). The percentages in Table 2 and Figure 6 are calculated relative to the total number of compounds per translator group (e.g., 18 spelling mistakes out of a total of 308 compounds for L1 Dutch translators equals 6%).

Table 2. Relative frequencies of compound inaccuracies per translator group

	L1 Dutch (n = 309)	L1 French (n = 214)
Spelling	18 (6%)	49 (23%)
Linking morpheme	1 (0%)	11 (5%)
Lexis	11 (4%)	13 (6%)
Grammar	2 (1%)	20 (9%)
Codeswitching	24 (8%)	16 (7%)
Word order	2 (1%)	2 (1%)

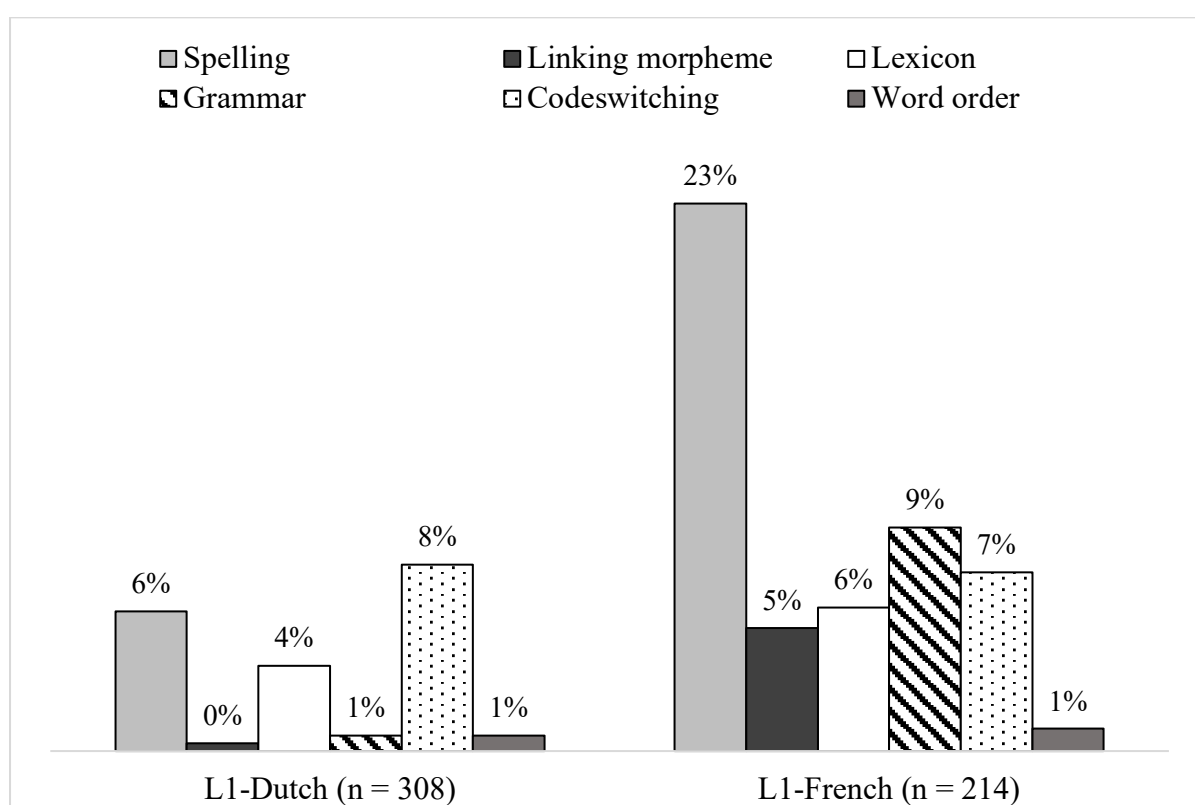


Figure 6. Relative frequencies of compound inaccuracies per translator group

The data in Table 2 and Figure 6 show that, overall, spelling mistakes are most common (n = 67, or 13%), e.g., *oost kunst* [cf. *oostkust*] ‘east coast’ in (17).

(17) *Ik was op de oost kunst* [cf., *oostkust*] van het land. (French *la côte ouest*)

‘I was on the **east coast** of the country.’

Codeswitching ($n = 40$, or 8%) is less common – e.g., *een firework* [cf. *vuurwerk*] in (18) – followed by grammatical mistakes ($n = 21$, or 4%) – e.g., *de kleine hotel voor de jeugd* ‘the small hotel for the youth’ (cf. the grammatically correct (but unidiomatic) phrase *het kleine hotel voor de jeugd* or the idiomatic compound *de jeugdherberg* ‘youth hostel’) – and lexical mistakes ($n = 20$, or 4%), e.g., *velden van zand* ‘fields of sand’ (cf. *zandstranden*, lit. ‘sand.beaches’).

(18) *Rond middernacht hadden de ouders een **firework*** [cf., *vuurwerk*] *gepland*. (French *un feu d’artifice*)

‘Around midnight, the parents had organized a **firework** [i.e., a fireworks display].’

(19) *We zijn eerst naar **de kleine hotel voor het jeugd*** [cf. *de jeugdherberg* ‘the youth hostel’, or] *gegaan*. (French *l’auberge de jeunesse*)

‘We first went to **the small hotel for the youth**.’

(20) *Ik hou ook van de vaste [sic] **velden van zand*** [cf., *de uitgerekte zandstranden* ‘the long sandy beaches’, lit. ‘sand.beaches’] *die daar te vinden zijn*. (French *les vastes plages de sable*)

‘I also love the solid **fields of sand** that are to be found there.’

Inaccuracies regarding linking morphemes ($n = 12$, or 2%) and word order ($n = 4$, or 1%) are, generally, not common, e.g., (21) and (22) respectively.

(21) *Gisteren ben ik naar Guillaume’s **verjaardagfeest*** [cf. *verjaardagsfeest*] *gegaan*! (French *fête d’anniversaire*)

‘Yesterday, I went to Guillaume’s **birthday party**!’

(22) *Op het einde hebben we een **vuurkamp*** [cf. *kampvuur* ‘campfire’] *opgebrand*. (French *feu de camp*)

‘At the end, we burnt up a **firecamp**.’

However, the likelihood of inaccuracies differed significantly between L1 Dutch and L1 French translators, specifically with respect to spelling ($p < .001$, Fisher’s exact test), the use of a linking morpheme ($p < .001$), and grammar ($p < .001$). In all cases, it is the L1 French translator group that is more likely to make mistakes. By contrast, for lexis ($p = .20$), codeswitching ($p = 1$) and word order ($p = 1$), no significant differences are found between the translator groups. For word order, this is because errors are generally rare for both translator groups. For codeswitching, however, translation ‘inaccuracies’ are relatively common with both L1 French ($n = 16$, or 7%) and L1 Dutch translators ($n = 24$, or 8%). This is because all cases of codeswitching, in the use of compounds, were coded as ‘inaccuracies’, which includes both less idiomatic uses of codeswitching (e.g., *firework* in (18)) and more idiomatic uses, such as *de eventmanager* in Dutch (rather than the less frequently used alternative *de evenementenplanner* ‘events.planner’) for French *la préparatrice de l’événement*. Interestingly, when using codeswitching, both L1 Dutch and L1 French translators are more likely to switch to English

than to French (but see, for instance, *kaskaden* for *watervallen* ‘waterfalls’ and *volcans* for *vulkanen* ‘volcanoes’ in (23)).

- (23) ***Kaskaden*** [cf. *watervallen* ‘waterfalls’], ***volcans*** [cf. *vulkanen* ‘volcanoes’], *grote groene plekken, met vrije schepen* [cf. *schapen*] *en coeien* [cf. *koeien*]...
‘Waterfalls, volcanoes, big green places, with free-ranging sheep and cows...’

This suggests that code-switching is more likely to result from borrowing or influence from English rather than from priming from French.

6 Discussion

This study focused on French-Dutch translation by novice translators, emphasizing the cross-linguistic differences between the two languages in the domain of compounding. Previous research has highlighted that differences in productivity within this field can pose obstacles from a language learning perspective. This follow-up study introduces an additional complication due to potential priming of the source language in translation, representing another form of constrained communication. We hypothesized that (i) novice translators would overgeneralize phrasal patterns over compounds, and (ii) this effect would be more pronounced in the L1 French translator group compared to the L1 Dutch group. Furthermore, we anticipated (iii) that L1 French translators would use more non-idiomatic multi-word expressions than native Dutch translators, and (iv) that less complex and more frequent (potential) Dutch target compounds would predict more accurate compound use in Dutch.

The results of our study largely confirm the hypotheses, but also introduce some nuances. Firstly, the study validates the overgeneralization of phrasal patterns in translation as a priming effect from French, the source language of the translation task. This phenomenon was observed across both translator groups, and, surprisingly, the effect was not stronger in the L1 French group than in the L1 Dutch group. This may indicate a strong tendency towards explicitation in translation, regardless of the native language. However, the data confirms that, overall, L1 French translators more frequently use non-idiomatic multi-word expressions. Focusing on the accuracy of the compounds used, we observed that the translations of L1 French translators tend to contain more inaccuracies in spelling, the use of linking morphemes, and grammar compared to the translations of the L1 Dutch translators. No notable differences were found between the groups regarding lexis, codeswitching, and word order. While the former inaccuracies could be considered true mistakes, as they violate the orthographic and grammatical system of Dutch, the lexical inaccuracies and codeswitching can be seen as more “creative” on the continuum, offering alternative expressions that aren't necessarily “wrong” but may sound less idiomatic due to their lower frequency, which influences their acceptability. Mistakes in word order were very infrequent, as also noted in Hendrikx & Van Goethem (2024). Finally, it was confirmed that the complexity and frequency of potential target compounds significantly influence the translators' choice to use a compound or a phrase. However, the study also emphasizes that individual variation plays a crucial role in this decision-making process. Additionally, L1 and education level contribute, but to a lesser degree. This indicates that while

general trends can be observed, individual translators' preferences and experiences heavily impact their translation choices. Notably, the fact that our regression model including individual differences can only explain 38% of the total variation implies that the majority of the variation in word formation accuracy is either random or influenced by factors that we have overlooked. The participants' translation expertise was not independently assessed, limiting insight into individual differences.

7. Conclusion

This study offers valuable insights into the translation behaviour of novice translators, confirming the influence of cross-linguistic differences and priming effects, while also highlighting the role of individual variation in translation decisions. The study underscores a continuum from 'creative solutions' - such as explication, simplification, and codeswitching - to true 'mistakes', including grammatical and spelling errors, incorrect or missing linking morphemes, and improper word order. This suggests that while novice translators may employ creative strategies to navigate translation challenges, they are also prone to making true language errors, even if they are native speakers of the target language. These findings contribute to a deeper understanding of the challenges faced by novice translators and underscore the importance of considering both linguistic and individual factors in translation research.

Building on these insights, several implications emerge for translator training and language pedagogy. Explicit instruction on Dutch compounding and linking morphemes could help reduce orthographic and grammatical errors among L1 French learners. Furthermore, raising awareness of priming effects may encourage novice translators to adopt more idiomatic solutions. Given that individual variation also played a role, training programs should also incorporate diagnostic assessments to tailor instruction to learners' cognitive profiles and prior experience. Such measures could foster both linguistic precision and strategic flexibility, equipping novice translators to navigate cross-linguistic challenges more effectively.

7 References

- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233–250). Benjamins. <https://doi.org/10.1075/z.64.15bak>
- Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355–387.
- Booij, G. (2019). *The morphology of Dutch* (2nd rev. ed.). Oxford University Press.
- De Sutter, G., & Lefer, M.-A. (2020). On the need for a new research agenda for corpus-based translation studies: a multi-methodological, multifactorial and interdisciplinary approach. *Perspectives-studies in Translation Theory and Practice*, 28(1), 1–23. <https://doi.org/10.1080/0907676x.2019.1611891>

- Don, J. (2009). IE, Germanic: Dutch. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding* (pp. 370–385). Oxford University Press.
- Ellis, N. C., & Wulff, S. (2008). Usage-based approaches to second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 75–93). Routledge. <https://doi.org/10.1075/sibil.54.03wul>
- Ferreira, V. S., & Bock, K. (2006). The functions of structural priming. *Language and Cognitive Processes*, 21(7–8), 1011–1029.
- Hüning, M., Vogl, U., Van der Wouden, T., & Verhagen, A. (2006). Nederlands tussen Duits en Engels. *Handelingen van de workshop aan de Freie Universität Berlin*. Stichting Neerlandistiek Leiden.
- Hendriks, I., & Van Goethem, K. (2024). Dutch compound constructions in additional language acquisition: A diasystematic-constructionist approach. *Constructions and Frames*, 16 (1), 64 - 99. doi:10.1075/cf.22023.hen
- Kruger, H., & Van Rooy, B. (2016). Constrained language: A multidimensional analysis of translated English and non-native indigenised varieties of English. *English World-Wide*, 37(1), 26–57.
- Lamiroy, B. (2011). Degrés de grammaticalisation à travers les langues de même famille. *Mémoires de la Société de linguistique de Paris*, 19, 167–192.
- Lanstyák, I., & Heltai, P. (2012). Universals in language contact and translation. *Across Languages and Cultures: A Multidisciplinary Journal for Translation and Interpreting Studies*, 13(1), 99–121.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227–234. <https://doi.org/10.1037/h0031564>
- Meunier, F., Hendriks, I., Bulon, A., Van Goethem, K., & Naets, H. (2023). MultINCo: Multilingual Traditional Immersion and Native Corpus. Better-documented multiliteracy practices for more refined SLA studies. *International Journal of Bilingual Education and Bilingualism*, 26(5), 572–589. <https://doi.org/10.1080/13670050.2020.1786494>
- Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39(4), 633–653.
- Prinzie, T., Suner Munoz, F., & Van Goethem, K. (2022). From crystal-clear to limpide: Translating English [Noun+Adj] compound adjectives with a figurative-intensifying noun into French. *Research in Language*, 20(4), 305–326. <https://doi.org/10.18778/1731-7533.20.4.01>
- Spina, S. (2025). Complexity and accuracy of verbal morphology in written L2 Italian: The role of proficiency and contingency. *International Journal of Learner Corpus Research*, 11(1), 114–144. <https://doi.org/10.1075/ijlcr.23032.spi>
- Tirkkonen-Condit, S. (2002). Metaphoric expressions in translation processes. *Across Languages and Cultures: A Multidisciplinary Journal for Translation and Interpreting Studies*, 3(1), 101–116. <https://doi.org/10.1556/Acr.3.2002.1.8>
- Toury, G. (1978). The nature and role of norms in translation. In L. Venuti (Ed.), *The translation studies reader* (pp. 198–211). Routledge.

- Uhrig, P. (2020). Creative intentions — The fine line between ‘creative’ and ‘wrong’. *Cognitive Semiotics*, 13(1), 1–19. <https://doi.org/10.1515/cogsem-2020-2027>
- Van Beveren, A., De Sutter, G., & Coleman, T. (2018). Questioning explicitation in translation studies: A multifactorial corpus investigation of the om-alternation in translated and original Dutch. *A paper presented at the 5th UCCTS Congress*, 12–14 September, Louvain-la-Neuve.
- Van Goethem, K. (2009). Choosing between A+N compounds and lexicalised A+N phrases: The position of French in comparison to Germanic languages. *Word Structure*, 2, 241–253.
- Van Goethem, K., & Amiot, D. (2019). Compounds and multi-word expressions in French. In B. Schlücker (Ed.), *Complex lexical units: Compounds and multi-word expressions* (pp. 127–152). De Gruyter. <https://doi.org/10.1515/9783110632446-005>
- Van Haeringen, C. B. (1956). *Nederlands tussen Duits en Engels*. Servire.
- Wen, Y., & van Heuven, W. J. B. (2017). Non-cognate translation priming in masked priming lexical decision experiments: A meta-analysis. *Psychonomic Bulletin & Review*, 24(3), 879–886.