

Linguists' Day 2024

Université de Mons

11 October 2024

Creativity and explicitation in translation: a case study in Dutch compounding

Isa Hendrikx, Wout Van Praet & Kristel Van Goethem

Outline

1. State of the art

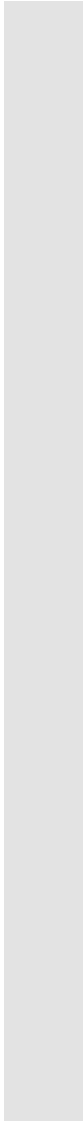
- Cross-linguistic differences: Dutch and French compound constructions
- Translation as 'constrained communication'
- Linguistic creativity

2. Aims of the study and research hypotheses

3. Corpus data and methods

4. Corpus results

5. Discussion and conclusions



1. State of the art

Cross-linguistic differences

Three important **differences** between compounding in Dutch and French:

- 1. Productivity:** Compounding is more productive in Dutch than in French (Van Goethem, 2009; Van Goethem & Amiot, 2019)
 - Du. *badkamer* vs Fr. *salle de bains* 'bathroom'
 - Du. *zoetwaterzwembad* vs Fr. *piscine d'eau douce* 'freshwater pool'
- 2. Word order:** Dutch has right-headed compounding, whereas French has left-headed compounding
 - Du. *bloemkool* vs Fr. *chou-fleur* 'cauliflower'
- 3. Demarcation compound/phrase:** The distinction between compounds and (lexicalized) phrases is more clearly established in Dutch than in French (spelling, prosody, linking elements, loss of adjectival inflection)
 - Du. *frisdrank* vs Fr. *boisson rafraîchissante* 'soft drink'

Previous study

Previous study on **learner data** in the MulTINCo corpus from 195 learners in French-speaking Belgium (Hendrikx & Van Goethem 2024):

- Compounds are more frequent in L1 Dutch than in L1 French
- French-speaking learners of Dutch overuse phrasal structures in contexts where a compound should be used (e.g., *lessen van zwembad* 'classes from swimming pool' instead of *zwemlessen* 'swimming classes')
- But they also use compounds in cases where a simplex word is appropriate (e.g., *kookman* 'cookman' instead of *kok* 'chef') and codeswitch within compounds (e.g., *verjaardagsgateau* 'birthday cake').



Present study: novice translators' use of Dutch compounds

Corpus-based Translation studies

- Baker (1993) 'Corpus Linguistics and Translation Studies. Implications and Applications'
 - **Internal developments** in the field of translation studies:
 - shift in focus from the source text to target system
 - shift from the normative concept of 'equivalence' to the descriptive concept of 'norms' (Toury 1978)
- **Translation universals**: 'features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems' (Baker 1993: 243).
 - Explicitation, simplification, normalization/conventionalization, ...

Translation as another form of 'constrained communication'

- Translation as another form of 'constrained communication' (Lanstyák & Heltai, 2012; Kruger, 2018; De Sutter & Lefer, 2019)
 - 'communication taking place under conditions where one or several of the potential limiting factors play a greater than average role' (Lanstyák and Heltai, 2012: 100),
 - i.e., writing in L2, editing, translation
 - **Priming** is pervasive in both mono- and multilingual behavior (Meyer & Schvaneveldt, 1971; Brysbaert, Van Dyck & Van de Poel, 1999)
 - Empirical process-oriented studies corroborate that even experienced translators tend to translate literally (Tirkkonen-Condit, 2002) before overruling these first translation options by more appropriate ones (Meyer & Schvaneveldt, 1971: 376)
- linguistic decisions of translators "should be analyzed as a **multifactorial phenomenon** rather than a monofactorial one (Van Beveren, De Sutter & Coleman 2018; see also De Sutter & Lefer 2016; Kruger 2018).

Linguistic creativity

- Uhrig (2020): 'the fine line between creative and wrong'
 - Criterion of “**deliberateness**”: creativity is intentional; mistakes are not
 - But **additional factors** should be considered:
 - **Frequency effects**: preemption or negative entrenchment (e.g., Sp. *vino tinto* 'coloured wine' blocks *vino rojo* 'red wine')
 - **Hearer expectations (context)**: “if the hearer thinks the competent native speaker wants to be funny, they will more readily accept the resulting utterance as creative than if the same utterance is produced by a struggling learner whose aim seems to be to produce a simple sentence that is compatible with the native norm” (Uhrig, 2020: 8)
- A **cline from creative to wrong 'constructions'** depending on multiple factors



2. Aims of the study and research hypotheses

Aims

- The **aim of the project** is fourfold:
 - i. to investigate **language-specific tendencies** towards the use of nominal compounds and their analytic equivalents
 - ii. to compare the use of Dutch compounds by **young French-Dutch translators** (whose L1 is Dutch or French)
 - iii. to reveal the **(creative) strategies** novice translators use to tackle problems caused by cross-linguistic differences
 - iv. to identify the **factors** involved in the choice between the use of compounds or phrases

Hypotheses

- i. **Overgeneralisation of phrasal patterns** in translation as a priming effect from French
 - E.g., *cours de natation* (FR) : **lessen van zwembad* vs *zwemlessen*
- ii. More **priming effects** in French-speaking translators than in Dutch speaking translators
- iii. More “**misuse**” in the non-native translators’ data
- iv. **Complexity (length) and frequency** of “potential” compound affect the translator’s choice to use a compound or a phrase



3. Corpus data and methods

Data collection and corpus sample

- French source texts from Multilingual Traditional, Immersion and Native Corpus (**MuLTINCo**) (Meunier et al., 2023)
 - Task: write an e-mail to a friend about a party or their holidays
 - 30 texts selected and edited (shortened) for the translation task
- **Translation corpus collected in 2024:**
 - Antwerp University, UGent, VUB
 - ULiège

	L1 Dutch	L1 French	Other L1
Number of participants	39	10	4
Number of nominal compounds	307	67	29
Average number of compounds per text	7.87	6.70	7.25

Other L1: Arabic (1), Spanish (2), Italian (1)

[Note: Spanish and Italian have similar word formation strategies as French]

Data annotation

Formal properties:

- **Compound length** (number of lexical components)
- **Compound structure** ([NN]_N, [VN]_N, [AN]_N, ...)

Accuracy:

- **Accuracy of the compound:** spelling mistakes, grammatical mistakes, lexical mistakes, linking morpheme, codeswitching (Fr/Eng), word order
- **Word-formation accuracy:** phrase instead of compound*, simplex instead of compound, etc.

***Checked in *Corpus Hedendaags Nederlands*: the compound had to be at least 10 times more frequent than the phrase to consider it “non-idiomatic”:**

- *mensen van mijn klas* (1 hit) vs *mijn klasgenoten* (123 hits) → non-idiomatic
- *in het centrum van de stad* (1076 hits) vs *in het stadscentrum* (956 hits) → accurate



4. Results

Results: compound length

Length compounds	L1-DU	L1-FR	L1 Other
2 components	262 (86%)	47 (70%)	21 (72%)
3 components	44 (14%)	20 (30%)	8 (28%)
4 components	1 (0%)	0	0
Total	307	67	29

$p = 0.008$

- **2 lexical components:**

- *Een **postkaart*** (L1-DU) [FR 'une carte postale']
- *Mijn **lievelingsmoment** van de dag* (L1-FR) [FR 'mon moment préféré de la journée.']

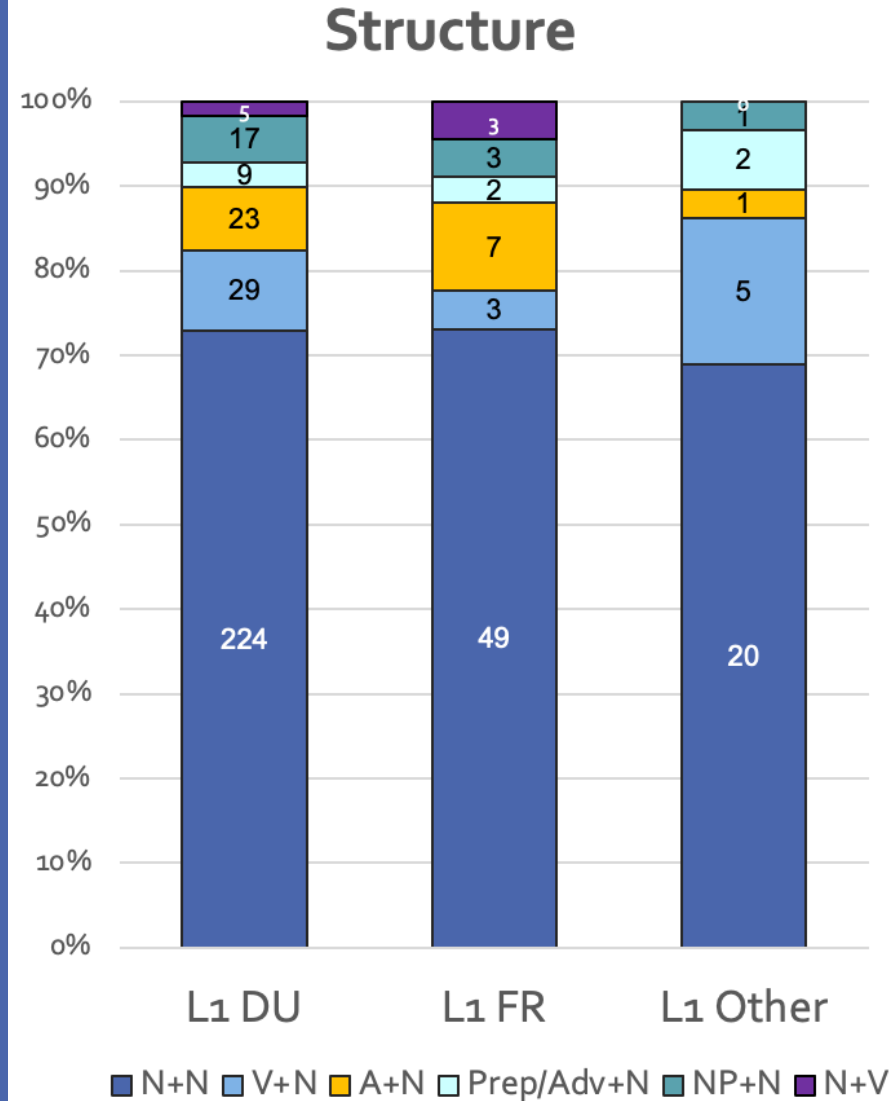
- **3 lexical components:**

- *warmwaterbronnen* (L1-DU) [FR 'sources d'eau chaude']
- *Mijn **lievelings waterbron*** (L1-FR) [FR 'Ma source préférée']

- **4 lexical components:**

- *zoogdierliefhebber* (L1-DU) [FR 'fan de mammifères']

Results: compound structure



[N+N]_N

- *Ik ben naar Guillaume's **verjaardagfeest** geweest.* (L1 FR) [FR 'fête d'anniversaire']

[V+N]_N

- *Ik heb zelfs **surflessen** kunnen volgen.* (L1 DU) [FR 'des cours de surf']

[Adj+N]_N

- *Het was niet het **hoogseizoen**.* (L1 DU) [FR 'la grande saison']

[Prep/Adv+N]_N

- *We hadden bijna altijd de zee op de **achtergrond**.* (L1 DU) [FR 'en arrière plan']

[NP+N]_N

- *Daarna stapten we naar het **driesterrenhotel**.* (L1 DU) [FR 'l'hôtel trois étoiles']

[N+V]_N

- *We gaven hem een **skydive** cadeau.* (L1 FR) [FR 'un saut en parachute']

Results: accuracy

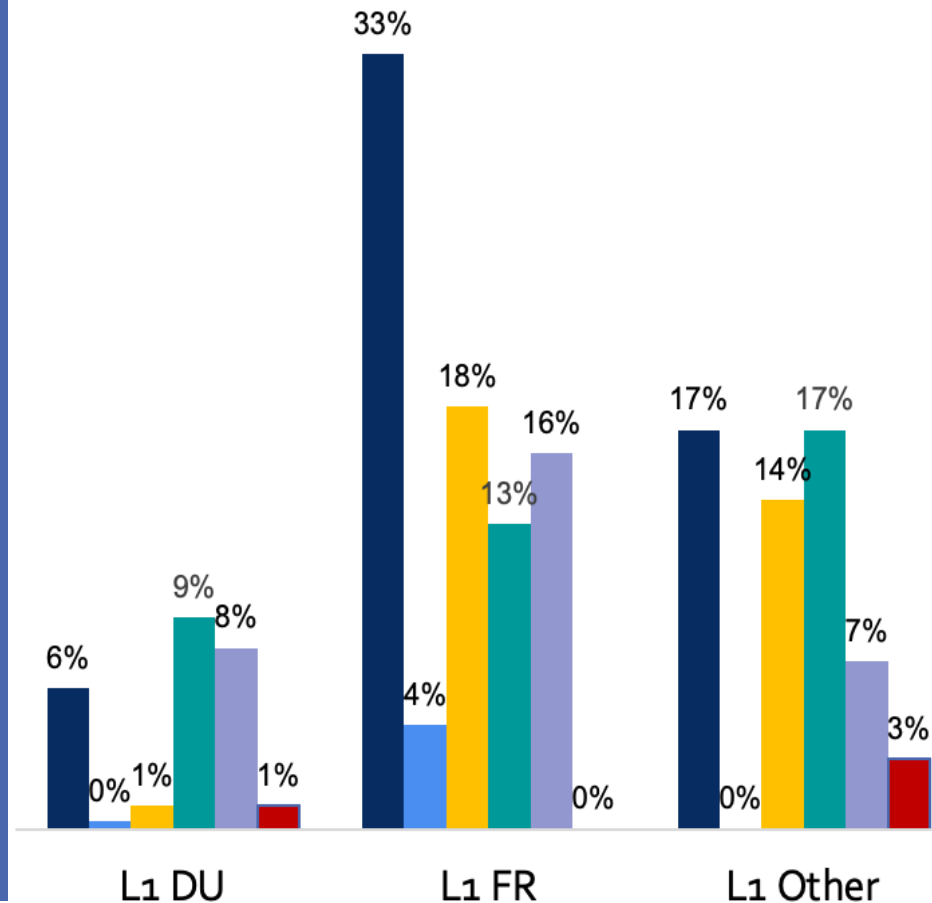
Accuracy of the compounds

Significant differences in freq. of errors in **compound** ($p < .001$) and **word formation** ($p < .001$)

	L1 DU	L1 FR	L1 Other
Inaccuracy in the compound	79 (26%)	57 (85%)	17 (59%)
Inaccuracy in the word formation	76 (25%)	42 (63%)	11 (38%)
Total number of compounds	307	67	29

Results: accuracy

Type 1 mistake (compound)



Spelling ***

- *Het was mijn **lievelings tijdstip**.* (L1 FR) [> mijn lievelingstijdstip; FR 'mon moment préféré']

Linking morpheme *

- *Een **verjaardagfeest**.* (L1 FR) [> verjaardagsfeest; FR 'une fête d'anniversaire']

Grammar ***

- ***De vuurwerk*** (L1 Other) [het vuurwerk; FR. 'le feu d'artifice']

Lexicon

- *Een **champagnetoren** van kristal* (L1 DU) [> een kristallen champagneglas; FR 'une flutte à champagne en cristal']

Codeswitching

- *Een **course relais*** (L1 FR) [> estafetteloop; FR 'des courses relais']
- *Een **firework*** (L1 FR) [> vuurwerk; FR 'un feu d'artifices']

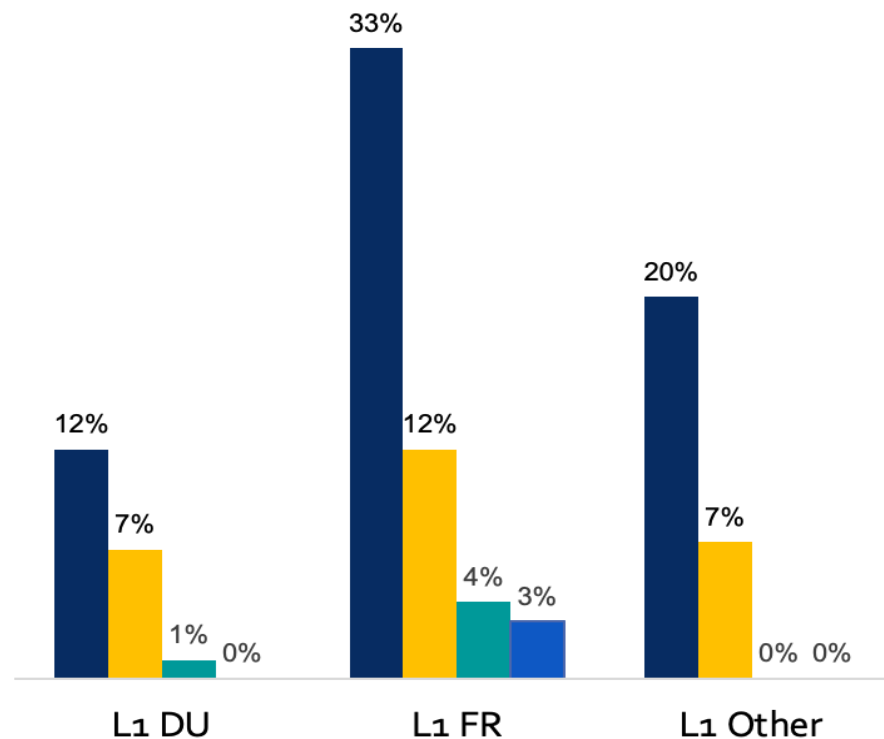
Word order

- *Een **activiteit "Jeux Olympiques"*** (L1 DU) [> een olympische spelenactiviteit, FR une activité nommée « Jeux-Olympiques »]

$p < 0.001$ *** ; $p < 0.01$ ** ; $p < 0.05$ *

Results: accuracy

Type 2 mistake (word formation)



Phrase (instead of compound) ***

(= explicitation)

- *Het **gooien van ballen met wat alcohol*** (L1 FR) [> bierpong; FR 'un beer pong']

Simplex (instead of compound)

(= simplification)

- *De **vakantie*** (L1 FR) [> *de zomervakantie*; FR 'les vacances d'été']

Compound (instead of simplex) *

(= explicitation)

- *De **feestzaal*** [> *de zaal*; FR 'la salle']

Omission (instead of compound)

- ... (L1 DU)
[> *die wolkenkrabbers*; FR 'ces gratte-ciels']

$p < 0.001$ ***; $p < 0.01$ **; $p < 0.05$ *

Results:

Influence of

- Frequency
- Length

Is there a relation between the use of a phrase instead of a compound and the frequency and length of this potential compound?

- **Compound frequency:** *des pains saucisses / brood met een worst erin (vs worstenbroodjes)* (CHN frequency: 1283 – relatively low frequency)
- **Compound length:** *au supermarché du coin / een supermarket die niet ver van ons was (vs buurtsupermarkt)* (3 lexical components – relatively complex compound)

→ Binary logistic regression to predict the occurrence of a phrase instead of a compound:

Independent variables	Exp (B)	Sig.	Nagelkerke R squares
Frequency in CHN	1	0.005**	0.19
Compound Length (2 components)	1,18E-11	0.015*	
Compound Length (3 components)	2,89E-11	1	
Compound Length (4 components)	5,35E-11	1	
Constant	1,9E+010	1	

Results:

- Frequency
- Length
- L1 Dutch

Is there a relation between the use of a phrase instead of a compound and the frequency and length of this potential compound?

Binary logistic regression to predict the occurrence of a phrase instead of a compound:

Independent variables	Exp (B)	Sig.	Nagelkerke R squares
Frequency in CHN	1	0.005**	0.20
Compound Length (2 components)	1,01E-11	0.049*	
L1 Dutch	0,48	0,015*	
Constant	1,9E+010	1	

Results: influence of

- Frequency
- Length
- Individual variation

Is there a relation between the use of a phrase instead of a compound and the frequency and length of this potential compound?

Binary logistic regression to predict the occurrence of a phrase instead of a compound:

Independent variables	Exp (B)	Sig.	Nagelkerke R squares
Compound Length (2 components)	2,93	0,002**	0.49
Frequency in CHN	1	0,003**	
Individual variation			
Constant	0,03	0,002**	

-> Not enough space to give the statistics for all ID's representing the individual variation

Results: discussion

- **Compound length and frequency** account for some variance in translators' choices ($R^2 = 0.19$).
- **L1 Dutch** adds minimal explanatory power, increasing the variance explained to $R^2 = 0.20$.
 - **Possible explanations** for the limited effect of L1 Dutch:
 - Insufficient data from L1 French speakers?
 - Other factors such as education level, exposure to the target language, or personal translation style.
- The variable with the greatest impact on translators' choices between compounds and phrases is **individual variation** ($R^2 = 0.49$).
 - This highlights the significant role of **individual differences** in translators' backgrounds and stylistic preferences.

Results: discussion

Examples of individual variation

ID13 (L1 Spanish, but living in Flanders) **produced 7 compounds in the translation:**

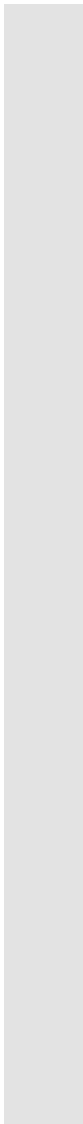
- *les joueurs de foot* - voetbalspelers
- *des autographes* - handtekeningen
- *pas un jour de pluie* - regendag
- *un empoisonnement alimentaire* - voedselvergiftiging
- *l'hôpital* - ziekenhuis
- *mes billets d'avion* - vliegtuigtickets
- *avion* - vliegtuig

ID62 (L1 Dutch, living in Flanders) **produced 7 times a phrase where a compound could have been used:**

- *2 derniers mois de vacances* - de laatste 2 maanden vakantie
- *les vastes plages de sable* - de vaste velden van zand
- *mes nombreux maillots de foot américain et de basketball* - mijn vele truitjes van Amerikaans football en basketbal
- *une des meilleures équipes de football américain* - een van de beste ploegen in het Amerikaans football
- *une équipe de football* - een ploeg van Amerikaans football
- *sa ville voisine* - de stad naast haar
- *une équipe de foot* - een ploeg van Amerikaans football



5. Conclusions



Conclusions

Hypotheses largely confirmed:

- ✓ **Overgeneralisation of phrasal patterns** in translation as a priming effect from French (= explicitation strategy)
 - In all translator groups, but stronger effect in non-native translators
- ✓ More **priming effects** in French-speaking translators than in Dutch-speaking translators
 - Double priming effect (source text + L1): e.g., more codeswitching to French
- ✓ More **“misuse”** in the non-native translators' data
 - Especially spelling and grammar mistakes in compound formation
 - Cline from 'creative solutions' (e.g., explicitation, simplification, codeswitching) to true 'mistakes' (e.g., grammatical and spelling mistakes, linking morpheme, word order)
- ✓ **Complexity and frequency** of “potential” compound affect the translator's choice to use a compound or a phrase
 - But: **individual variation** plays a more important role!

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). Philadelphia, PA/Amsterdam: Benjamins.
- Booij, G..2002. Constructional idioms, morphology, and the Dutch lexicon. *Journal of Germanic Linguistics*, 14(4), 301-329.
- Booij, G..2019. *The Morphology of Dutch*, Second, revised edition. Oxford UK: Oxford University Press.
- Brysbaert, M., G. Van Dyck & M. Van de Poel. 1999. Visual Word Recognition in Bilinguals: Evidence from Masked Phonological Priming. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 137–148.
- De Sutter, G. & M.-A. Lefer. 2019. On the need for a new research agenda for corpus-based translation studies: a multi-methodological, multifactorial and interdisciplinary approach, *Perspectives*, DOI: 10.1080/0907676X.2019.1611891
- Hendrikx, I. & Van Goethem, K. 2024. Dutch compound constructions in additional language acquisition: a diasystematic-constructionist approach. *Constructions and Frames*.
- 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*. Leiden: Stichting Neerlandistiek Leiden.
- Kruger, H. 2018. Expanding the third code: Corpus-based studies of constrained communication and language mediation. In S. Granger, M.-A. Lefer, & L. Penha-Marion (Eds.), *Book of Abstracts. Using Corpora in Contrastive and Translation Studies Conference* (5th edition). CECL Papers 1 (pp. 9–12). Louvain-la-Neuve: Centre for English Corpus Linguistics/Université catholique de Louvain.

References

- 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*, 13(1), 99–121.
- Meyer, D. E. & R. W. Schvaneveldt. 1971. Facilitation in Recognizing Pairs of Words: Evidence of a Dependence between Retrieval Operations. *Journal of Experimental Psychology*, 90, 227–234.
- Meunier, F., Hendrikx, I., Bulon, A., Van Goethem, K. & Naets, H. 2023. MULTINCo: Multilingual Traditional Immersion and Native Corpus. Better-documented multi-literacy practices for more refined SLA studies. *International Journal of Bilingual Education and Bilingualism* 26:5. 572-589.
- 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.
- Toury, G. 1978. The Nature and Role of Norms in Translation. In: L. Venuti (Ed.), *The Translation Studies Reader*. London. New York: Routledge. 198–211.
- Uhrig, Peter. 2020. Creative intentions — The fine line between ‘creative’ and ‘wrong’ . *Cognitive Semiotics* 13: 1, pp. 1-19.
- 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. Unpublished conference 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. & D. Amiot. 2019. Compounds and multi-word expressions in French. In: B. Schlücker (Ed.), *Complex lexical units: compounds and multi-word expressions*. Berlin: De Gruyter. 127-152.
- Van Haeringen, C. B. 1956. *Nederlands tussen Duits en Engels*. The Hague: Servire.



Cette photo par Auteur inconnu est soumis à la licence [CC BY](#)

Results

Type 1 mistake (compound)

- **Spelling** and **grammar** main differences between L1 DU and other L1s **lexical?**

= areas that are most strictly codified and, hence, allow for least 'creativity' in translation

- Yet, spelling errors also common with L1 DU translators, due to 'English' spelling strategies and typos

een sportauto salon (L1 DU) > een sportautosalon
FR 'un salon de voiture de sport'

de jeugdherbreg (L1 DU) > de jeugdherberg
FR 'l'auberge de jeunesse'

Results

Type 1 mistake (compound)

- **Codeswitching** also common across groups, but significantly more so with L1 FR translations

→ L1 DU codeswitch to EN:

e.g., *Er was ook een **photobooth** > fotohokje*
FR 'Il y avait aussi une cabine photo'

→ L1 FR codeswitch to FR and EN:

e.g., *De landschappen waren mooi. **Kaskaden**, volcans, etc.*
> watervallen

FR 'Les paysages étaient magnifiques. Cascades, volcans, etc.'

e.g., *Rond middernacht hadden de ouders een **firework** gepland.*
> vuurwerk

FR 'Vers minuit, les parents avaient prévu un feu d'artifice'

Results

Type 2 mistake (word formation)

- All groups more likely to use phrases or simplex words instead of compounds

e.g., FR source text: *Notre programme de la journée était vraiment très fatigant.* [DU *ons dagprogramma*]

<> L1 DU *Onze **programma's overdag** waren vermoeiend.*

- L1 FR translators use this strategy significantly more so than L1 DU

e.g., FR source text: *un pain saucisse* [DU *een worstenbroodje*]

<> L1 FR *een broodje met een worst ertussen*
'un sandwich avec une saucisse dedans'