Climate change and *Harvard students*: English noun sequences and their German and Swedish correspondences

Jenny Ström Herold[1] and Magnus Levin[1]

Abstract

This study explores English noun sequences such as *climate change*, with a common noun modifier, and *Harvard students*, with a proper noun modifier, contrasting German and Swedish. The material is provided by the Linnaeus University English–German–Swedish corpus (LEGS), a multi-directional 5-million word non-fiction corpus. The results show that the most common type of translation correspondence – regardless of translation direction – is the German and Swedish (solid) compound noun (*world war* > *Weltkrieg/världskrig*). When specifically focussing on English proper noun modifiers, it is, however, evident that these are less likely to produce compound nouns in translations, due to language-internal preferences in German and Swedish. Apart from the formal properties of correspondences, this study also takes semantics into account. We show that some types of semantic relations between the head and its modifying noun, such as Composition, which identifies the material of the head noun (*silk cloth*), are more likely to be rendered as compound nouns in German and Swedish. Amongst the non-compound correspondences in German and Swedish, post-modifying prepositional phrases are one of the more prominent alternatives (*climate signal* > *signal från* ['from'] *klimatet* [Swedish]). This result is in line with our previous findings (Ström Herold and Levin, 2019; and Levin and Ström Herold, 2024), suggesting that Swedish, more than German, favours post-modification. Amongst the notable translation effects, we observe how translators sometimes make the content more explicit through the addition of a noun, but also that the opposite applies.

[1] Linnaeus University, Institutionen för språk / Department of Languages, SE– 351 95 Växjö, Sweden.
Correspondence to: Jenny Ström Herold, *e-mail*: jenny.strom.herold@lnu.se

Corpora 2025 Vol. 20 (3): 409–438
DOI: 10.3366/cor.2025.0347

www.euppublishing.com/cor

Keywords
compound noun, explicitation, implicitation, noun sequence, semantic relations, translation

1. Introduction

Different languages favour different ways of packaging information in phrases and clauses. One such prominent area of differing information packaging principles relates to the noun phrase, where even closely related languages may have partly different preferences (see, for example, Fabricius-Hansen [1998]). This study explores one such domain: English noun sequences and their German and Swedish correspondences. We set out to investigate both whether German and Swedish compound nouns are the closest structural correspondences to English noun sequences, and also what other non-compound structures, such as paraphrases with pre- or post-modifiers, are used as translation equivalents.[2] Noun sequences and compound nouns are particularly relevant to study from the point of view of information packaging, as these are maximally condensed structures, highly dependent on the context, and pose a number of interpretive and structural challenges for readers as well as translators. What, for instance, is meant by a *potato war* or a *museum dose*? The use of English noun sequences in context is illustrated in Example 1, which is an extract from the non-fiction corpus LEGS to be considered further in Section 3.

(*1*) [...] when the *marshmallow experiments* were well under way, I vaguely remembered Homer's tales. I also wondered whether Adam and Eve might have held on to paradise longer if they had had plans [...] to help them resist snakes and *apple temptations*. I started to think about the *Bing preschoolers*: how would they deal with a powerful tempter [...]?
(LEGS, English original)

As the three cases in this example show, noun sequences contain only content words, with no function words to show the semantic relations between the lexical items, the noun head and the pre-modifying noun. Thus, in the words of Smitterberg (2021: 188), sequences are a 'sign of densification' of information and may, according to Biber *et al*. (2021 [1999]: 584), 'express a bewildering array of logical relations'. Not surprisingly, pinning down the meaning relationships in our three instances in Example 1 proves to be partly challenging. The sequence *marshmallow experiments* might be interpreted as some kind of About(ness) relation, paraphrased

[2] In this paper, we rely on orthography for the definitions: English juxtaposed nouns such as *spray bottle* are called noun sequences (see, for example, Biber *et al*. [2021 (1999): 584]) and when two or more nouns are written as one word, as in German and Swedish, they are referred to as noun compounds (e.g., *Sprühflasche* [Ge.] / *sprejflaska* [Sw.]).

as ‘the experiments involve marshmallows’. *Bing preschoolers* is likely to refer to pre-schoolers ‘affiliated’ with the Bing Nursery School. *Apple* in *apple temptations* might denote an Agent (‘apples tempt someone’), an Instrument (‘tempting someone with apples’) or, albeit far-fetched, ‘material’ (‘temptations consist of apples’).

Little is still known about English noun sequences in contrast with other languages, which applies both to contrastive and translation studies. Amongst the rare exceptions is Berg (2016, 2017), who compares English noun sequences with German solid noun compounds, considered to be its closest structural correspondence, but also other translation options. Such select studies are, however, often largely restricted because of the limited corpus material examined. Also, they typically compare only two languages, such as English and German (Berg, 2016, 2017; and Kosmata and Schlücker, 2022) or German and Dutch (De Metsenaere *et al.*, 2016). In this study, we take a broader multi-directional trilingual approach, targeting English, German and the under-studied Swedish language. As closely related Germanic languages, these have access to similar ‘noun phrase templates’, meaning that they in theory have the same options for creating noun sequences or noun compounds, but also have the same pre- and post-modifier types at their disposal. This prerequisite is important when trying to disentangle true language preferences and translation-induced changes in the domain of information packaging. Our previous studies on noun phrase features in the three languages (Ström Herold and Levin, 2019; and Levin and Ström Herold, 2024) have, for example, indicated that Swedish has a stronger preference for using post-modification than German, and it remains to be seen if this holds true also for the translations of English noun sequences.

This trilingual study continues our explorations of noun phrases in English, German and Swedish (see, for example, Levin and Ström Herold [2024]) and aims to investigate English noun sequences and their German and Swedish correspondences. We here use Johansson’s (2001) term ‘correspondence’ to refer to structures found in either German or Swedish originals and translations corresponding to English noun sequences. Our study addresses the following research questions:

- What are the German and Swedish translation correspondences of English noun sequences, and how are these distributed?
- How do the categorial status of the pre-modifiers (common or proper noun) and the semantic relationship between modifiers and heads affect the distributions of correspondences?
- What do the German and Swedish correspondences tell us about language preferences, and/or translation effects?

The paper is structured in the following way. Section 2 presents the previous work on English noun sequences in contrast, and Section 3 gives an overview of the LEGS corpus, and the methods of data retrieval and classification.

Section 4 starts with the findings of the English originals, then moves on to the German and Swedish correspondences and finally the translations into English.

2. Background

There is an abundance of monolingual studies on complex noun phrase structures, such as noun sequences, but this applies mainly to English. For English, it has been established that the noun phrase is undergoing rapid changes, the most pervasive one being the surge of noun sequences such as *air flow limitations* (e.g., Biber *et al.* [2009], Biber and Gray [2016] and Smitterberg [2021: 195]) and *the Bush administration* (Rosenbach, 2007). Biber *et al.* (2009: 184) attribute this to an 'informational explosion', which has triggered the need 'to communicate information as efficiently and economically as possible'.

As mentioned in Section 1, the 'downside' to this economical compression is the implicitness and vagueness of the meanings of noun sequences, since they lack any overt markers of the semantic relation holding between the nouns. Biber *et al.* (2021 [1999]: 582) illustrate this with the example *elephant boy*, which is highly indeterminate and could mean, for example, 'a boy who resembles an elephant', 'a boy who rides on an elephant' and 'a boy who takes care of elephants'. According to Biber and Gray (2016: 176), there has also been a recent expansion in the types of semantic relations that noun sequences may express, and Rosenbach (2007: 163) notes an increase (from 1650 to 1999) in animate nouns in the pre-modifier slot – for example, proper names (*Bush administration*) (see also Breban [2018]).

Noun sequences are also recursive in the sense that, in theory, they may be indefinitely long. However, Leech *et al.* (2009: 217), using a 1991 press corpus (F-LOB), find that the lion's share, about 85 percent, are two-part sequences and that longer three(+)-part sequences (such as *conversation district supervisors*) are comparatively rare (see also Berg [2017]). Still, Biber and Gray's (2016: 177) findings suggest that noun sequences are becoming more frequent, at least since the second half of the twentieth century (but see also Smitterberg [2021: 195] on their doubling in frequency in nineteenth-century newspaper English).

The semantic typologies proposed for noun sequences are plentiful – and some are more fine-grained than others. For example, Berg (2016) distinguishes twenty-five semantic relations, Smitterberg (2021: 206–207) decides on nineteen, Biber *et al.* (2021 [1999]: 584–5) settle on fifteen, and Bauer and Tarasova ([2013], based on Levi [1978]) work with twelve different types. The labels used for these meaning relations are just as bountiful. For instance, the category that Biber *et al.* (2021 [1999]: 584) call Composition (*glass windows*), Smitterberg (2021: 207) labels Of (*iron wheels*). For our purposes, we decided on a rather down-sized typology of fourteen semantic relations. These include relations such as Kind, Location and Purpose, and will be presented in more detail in Section 3. Studies have shown that distributions of semantic relations are text/topic-dependent: 'different genres favour specific types of relation between modifier and head' (Smitterberg, 2021: 214). This study presents results on the distributions of the semantic relations in Present-Day popular non-fiction.

As for German and Swedish, comparative large-scale studies on noun phrase packaging are scarce. Some rare exceptions include Carlsson (2004) on nominal and verbal style in German and Swedish newspapers, and Liebel (2009) on creative (*ad hoc*) compounds in Swedish–German fiction translation. These studies do not specifically delve into noun phrase packaging and clear insights into language-dependent modification preferences are not produced. However, both German and Swedish are known to be productive compounding languages (see Liebel [2009: 29]), and noun compounds (see Footnote 2) can in many ways be regarded as the closest structural correspondence to English noun sequences. Indeed, judging from Ström Herold and Levin's (2019) study on the translation of English noun sequences with proper noun pre-modifiers, noun compounds are the most frequent German and Swedish translation correspondents.

Noun compounds are – just like noun sequences – recursive: see the German stock example *Donaudampfschiffahrtsgesellschaftskapitän* ('Danube steam shipping company captain'), but bi-partite NN compounds seem to predominate. In Berg's (2017) study of German noun compounds in a fiction corpus, they amount to 90 percent, strikingly similar to Leech *et al.*'s (2009: 217) findings for English noun sequences in newspaper texts, mentioned above. German compounds are just as implicit as English noun sequences, one example being *Fischfrau* ('fish woman'), which can be ascribed many different interpretations such as 'woman who sells fish' and 'woman who resembles a fish' (Heringer, 1984: 2). Semantically, compounds are as versatile as noun sequences (Berg, 2016).

Despite this close 'correlation', findings from corpus-based translation studies show that noun compounds are not exclusively used as translation correspondences. Berg (2017), conducting a bi-directional English–German translation study of two-part noun sequences in a (narrowly sampled) fiction corpus, finds that, from English to German, 69 percent are rendered as noun compounds, but the percentage is lower in the other direction, where German noun compounds are rendered as English noun sequences in 48 percent of the cases.

Kosmata and Schlücker (2022), conducting a bi-directional English–Dutch–German translation study of sequences with proper nouns modifiers (*Wagner adoration*) in the Europarl corpus, find even lower numbers. Here, only 35 percent are rendered as noun compounds in German translations, but the percentage is slightly higher in translations into English (40 percent). Ström Herold and Levin (2019), also investigating sequences with proper nouns modifiers, find a broad spectrum of German and Swedish correspondences (fourteen types in total), such as prepositional post-modifiers (*the Norway fiasco* > *fiaskot i* [‘in’] *Norge* [Swedish translation]) and close appositions (*the Merkel government* > *die Regierung Merkel* [‘the government Merkel’] [German translation]). Compounds, however, are the most frequent correspondences – 44 percent in German and 32 percent in Swedish translations. Occasionally, translators may also omit parts of the noun sequences, in our study referred to as ‘reduction’. Kosmata and Schlücker (2022: 93–94) find cases where the common noun head has been omitted (*Tsjernobyl-centrale* [‘Chernobyl power plant’] [Dutch original] > *Chernobyl* [English translation]), but also reverse cases where the translator has left out the proper noun modifier (*Irak-Konferenz* [‘Iraq conference’] [German original] > *conference* [English translation]). Such instances of reduction seem, overall, to be exotic, but the former occurs slightly more often in their corpus, regardless of language and translation direction.

In our study, English noun sequences with both common and proper noun pre-modifiers will be investigated in terms of what kinds of German and Swedish structures are used to render these in translations (i.e., their translation correspondences), and the frequencies of these. Considering that previous studies have revealed some translational preferences due to the categorial status of the noun pre-modifier, we deem this to be one important variable. Also, the semantic relation will be considered as a factor even if studies, such as Berg (2016), have indicated that English noun sequences and German noun compounds more or less share the same inventory of meaning relations. We will also include comparisons of how noun sequences in original English texts behave as opposed to noun sequences used in translated English texts (from German and Swedish).

3. Material and method

This study is based on material from the Linnaeus University English–German–Swedish corpus (LEGS) (for further information, see Levin and Ström Herold [2024]), which consists of English, German and Swedish popular non-fiction texts from the 2000s. The non-fiction sub-genres covered in the corpus range from biographies, popular history and popular science to cookbooks and instructive self-help books. These sub-genres have perhaps more in common than can be at first surmised. Popular science often contains biographical sections, modern cookbooks are full of nutritional data, and self-help books can be more scientific than popular science. Therefore, we do not systematically investigate the potential differences between the sub-genres. The genres included are ‘popular enough’ to be translated between the three languages – a state of affairs which is otherwise largely restricted to fiction. Compared to fiction, non-fiction can be assumed to contain a higher proportion of nouns and noun sequences (Biber *et al.*, 2021 [1999]: 586; and Leech *et al.*, 2009: 216–18) and to be less focussed on authors’ creativity

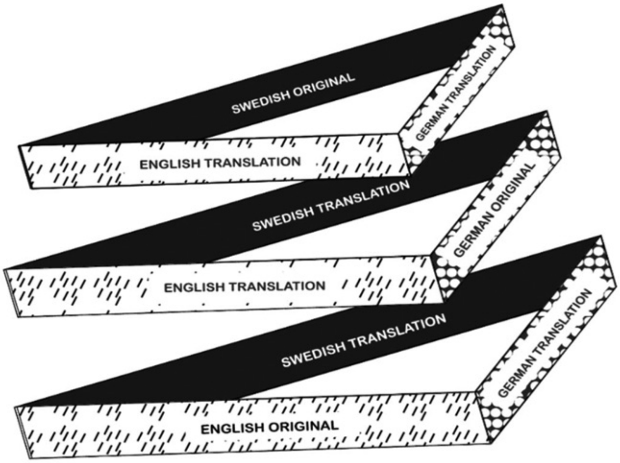

Figure 1: The structure of the Linnaeus University English-German-Swedish corpus (LEGS).

and more on conveying information, which is why we consider a non-fiction corpus to be particularly productive for our endeavour.

In the corpus, there are two translations for each original, so that, for instance, an English original has both a German and a Swedish translation. The corpus is multi-directional, which means that it is possible to compare, for example, original English texts with translated English texts from both German and Swedish, as illustrated in Figure 1.

There are obvious advantages of multi-directional translation corpora for contrastive research. Translation corpora contain 'texts which are intended to express the same meanings', according to Johansson (2007: 9), which enables us to compare like with like. Such cross-linguistic comparisons may in turn uncover language patterns that would otherwise remain hidden. However, we must also be attentive to possible source language interference – 'shining-through' effects (Hansen and Hansen-Schirra, 2012) – when interpreting the data.

This study is based on a corpus version of LEGS containing twelve English books (translated by sixteen German and fourteen Swedish professional translators), ten German books (two co-authored books with thirteen translators) and eleven Swedish books (eleven translators). Each author or translator is only represented once in the corpus, thereby minimising the influence of author- or translator-specific usages. In all, there are about 5 million words in the corpus, with the word frequencies for each sub-corpus shown in Table 1.

For retrieving the data, we searched a tree-tagged LEGS version with AntPConc (Anthony, 2017), using the search string *_N* *_NN*. This means that noun sequences (i.e., juxtaposed nouns [Smitterberg, 2021]) as

		English translations	German translations	Swedish translations
English originals	613,000	–	594,000	592,000
German originals	410,000	468,000	–	412,000
Swedish originals	471,000	516,000	462,000	–

Table 1: Word frequencies in the LEGS sub-corpora.

in *rain forest*, are included, whilst variant writings (i.e., *rainforest*, *rain-forest* [Bauer, 2019: 46]), are excluded. The searches only retrieved sequences with common noun heads, thereby excluding instances such as *The Bitcoin Foundation* where the noun head is capitalised, but including *the Bitcoin community* where it is not. The modifiers, however, may be both common and proper nouns. We randomly retrieved 1,000 instances[3] each from the English originals, and from English translations from German and from Swedish – in all, 3,000 English noun sequences with their German and Swedish correspondences.

The following formal and semantic variables were used for classifying the English noun sequences:

(*1*) noun sequence length: N N, N N N, *etc*.,
(*2*) modifier category: proper noun *versus* common noun,
(*3*) semantic relation holding between head and modifier.

The translation correspondences were classified as regards their syntactic form (e.g., compound noun or genitive phrase), but some comparisons will be made based on the semantic relations between the heads and their modifying nouns.

The semantic classification of the English noun sequences was carried out jointly by the two authors. No systematic inter-rater agreement analysis was carried out, but the two project members discussed unclear instances and categorisation issues in several meetings. As mentioned above, we decided on fourteen different relations, based on the typologies by Biber *et al*. (2021 [1999]: 584–585), Smitterberg (2021: 206–7) and Kosmata and Schlücker (2022). These are presented and exemplified in Table 2.

Classifying such large numbers of examples based on subtle semantic nuances is quite a complex undertaking. As pointed out in previous studies (Biber *et al*., 2021 [1999]: 568; Smitterberg, 2021: 207; and Biber *et al*.,

[3] From about 14,000 English original concordance lines, every tenth was selected. The precision for the English originals was 0.73. False positives included mistagged instances and tokens where two nouns happen to co-occur (e.g., *the way administrators responded*). Recall was high, reaching 0.96 (280/299 tokens sampled). Many of the false negatives were connected to the *-ing* ending (e.g., *a resting place* being classified as containing an adjective).

Semantic relation category	Example from LEGS
About	*revenge scenario*
Affiliation	*NSA whistleblowers*
Commemorative	*Molotov cocktails*
Composition	*cotton uniforms*
Extended	*potato war*
Genitive	*rat tail*
Kind	*flu viruses*
Label	*Bösendorfer pianos*
Location	*Hampshire countryside*
Object	*impulse control*
Partitive	*family member*
Purpose	*research plans*
Subject	*climate change*
Time	*spring bulb*

Table 2: The semantic relations identified in English noun sequences in LEGS.

2024: 104), there are many borderline instances. For instance, Smitterberg (2021: 207) notes that the semantic relation in *ankle bone* can be classified as Genitive, Partitive, or even Location. In our own data, we find examples such as *Southampton airport*, which we ended up classifying as Location, though Genitive could arguably also have been used. For *prisoner riot*, ultimately labelled as Subject, we could potentially have opted for Genitive or Kind. The Extended group contains metaphorical or abstract links (see Section 4.2). The relations Affiliation, Commemorative and Label, drawn from Kosmata and Schlücker's (2022) study on proper noun modifiers, were restricted to occurrences with proper noun modifiers.

For pre-modifiers ending in *-ing*, we follow Smitterberg (2021: 191). According to this approach, noun phrases which cannot be paraphrased into a subject relative clause are included in the study. When they can be rephrased, they were deemed to be too verbal and hence were excluded. Smitterberg gives a *smoking room* (*'a room that is smoking') as an example to be included and *a trifling act* ('an act that is trifling') as one to be discarded.

For statistical analysis, we used chi-squared tests for distributions. These calculations involved 2×2 contingency tables, setting the threshold at 0.05 for statistical significance. For effect size, we used Cramér's *V*. For the calculation of the agreement of the German and Swedish correspondences of English original sequences (see Section 4.2), we used Cohen's Kappa (Cohen, [1960]; for an application in a similar context, see Ström Herold and Levin [2019]).

In the following section, we turn to our corpus findings on the forms and semantics of noun sequences and corresponding structures in English and German.

4. Results

Section 4.1 starts by describing the data in the English originals. These will serve as the starting point of the analysis of the correspondences found in the German and Swedish translations and originals. Section 4.2 thus presents the findings on the structures identified in the German and Swedish translations. Section 4.3 takes the opposite approach of investigating translations into English noun sequences, and, finally, Section 4.4 focusses on cases where the translations have been made either more or less explicit due to nouns being included or excluded in the sequences or compounds.

4.1 Noun sequences in English source texts

In the LEGS material, most English source-text noun sequences are of the minimum length, two nouns: 899 of the 1,000 analysed occurrences. With nine out of ten sequences being of the minimal two-part type, the results from our non-fiction corpus are very similar to Berg's (2017) fiction and Leech *et al.*'s (2009) press corpus. Thus, the average length of noun sequences does not seem to differ much between genres. Although the vast majority of the sequences are two-part, there is also evidence of recursiveness in the LEGS data with one instance of an eight-part sequence.

Since most noun sequences in our material consist of only two nouns, our paper will mainly focus on these (but see Section 4.3 for a comparison of noun sequence lengths in English originals with translations into English). Of the 899 two-part sequences, 139 (15.5 percent) contain proper noun modifiers. So, although common nouns (e.g., *plant labels*) are the most frequent, proper nouns represent a sizeable minority in the originals. As in Kosmata and Schlücker's (2022: 88) Europarl data, ergonyms (i.e., names of objects and organisations [*CIA director*]), and toponyms (i.e., place names [*Lockerbie victims*]), account for a large majority of all proper nouns in LEGS. Anthroponyms (i.e., personal names such as *the Adamczyk family*) and praxonyms (i.e., names of events such as *Christmas tree*), are much rarer. Hence, in spite of the increase in animate noun modifiers (Rosenbach, 2007: 163), other kinds of proper nouns seem to predominate in actual usage.

Turning to the semantic relations in the two-part English source-text sequences, we see in Figure 2 the distributions across the fourteen different relations. The three most frequent ones represent half of all the tokens.

The distributions of the semantic relations indicate that, on the one hand, a large number of semantic relations are possible and occur in actual usage, and, on the other, that only three relations – Kind (*gall bladder*), Purpose (*war elephants*) and Genitive (*newspaper headlines*) – account for a large proportion. Of these, Kind is the most frequent, representing one in four tokens. This relation typically concerns sub-groups of the category denoted by the head noun, such as *stem cell* (a stem cell is a kind of cell), and these

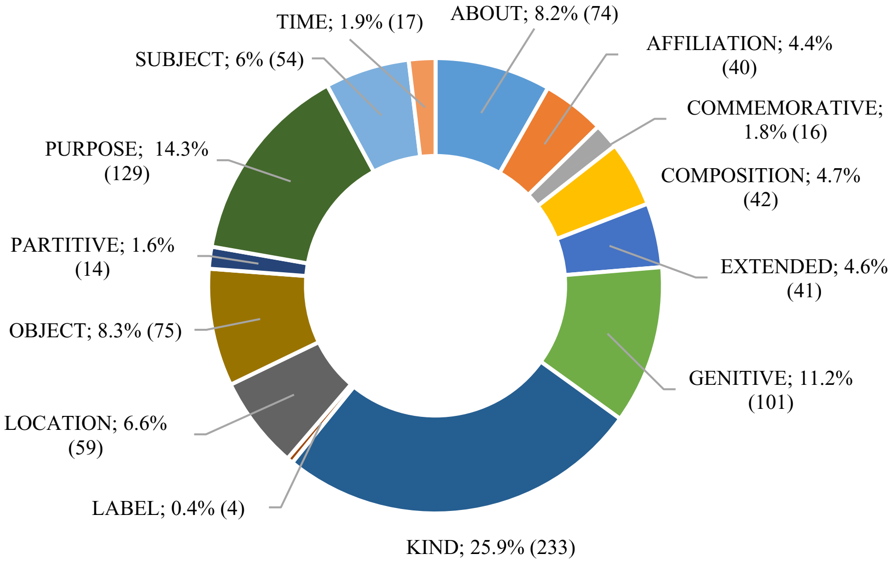

Figure 2: Semantic relations in English two-part noun sequences.

sequences often have the status as terms. As noted by Smitterberg (2021: 214), different genres may prefer different types of relations and the fact that Kind is frequent in our material may be explained by the popular science and popular history texts included in LEGS. The second most common relation expresses the Purpose of the head noun, as in *trade route* ('the purpose of the route is trade'), which in Smitterberg's (2021: 209–10) Late Modern English corpora is the most frequent one (his label being For). The third-most frequent relation is Genitive, expressed in examples such as *tongue length*, being semantically similar to a genitive determiner ('tongue's length') or an *of*-phrase ('length of the tongue').

This section has shown that the typical English original noun sequence consists of two parts which are common nouns, and that a few semantic relations account for a large number of the tokens. The next section concerns the translation correspondences of English noun sequences.

4.2 German and Swedish translations of English noun sequences

This section first analyses the German and Swedish translation correspondences, taking the semantic relations into account. Then we focus on the comparison of translations of common and proper noun modifiers, and finally we delve deeper into those instances that are not translated into compound nouns.

Correspondence type	Source-text example	Translation correspondence
(*i*) compound noun	*marshmallow experiments*	*Marshmallow-Experimente* (Ge.)
(*ii*) prepositional phrase	*Highclere gardeners*	*trädgårdsmästare på Highclere* ('gardeners at Highclere') (Sw.)
(*iii*) genitive attribute	*the nest entrance*	*boets ingång* ('nest-GEN.DEF.SG entrance') (Sw.)
(*iv*) adjective phrase	*plant matter*	*organisches Material* ('organic material') (Ge.)
(*v*) reduction (only translations from English)	*world peace*	Ø *Frieden* ('peace') (Ge.)

Table 3: The top five correspondences types of English source-text sequences.

Table 3 gives examples of the top five correspondence types[4] identified in German and Swedish originals and translations. Taken together, these five cover about 90 percent of all translations of noun sequences.

As already stated in Section 2, German and Swedish compound nouns can be argued to be the structurally closest correspondence to English noun sequences, and they are, as will be seen in Figure 3, quite frequent as well. German and Swedish compounds are either written solid, or less often, as in (*i*) in Table 3, with hyphen(s). In prepositional phrases, exemplified in (*ii*), the translators shift the pre-modifying noun into a post-modifier introduced by a preposition. The Swedish genitive attribute in (*iii*) renders the English common-case pre-modifier as a genitive (the genitive being pre-posed in Swedish and post-posed in German). In some instances, the pre-modifying English noun has an adjective as a correspondent in the target language, as in (*iv*). Reductions (*v*) occur when translators omit a noun in a noun sequence (for further information, see Section 4.4). The various correspondences of English two-part sequences will be discussed further in the following, but first we will compare the proportions of the most frequent correspondence type, the compound noun with non-compounds, in German and Swedish translations. This is illustrated in Figure 3.

The proportions of compounds are strikingly similar across the target languages in LEGS (German 69.1 percent; Swedish 67.5 percent). This also

[4] The six rarer categories, later referred to as 'Other', are: apposition, extended attribute, omission, relative clause, simple noun and verbal paraphrase.

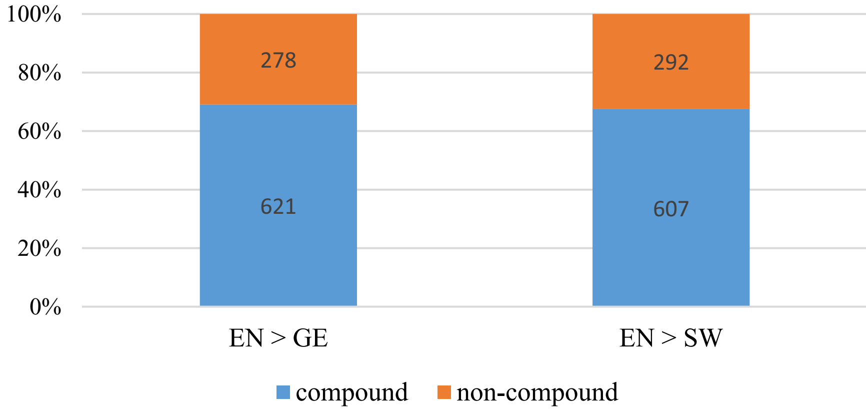

Figure 3: German and Swedish compounds and non-compounds as translation correspondences.

holds true across genres, if we compare with Berg's (2017) 69 percent compounds in German fiction translated from English. The fact that both target languages to a great extent resort to nominal compounding is perhaps not surprising as both German and Swedish are compound-productive languages (Liebel, 2009: 29). There are also translational benefits that come with these close item-for-item renderings: (*1*) they keep the semantic relation between the noun parts as vague as in the source-text noun sequences, meaning that translators avoid pinning down a specific semantic reading (which would likely be necessary with, for example, a prepositional phrase post-modifier), and (*2*) they render the semantic import as economically as noun sequences (Biber *et al.*, 2009: 184), which is obviously a guiding principle in non-fiction writing. Another contributing factor is certainly the fact that many sequences constitute terms, as mentioned in Section 4.1, in connection with the Kind relation. For example, the English terms *stem cell* and *fruit fly* have established term equivalents in German (*Stammzell* and *Taufliege*) and Swedish (*stamcell* and *bananfluga*), which happen to be compounds. For the translation of such fixed terms, the translators' freedom is generally quite restricted (see Thelen [2015: 352]).

The German and Swedish translations of noun sequences were also investigated in terms of their 'translational agreement'. We approach this as an inter-rater agreement matter, as two expert groups translated the same 899 two-part English noun sequences, and compare their choices at the structural level of the target forms. In other words, the choice of correspondence type (compound noun or non-compound) of the translated forms by the German and the Swedish translator was tested by calculating the level of agreement between the two groups of translators. The coefficient Kappa (Cohen, 1960) was used to calculate the agreement score with a confidence level of 95

Semantic relation	Kappa
About	0.35
Affiliation	0.36
Commemorative	0.37
Composition	0.68
Extended	0.37
Genitive	0.33
Kind	0.36
Label	0.00
Location	0.49
Object	0.47
Partitive	–0.13
Purpose	0.39
Subject	0.45
Time	0.27
Overall	0.41

Table 4: The agreement between German and Swedish translations for each semantic relation: compounds and non-compounds.

percent.[5] The results for the overall agreement score and the score in the separate semantic relation categories are shown in Table 4.

As shown in the table, the overall agreement score is not high (0.41), and according to Landis and Koch (1977), the agreement between the German and the Swedish translators cannot even be considered moderate (0.41 to 0.60). However, in some categories, a higher level of agreement between the two groups of translators was achieved: in the Composition category, there is substantial agreement (0.68), and the Location (0.49), Object (0.47) and Subject (0.45) categories show moderate agreement. Moving beyond the binary compound/non-compound division, we compared the agreement between German and Swedish translations for all eleven correspondence types (see Table 3 and Footnote 2). The same approach as in Table 4 was followed to identify any patterns at the structural level in the translation into German and Swedish correspondences. The results are given in Table 5.

As in Table 4, the overall agreement score is low (0.34), and this can be due to different factors. The first concerns methodological reasons: a large number of categories results in many frequent and less

[5] Cohen's Kappa is a standard metric for the calculation of the inter-annotator agreement, but alternative measures can also be considered (e.g., Gwet [2002]), and concerns about the interpretation of the results and the acceptable level of reliability (see Simaki *et al*. [2023]) can be addressed.

Semantic relation	Kappa
About	0.27
Affiliation	0.45
Commemorative	0.31
Composition	0.61
Extended	0.41
Genitive	0.27
Kind	0.27
Label	0.00
Location	0.42
Object	0.29
Partitive	–0.08
Purpose	0.42
Subject	0.30
Time	0.36
Overall	0.34

Table 5: The agreement between German and Swedish translations for each semantic relation: all eleven correspondence types.

frequent categories (non-balanced dataset), which affects the results. The second concerns language-specific reasons: the German and the Swedish use different correspondence types when it comes to the translation of the given noun sequences. So, it can be concluded that the semantic relation is not a very important feature when it comes to the structural choices of the German and Swedish translators. However, as shown in Tables 4 and 5, there are some cases where a higher agreement score is reached, and of these Composition, Location and Affiliation will be discussed below.

Composition reaches the highest level of agreement in both Table 4 and Table 5. In both the German and Swedish translations compounds are favoured, as in this example.

(2) *olive oil* > *Olivenöl* (Ge.); *olivolja* (Sw.) [“olive-oil”] (Composition)

When an entity is composed of some substance or material, a compound is a likely choice of construction, as demonstrated in Example 2. In the very rare cases where Composition yields a non-compound in the translations, the material is to be figuratively interpreted as in *iron determination*, translated into Swedish as *en vilja av stål* (‘a will of steel’) (see also Smitterberg [2021: 207]).

The Location relation also produces agreeing translation choices in German and Swedish (the second highest in Table 4), though slightly differently than discussed above, as shown in Example 3.

(3) *the Oakland airport > der Flughafen in Oakland* (Ge.); *flygplatsen i Oakland* ["the airport in Oakland"] (Location)

In contrast to Composition, translators here seem comfortable with making the semantic relations more explicit through the use of prepositions, which also relates to the categorial status of the pre-modifier as a proper noun. When the pre-modifier expressing a Location relation is a common noun, there are also some instances where the two translators opt for compounds, as seen in Example 4.

(4) *tomb frescos > Grabfresken* (Ge.); *gravmålningar* (Sw.) ["tomb frescos/paintings"] (Location)

Affiliation, which reaches the second highest level of agreement (see Table 5), shares key features with the Location category in the sense that they both often (and always) contain proper noun pre-modifiers and often lead to prepositional phrases in the German and Swedish translations.

(5) *Apple programmer > Programmierer bei Apple* (Ge.); *programmerare på Apple* (Sw.) ["programmer at Apple"] (Affiliation)

In Examples 3 and 5, we related the use of English proper nouns to the choice of German and Swedish correspondence types. The categorial status of the English modifiers are correlated in Figure 4 with the German and Swedish compound and non-compound correspondents. Three significant differences emerge: between common and proper nouns in the two target languages and between proper nouns in the two target languages.[6]

As can be surmised from Figure 4, for both German[7] and Swedish[8] translations, compounds are less common correspondences for sequences with proper noun modifiers than for sequences with common noun modifiers, compounds being used in 72.1 percent with common nouns in German and 75.3 percent in Swedish, but only in 52.5 percent and 37.4 percent with proper nouns. The two target languages therefore share a slight aversion to forming compounds with proper nouns as the left-hand elements – at least in comparison with English noun sequences. This was partly to be expected considering previous studies. For example, Kosmata and Schlücker (2022) find that only about one-third of their English sequences with proper noun

[6] In this figure and in the following, one asterisk (*) indicates a *p*-value of 0.05, two asterisks (**) indicate a chi-square of less than 0.01, and three asterisks (***) denote less than 0.001.
[7] $\chi^2 = 20.20$; $p = 0.000006980073$; Cramer's V (effect size) = 0.1532.
[8] $\chi^2 = 77.15$; $p = 1.581307 \times 10^{-18}$; Cramer's V (effect size) = 0.3002.

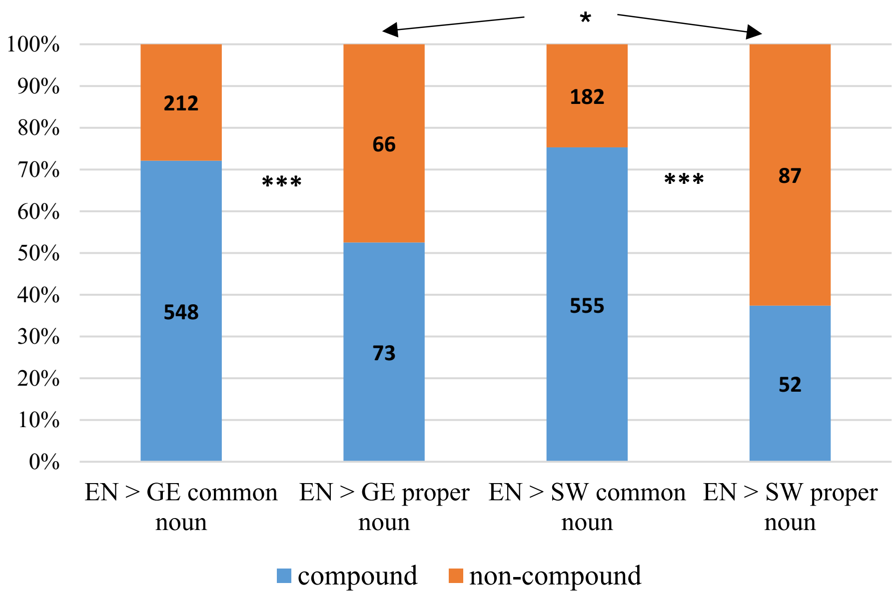

Figure 4: German and Swedish translations of English common and proper noun modifiers.

modifiers are rendered as noun compounds in German translations. Also, Zifonun's (2010: 169; 179) finding that proper name modifiers are generally less common in German noun compounds (e.g., *Kohl-Regierung*), although on the increase, than in English noun sequences supports this trend. Figure 4 also highlights a difference between the target languages: Swedish produces significantly fewer compounds than German[9] when the English sequence contains a proper noun modifier. The main reason for this lower incidence in Swedish is that German more readily creates compounds with acronyms (see Ström Herold and Levin [2019: 841]). An instance of this is given in Example 6, where the Swedish rendering instead contains a genitive attribute:

(6) *NATO bombs* > *NATO-Bomben* (Ge.); *Natos* [GEN.] *bomber* (Sw.)

It should be noted that both noun sequences and compounds have the potential to express typification. This is also addressed by Campe (2010: 210), her German example being *Brecht-Theater* ("Brecht theatre"), which can be interpreted as 'theatre typical of/in the way of Brecht', a reading not provided by the German genitive paraphrase *Brecht's Theatre*. Thus, in Example 5, both *NATO bombs* and *Nato-Bomben* may activate the typification reading – a reading which is absent from the Swedish correspondent with a genitive

[9] $\chi^2 = 5.81$; $p = 0.01589566$; Cramer's V (effect size) = 0.1519.

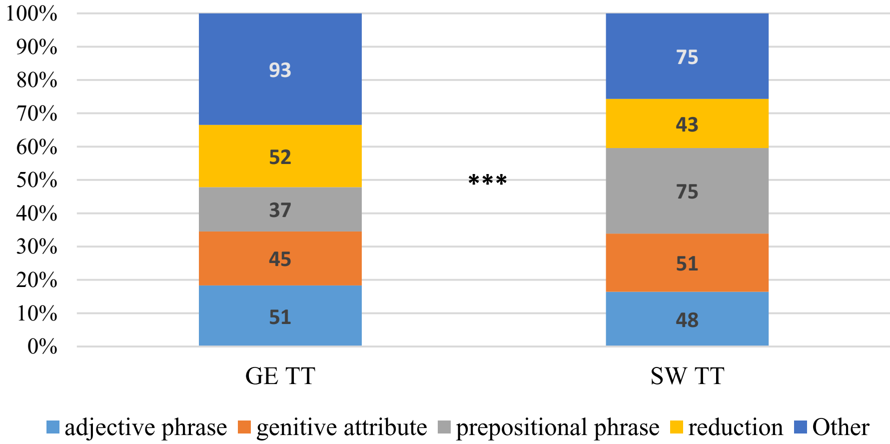

Figure 5: German and Swedish non-compound translation correspondences.

attribute: *Natos bomber*. As such, there is also a slight change in semantics in the Swedish translation.

The comparisons of the target languages have so far mostly addressed compounds, because of their high frequencies. To obtain a higher resolution of the non-compound correspondences in German and Swedish translations, we present these in Figure 5. The most prominent of these categories were introduced in Table 3. The Other category subsumes a handful of minor correspondence types, most of which have a rather loose formal relation to the source-text sequence.[10]

The target languages are similar in their preferences, overall, but compared to German, Swedish shows a stronger predilection for prepositional phrases (13.3 percent of German and 25.7 percent of Swedish non-compounds).[11] This is significant[12] for Affiliation and Location combined where more than one-quarter of all instances (20/75) are rendered as prepositional phrases in Swedish. Example 7 provides an instance of an English noun sequence of the semantic relation Affiliation rendered as a Swedish prepositional phrase, German instead resorting to a compound.

[10] In the Other category, the rare verbal paraphrases (twenty-seven German; sixteen Swedish) deserve to be mentioned: here translators go beyond the confines of the noun-phrase structure either because no suitable one can be created (*guardian lions are found at the entrance* > *bewachen Löwen die Eingänge* ['lions guard the entrances'] [Ge.]) or because a more verbal mode of expression conforms to the target-language norms (*during work hours* > *när jag jobbar* ['when I work'] [Sw.]). For the Swedish preference for verbal modes of expression, see Carlsson (2004).

[11] $\chi^2 = 13.04$; $p = 0.0003045429$; Cramer's V (effect size) = 0.1557.

[12] $\chi^2 = 18.76$; $p = 0.00001484832$; Cramer's V (effect size) = 0.1509.

(7) *Heartland colleagues* > *Heartland-Kollegen* (Ge.); *kolleger vid* ["at"] *Heartland* (Sw.)

The genitive attribute is evenly used in German and Swedish translations. In our material, this correspondence type occurs frequently with the semantic categories Genitive, Commemorative and Affiliation amounting to almost half of all instances in German translations (22/45) and two-thirds of the instances in Swedish translations (34/51). The fact that the semantic relation Genitive is associated with the genitive form is quite expected, and case of this is shown as Example 8.

(*8*) *prosecution witnesses* > *Zeugen der Anklage* ["witnesses the-GEN.SG prosecution"] (Ge.); *åklagarens vittnen* ["prosecutor-GEN.DEF. SG witnesses"] (Sw.)

The Commemorative and Affiliation relations, as previously mentioned, only apply to noun sequences with proper noun modifiers. Example 9 illustrates a noun sequence classified as expressing Affiliation:

(*9*) *former CIA director* > *CIA:s före detta direktör* ["CIA-GEN. former director"] (Sw.)

Here, the interchangeability of pre-modifying nouns and the *s*-genitive is worth mentioning. For example, Biber and Gray (2016: 171) note that the noun sequence *the Communist Party chief* could be paraphrased as *the Communist Party's chief*. It can be assumed that this is also one of the reasons why the genitive attribute is often a readily available translation option for German and Swedish translators. In our material, the genitive attribute is more common with proper noun than with common noun pre-modifiers in both target languages. For German, the genitive attribute accounts for 13.7 percent for proper nouns and only 3.4 percent for common nouns. For Swedish the difference is even greater, with proper nouns amounting to 21.6 percent and common nouns to 2.8 percent. In Kosmata and Schlücker's (2022: 93) study on the translation of English noun sequences with proper noun modifiers into German, the genitive attribute was also one of the most important competitor constructions to noun compounds.

The use of adjective phrases is also quite balanced between the two target languages. It is difficult to pinpoint any specific preferences in terms of semantic relations. However, occasionally adjectives show up in translations of noun sequences of the Extended type. Here, as previously mentioned, the relation between the nouns may be partly nebulous or metaphorical in nature. In the following example, both the German and Swedish translator resort to a more neutral descriptive adjective, clarifying the semantic import.

(*10*) *signature style* > *typischer Stil* (Ge.); *typisk stil* (Sw.) ["typical style"]

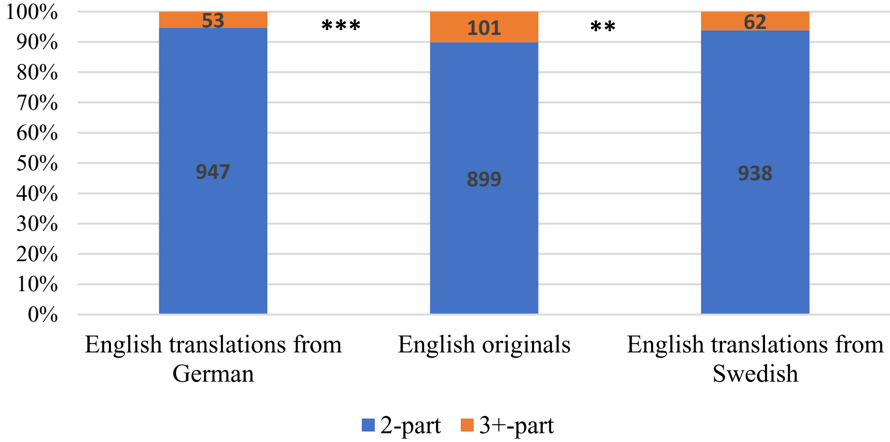

Figure 6: Noun sequence length in English originals and English translations.

Finally, the correspondence type reduction, which involves the omission of one of the nouns in the noun sequence, represents a sizeable proportion, most notably seen in German translations. Cases of reduction will be further discussed in Section 4.4.

As seen from the results, there is a strong connection between English noun sequences and German and Swedish compounds. At the same time, sequences involving proper noun modifiers are less likely to be translated into compounds. The findings also highlight the importance of taking into account the semantic relations between the nouns. In the following, we will relate these findings to those for the translations from German and Swedish into English noun sequences.

4.3 English noun sequences in translations from German and Swedish

This section discusses the effects of translation directions, exploring what type of German and Swedish structures trigger the use of English noun sequences. We begin by comparing the distributions of two-part and three+-part sequences in English originals and English translations (see Figure 6).

English originals contain significantly more three+-part sequences than both translations from German[13] and Swedish.[14] It seems reasonable to expect noun sequences, which are maximally condensed and non-transparent structures, to be more frequent in originals than in translations, which, in contrast, tend to be more elaborate and explicit (see Baker [1993]).

Perhaps unexpectedly, when translators into English use sequences of three or more nouns, the German and Swedish source-text structures do not

[13] $\chi^2 = 15.54$; $p = 0.00008074374$; it should be noted that the effect size (Cramer's V) is low: 0.0900.
[14] $\chi^2 = 9.64$; $p = 0.001898718$; here, too, Cramer's V is low: 0.0713.

generally comprise compounds consisting of three or more nouns, but of other structures (in German originals, only nineteen of fifty-three [36 percent] and, in Swedish originals, twenty of sixty-two tokens [32 percent]). Two examples of this are given below.

(11) *der Pariser Kunsthändler* ["the-DEF. Paris-ADJ art-dealer"] (Ge.)> *the Paris art dealer*

(12) *vingfrekvensen hos insekter* ["wing-frequency-DEF.SG among insects"] (Sw.)> *insect wing frequencies*

In Example 11, the German original deonymic adjective *Pariser* modifying the compound *Kunsthändler* is rendered as a three-part noun sequence. For the Swedish compound noun *vingfrekvensen* in Example 12, which is followed by the prepositional post-modifier *hos insekter*, the English translator shifts the prepositional phrase into a noun pre-modifier, creating a dense, less explicit three-part noun sequence.

Previous findings (Leech *et al.*, 2009: 217; and Berg, 2017) indicate that at least English and German rarely go beyond two-part noun compounds or sequences, but our findings seem to suggest that three+-part structures are more frequent in English. This difference between English and German is supported by Berg *et al.*'s (2012) study, but for Swedish there is a true lack of quantitative data. This is something we plan to explore in a forthcoming study.

Like long noun sequences, proper noun modifiers are also less frequent in translations into English than in original English, as seen in Figure 7.

Proper nouns as modifiers are significantly more frequent in English originals (15.5 percent) than in English translations from German (5.6 percent)[15] and Swedish (9.1 percent).[16] Of the noun sequences, most translations from Swedish and about half from German are based on compound nouns with a proper noun as their left-hand element. In the Swedish data in particular, but also to a lesser extent in the German originals, it is obvious that some source-text noun phrases are loan translations of English origin, referring to culture-specific institutions and concepts. This suggests that the translators are in fact "back-translating" these item-for-item into English, as in, for example, *Hollywoodproducenter* (Sw.) > *Hollywood producers*, *Vietnamveteran* (Sw.) > *Vietnam veteran* and *Youtube-Video* (Ge.) > *YouTube video*.

Turning to the semantic relations in English originals and English translations, Figure 8 shows that there is a high degree of similarity between the distributions.

[15] $\chi^2 = 47.11$: $p = 6.715917 \times 10^{-12}$; Cramér's $V = 0.1615$.

[16] $\chi^2 = 16.97$; $p = 0.0001206957$; for Swedish originals, the effect size (Cramér's V) is low: 0.0913.

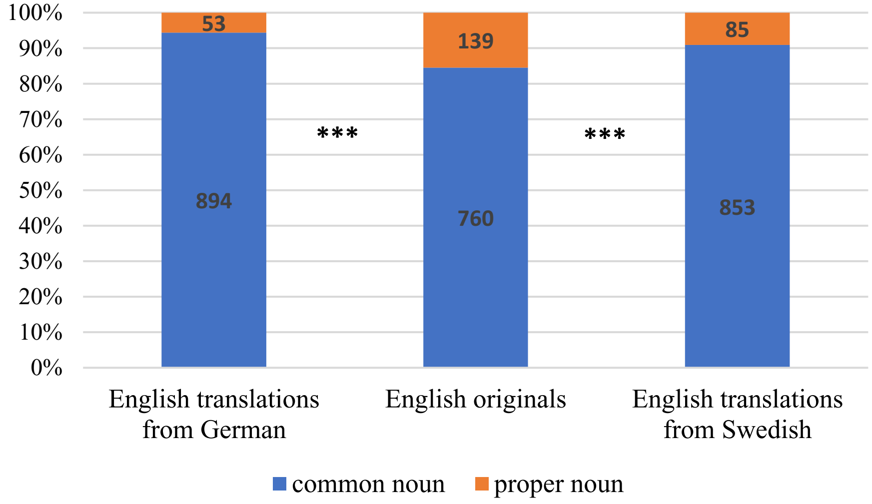

Figure 7: Proper noun modifiers in two-part sequences in English originals and English translations.

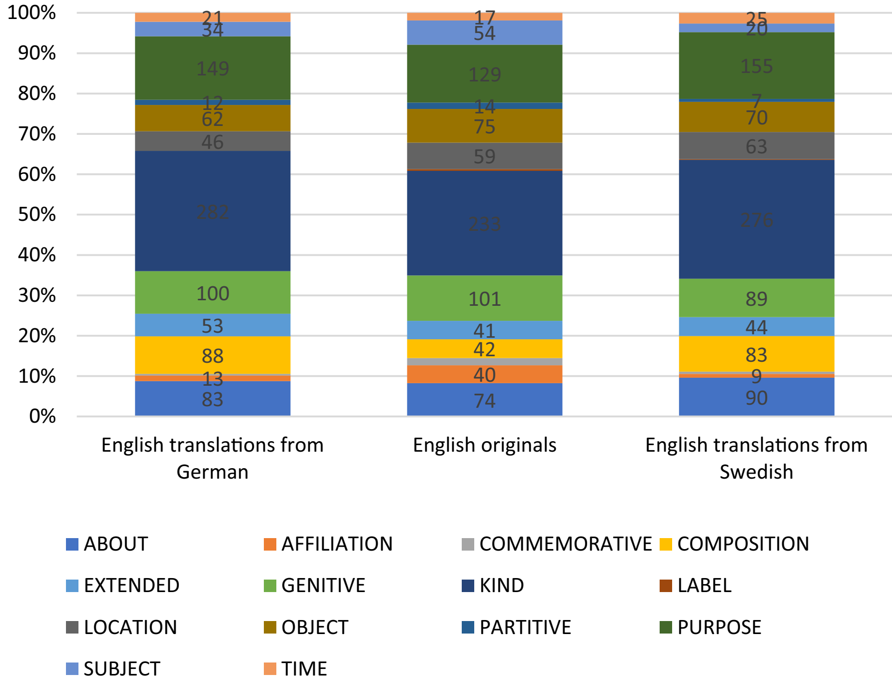

Figure 8: Semantic relations in English originals and translations into English.

In all three English sub-corpora, Kind is the most frequent relation followed by Purpose. Genitive comes in third place in the English originals and translations from German, and only just misses out on this position in the translations from Swedish. Since a large majority of the semantic relations are carried over from the source text to the translation (Kind: *Ölfirma* [Ge.] > *oil company*; Purpose: *skyddsnät* [Sw.] > *safety net*), it can be assumed that the three languages share preferences for semantic relations in the non-fiction genre (see Berg [2016]). However, there are cases where the semantic relation in the source-text sequence and the translation correspondent changes, as shown in the following examples.

(13) *foil lids* (Composition) > *Deckelfolie* ["lid-foil"] (Ge.) (Purpose)

(14) *bakplåtspapper* ["baking-tray-paper"] (Sw.) (Location) > *parchment paper* (Extended)

(15) *silk moth* (Kind) > *Seidenspinner* [lit. "silk-spinner"] (Ge.) (Object)

(16) *trafikolyckor* ["traffic-accidents"] (Sw.) (Kind) > *road accidents* (Location)

In Example 13, we see that the head and the modifier change places in the German version, thereby altering the semantic relation from what material the lids consist of to the purpose of the foil. In Example 14, Swedish highlights where the paper is placed, whilst English focusses on what it resembles: parchment. The English *silk moth* in Example 15 is a kind of moth, whilst the German *Seidenspinner* expresses an object relation. Finally, the Swedish compound in Example 16 displays a Kind relation that also occurs in English (*traffic accidents*), but the translator still chooses the Location relation, as in 'accidents that occur on the roads'. The parallel nature of the translation corpus enables the discovery of such cross-linguistic discrepancies that ultimately boil down to how languages sometimes lexicalise the world in different ways.

The two English translation sub-corpora also enable us to investigate which German and Swedish source-text structures trigger the use of English noun sequences. Figure 9 includes the source-text structures for English two-part noun sequences. Some minor categories have been subsumed under the label Other,[17] and the category extension will be discussed separately in Section 4.4. The main findings in Figure 9 will be compared with those discussed in Figures 3 and 5.

As seen in Figure 9, there is a strong preference in both languages for compound nouns being the underlying source-text structures for English noun sequences in translations: 72.8 percent (689/947) in German and 76.5

[17] Amongst these, the simple noun correspondences are noteworthy. Here, the noun sequence is rendered as a simple noun that is not the result of a reduction, so that for instance *role model* corresponds to the Swedish simplex *föredöme*.

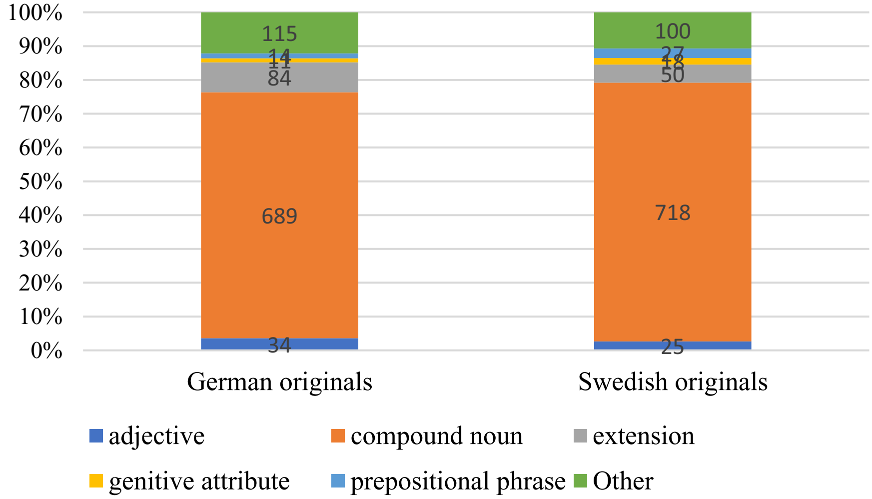

Figure 9: German and Swedish source-text structures rendered as English two-part noun sequences.

percent (718/938) in Swedish. In the English–Swedish language pair, it is more likely that an English noun sequence is translated from a Swedish compound (76.5 percent) than a noun sequence being translated into a Swedish compound noun (67.5 percent; see Figure 3).[18] For Swedish then we can establish a translation-related effect, or rather a translation-direction-related effect: it is less likely that an English noun sequence is translated into a Swedish compound than an English noun sequence being a rendition of a Swedish compound (see Figure 5). This is, at least in part, an effect of some correspondences being more likely in one translation direction than in another. For Swedish, this applies to prepositional phrases and genitive attributes. The point here is that, for instance, a Swedish prepositional post-modifier such as *kampanjer på internet* ('campaigns on the internet') is rarely translated into the less explicit noun sequence *Internet campaigns*, but that an English noun sequence, such as (*ii*) in Table 3, is more likely to be rendered as a prepositional phrase in Swedish. Translators thus more often 'unpack' condensed noun phrases than they make elaborate structures more compact (see Baker [1993]).

This section has highlighted similarities and differences between originals and translations. Some features, such as three+-part English noun sequences and proper noun modifiers, are more frequent in originals than in translations, whilst others, such as the most frequent semantic relations are the same across originals and translations. These differences suggest that there are slightly different preferences in the languages regarding the more formal

[18] $\chi^2 = 18.16$; $p = 0.00002033154$; Cramer's $V = 0.1006$.

	English to German	English to Swedish	Examples
Head noun omitted	27	20	*tank reinforcements* > *stridsvagnar* (‘tanks’) (Sw.)
Modifier noun omitted	25	23	*baby boy* > *pojke* (‘boy’) (Sw.)
Total	52 (5.8 percent)	43 (4.8 percent)	

Table 6: Reductions in translations from English.

	German to English	Swedish to English	Examples
Head noun added	46	22	*Salat* (Ge.) > *salad leaves*
Modifier noun added	38	28	*bolag* (‘companies’) (Sw.) > *record companies*
Total	84 (8.9 percent)	50 (5.3 percent)	

Table 7: Extensions in translations into English.

features (i.e., noun sequence length and categorial status of the pre-modifier), but that the core semantics is generally similar.

Reductions and extensions are the focus of the final sub-section, taking into account language preferences, translation effects and translation directions in relation to the final research question.

4.4 Translation effects: reductions and extensions

Translators into German and Swedish sometimes reduce English noun sequences to simple nouns, and translators into English may ‘extend’ simple German or Swedish source-text nouns by adding a noun. Tables 6 and 7 present the distributions of such reductions and extensions in the different sub-corpora and include illustrative examples.

Translated texts are generally characterised by explicitation (extension) rather than implicitation (reduction) (see Baker [1993]), and this is given some support in Tables 4 and 5: extensions are slightly more common than reductions in translations. Another finding is that it is not only modifiers that are omitted and added, but also heads – and in virtually equal proportions. In contrast, Kosmata and Schlücker (2022: 93–4) found for translations from English that proper noun modifiers are retained three to five times more often than common noun heads. Their high retention rate for modifiers in comparison to our data in Table 4 is probably an effect of their study having a narrower scope, restricting it to proper noun modifiers in EU texts. Language-specific preferences probably also play a role: Nesset (2018) found that as many as 46 percent of his Norwegian compound nouns corresponded to non-compound nouns in Russian originals and translations.

When translators reduce or expand noun phrases, two main – and partly contradicting – factors seem to be involved: a possible wish to reduce textual redundancy present in the source text, or the wish to be more explicit, possibly introducing some degree of redundancy in the translation. The first case is illustrated in Example 17.

(*17*) [. . .] to "occupy" EDL *Twitter* users' timelines – using several fake *Twitter accounts* > die Zeitleiste der EDL-Nutzer auf *Twitter* zu „besetzen", [. . .] über gefälschte Ø *Accounts* (Ge.)

The English original mentions *Twitter* twice within a few words, and the translator omits the second occurrence without major loss of information or clarity, considering the close proximity of the first occurrence. Example 18, taken from a book focussing on forests and trees, illustrates the reverse case.

(*18*) Und das können Sie selbst mit einem Blick nach oben in *die Kronen* nachvollziehen. (Ge.) > You can check this out for yourself simply by looking up into *the forest canopy*.

Here, the English translator chooses to verbalise the prominent text topic: forest. As seen, the German source text introduces the simple noun *Kronen* although conventionalised compounds such as *Baumkronen* or *Kronendach* could have been used. The English translator instead explicitly and somewhat pleonastically anchors the description in the forest, creating the noun sequence *forest canopy*. Other examples of extensions relate to cultural features, as in the German *im Harz > in the Harz mountains*, where a descriptive noun is added to the name of the German range to assist readers.

Finally, there is an interesting tendency for English translators to add abstract noun heads to simple nouns in the German and Swedish originals, as in the Swedish *destillation > distillation process* and the German *Wurzeln* ('roots') > *root systems*. In German contexts, compounds such as *Destillationsprozess* are referred to as explicative compounds ('Explikativkomposita', see Lehmann [2023]). The added head of such compounds, or sequences, is a hypernym to the left-hand element and are often considered pleonastic. Thus, the added noun *process* in *distillation process* does not contribute greatly to making the translation more explicit; rather, such extensions seem to follow the general English trend of frequently packing information into noun sequences – even in cases where it is not strictly necessary. The general pervasiveness of English noun sequences addressed in the introduction is thus also reflected in translations.

5. Conclusion

This study of English noun sequences in contrast with German and Swedish proved to be challenging in different ways. The inherent vagueness of noun sequences, but also of compounds, is a hurdle for translators and researchers alike. The many different proposed classifications of semantic relations in noun sequences are a testament to this (e.g., Biber *et al.* [1999 (2021)], Berg [2016] and Smitterberg [2021]), and this fact also makes it partly problematic to compare our findings to theirs.

Still, this study has found a strong link between English noun sequences, and German and Swedish compounds, corroborating previous initial findings (see, for example, Berg [2017]). Depending on the translation direction, the correspondence rates range from two-thirds to three-quarters. Thus, it is reasonable to treat these noun sequences and compounds as each other's closest cross-linguistic correspondences. The high degree of sequence–compound correspondences is at least partly due to the domain of terminology in the non-fiction genre investigated, terms often being established as condensed noun phrases. Thus, the non-fiction genre as such also restricts translators' freedom to a great extent. There is nevertheless variation found, since compounds are rarer as correspondences of sequences with proper noun modifiers, especially in Swedish (see Ström Herold and Levin [2019]).

Among the non-compounds used as translation correspondences, we find support for a previously indicated preference in Swedish (compared to German) of using post-modifying prepositional phrases (Levin and Ström Herold, 2024). Post-modifying prepositional phrases seem to be most typical for the relations Location and Affiliation which typically involve English proper noun modifiers. This indicates that there is a degree of interrelatedness between the categorial status of the modifiers and their translation correspondences. However, the semantic relations were found to have less influence on the choice of correspondence types, with some exceptions with 'moderate' agreement scores.

The use of a multi-directional translation corpus produces findings that otherwise would be hard to obtain. A major advantage with our LEGS corpus is that it allows the comparisons of German and Swedish correspondences of one and the same English structure, whereby the English original ensures that the two translations relate to 'the same' content. The material also allows the conclusion that English noun sequences tend to be longer in originals than in translations. To exemplify, the longest sequence in English originals consists of eight parts (*a Ruckus motorcycle atv car valve spring compressor tool*) whilst the translation sub-corpora only reach four items (e.g., *US child custody cases*). The longer the sequences, the more condensed and less semantically explicit they become. As such, it makes intuitive sense that the longest sequences are found in originals and not in translations. Translation effects are also seen in those cases we call extensions, where the translator adds a noun to clarify culture-specific content to new target-text readers.

Even though our study has highlighted the strong correlations between noun sequences and compounds, there are many questions remaining. One of these pertains to the lack of comparative studies focussing on the (morphological) complexity of noun sequences and compounds. Berg (2016) is the only exception for the language pair English–German, but more languages and genres need to be considered. The ongoing surge in English (at least) of noun sequences illustrates the importance of further studies, not least taking into account the degree of termhood of the items involved, since this has a bearing on how free translators can be in their choices.

Acknowledgments

This research was funded by the Swedish Research Council project 'Translating Complexity: Noun Phrases in English–German–Swedish Contrast' (reference number 2022–01739).

References

Anthony, L. 2017. AntPConc. (Version 1.2.1.) (Computer Software.) Tokyo, Japan: Waseda University. Available online at: https://www.laurenceanthony.net/software.

Baker, M. 1993. 'Corpus linguistics and translation studies: implications and applications' in M. Baker, G. Francis and E. Tognini-Bonelli (eds) Text and Technology: In Honour of John Sinclair, pp. 233–50. Amsterdam: John Benjamins.

Bauer, L. 2019. 'Compounds and multi-word expressions in English' in B. Schlücker (ed.) Complex Lexical Units: Compounds and Multi-Word Expressions, pp. 45–68. Berlin: De Gruyter.

Bauer, L. and E. Tarasova. 2013. 'The meaning link in nominal compounds', SKASE Journal of Theoretical Linguistics 10 (3), pp. 1–17.

Berg, T. 2016. 'The semantic structure of English and German compounds: same or different?', Studia Neophilologica 88 (2), pp. 148–64.

Berg, T. 2017. 'Compounding in German and English: a quantitative translation study', Languages in Contrast 17 (1), pp. 43–68.

Berg, T., S. Helmer, M. Neubauer and A. Lohmann. 2012. 'Determinants of the extent of compound use', Linguistics 50 (2), pp. 269–303.

Biber, D. and B. Gray. 2016. Grammatical Complexity in Academic English: Linguistic Change in Writing. Cambridge: Cambridge University Press.

Biber, D., J. Grieve and G. Iberri-Shea. 2009. 'Noun phrase modification' in G. Rohdenburg and J. Schlüter (eds) One Language, Two Grammars?

Differences between British and American English, pp. 182–93. Cambridge: Cambridge University Press.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 2021 (1999). Longman Grammar of Spoken and Written English. Harlow: Longman.

Biber, D., B. Szmrecsanyi, R. Reppen and T. Larsson. 2024. 'Expanding the scope of grammatical variation: towards a comprehensive account of genitive variation across registers', English Language and Linguistics 28 (1), pp. 95–133.

Breban, T. 2018. 'Proper names used as modifiers: a comprehensive functional analysis', English Language and Linguistics 22 (3), pp. 381–401.

Campe, P. 2010. 'Syntactic variation in German adnominal constructions: an application to the alternatives "genitive", "apposition" and "compound"' in A. Plewnia and A.N. Lenz (eds) Grammar between Norm and Variation, pp. 193–218. Frankfurt: Peter Lang.

Carlsson, M. 2004. Deutsch und Schwedisch im Kontrast: Zur Distribution nominaler und verbaler Ausdrucksweise in Zeitungstexten ('German and Swedish in Contrast: On the Distribution of Nominal and Verbal Modes of Expression in Newspaper Texts'). Göteborg: Acta Universitatis Gothoburgensis.

Cohen, J. 1960. 'A coefficient of agreement for nominal scales', Educational and Psychological Measurement 20 (1), pp. 37–46.

De Metsenaere, H., S. Vandepitte and M. Van de Velde. 2016. 'Dutch and German noun–noun compounds in translation', Lebende Sprachen 61 (1), pp. 175–205.

Gwet, K. 2002. 'Kappa statistic is not satisfactory for assessing the extent of agreement between raters', Statistical Methods for Inter-rater Reliability Assessment 1, pp. 1–5.

Fabricius-Hansen, C. 1998. 'Informational density and translation, with special reference to German – Norwegian – English' in S. Johansson and S. Oksefjell (eds) Corpora and Cross-linguistic Research: Theory, Method, and Case Studies, pp. 197–234. Amsterdam: Rodopi.

Hansen, S. and S. Hansen-Schirra. 2012. 'Grammatical shifts in English–German noun phrases' in S. Hansen, E. Steiner and S. Neumann (eds) Cross-Linguistic Corpora for the Study of Translations: Insights from the Language Pair English–German, pp. 133–45. Berlin: De Gruyter.

Heringer, H.J. 1984. 'Wortbildung: Sinn aus dem Chaos' ('Word formation: making sense out of chaos'), Deutsche Sprache 12, pp. 1–13.

Johansson, S. 2001. 'The German and Norwegian correspondences to the English construction type *that's what*', Linguistics 39 (3), pp. 583–605.

Johansson, S. 2007. Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies. Amsterdam: John Benjamins.

Kosmata, E. and B. Schlücker. 2022. 'Contact-induced grammatical change? The case of proper name compounding in English, German, and Dutch', Languages in Contrast 22 (1), pp. 77–113.

Landis, J.R. and G.G. Koch. 1977. ‘The measurement of observer agreement for categorical data’, Biometrics 33 (1), pp. 159–74.

Leech, G., M. Hundt, C. Mair and N. Smith. 2009. Change in Contemporary English: A Grammatical Study. Cambridge: Cambridge University Press.

Lehmann, C. 2023. ‘Explikativkomposition’ (‘Explicative compounding’) in I. Trost (ed.) Remotivierung: Von der Morphologie bis zur Pragmatik (‘Remotivation: From Morphology to Pragmatics’), pp. 123–40. Berlin: De Gruyter.

Levi, J. 1978. The Syntax and Semantics of Complex Nominals. New York: Academic Press.

Levin, M. and J. Ström Herold. 2024. ‘English complex premodifiers and their German and Swedish correspondences: the case of hyphenated premodifiers in a non-fiction corpus’, Languages in Contrast 24 (1), pp. 5–32.

Liebel, D. 2009. Tageslichtfreude und Buchstabenangst: Zu Harry Martinsons dichterischen Wortbildungen als Übersetzungsproblematik (“‘Daylight-joy” and “Letter-fear”: On Harry Martinson’s Poetic Word Formations as Translation Problem’). Umeå: UmeåUniversity.

Nesset, T. 2018. ‘When a single word is enough: Norwegian compounds and their Russian counterparts’, Slovo: Journal of Slavic Languages, Literatures and Cultures 59, pp. 61–72.

Rosenbach, A. 2007. ‘Emerging variation: determiner genitives and noun modifiers in English’, English Language and Linguistics 11 (1), pp. 143–89.

Simaki, V., E. Seitanidi and C. Paradis. 2023. ‘Evaluating stance annotation of Twitter data’, Research in Corpus Linguistics 11 (1), pp. 53–80.

Smitterberg, E. 2021. Syntactic Change in Late Modern English: Studies on Colloquialization and Densification. Cambridge: Cambridge University Press.

Ström Herold, J. and M. Levin. 2019. ‘*The Obama presidency*, *the Macintosh keyboard* and *the Norway fiasco*: English proper noun modifiers in German and Swedish contrast’, English Language and Linguistics 23 (4), pp. 827–54.

Thelen, M. 2015. ‘The interaction between terminology and translation: or where terminology and translation meet’, Trans-kom 8 (2), pp. 347–81.

Zifonun, G. 2010. ‘Von Bush administration zu Kohl-Regierung: Englische Einflüsse auf deutsche Nominalkonstruktionen?’ (‘From Bush administration to Kohl-government: English influence on German nominal constructions?’) in C. Scherer and A. Holler (eds) Strategien der Integration und Isolation nicht-nativer Einheiten und Strukturen (‘Strategies for Integration and Isolation of Non-Native Units and Structures’), pp. 165–82. Berlin: De Gruyter.