

## Observing Eurolects: The Case of Finnish

### **Abstract**

Finnish is a non-Indo-European language and a language with strong purist traditions. The comparison of the language of EU directives with that of the measures of national implementation has shown that the influence of English-language versions of EU directives on Finnish-language ones is limited. Only some rare lexical borrowings and indirect influence of the structure and style of English on the Finnish-language versions was found. Many of the features of the legal language appear alike in the directives and national legislation. Still, the writers of the directives and the writers of Finnish legislation often make different choices. The root of these differences is probably interference from the English-language versions. The claim that a Finnish Eurolect does exist still has grounds.

### 1. Introduction and background

Finnish translations of European Union directives have their basis in the *acquis* translated at the Ministry of Justice before Finland joined the EU in 1995. As a member state-to-be, Finland was obliged prior to membership to ensure the translation of the *acquis communautaire* into Finnish. For this purpose, a special unit was established at the Ministry of Justice, where Finnish translators and lawyers conducted the translation work with French as their primary source language.

The principles adopted by the translation unit laid the foundations for the translation of EU directives into Finnish, and they continue to have an impact on how EU directives are translated today, as certain set phrases of directives, for example, as well as the wording of

treaties and many key EU terms recur from one directive to another. For instance, the unit's decision to avoid loan words which disagree with the Finnish sound structure or are instantly recognized as foreign by Finnish speakers (e.g. *kertomus* for *raportti* 'report', *vuoropuhelu* for *dialogi* 'dialogue') – a position that also applies to national legal texts – has up to the present limited the ways in which EU vocabulary has been coined.

The conventions of conveying EU directives in Finnish were thus created by marrying the characteristics of the source languages with Finnish legal language. The principles of the Finnish language of law were applied whenever possible. EU directives, in fact, feature a number of conventions typical of Finnish legal language, especially those which distinguish the language of law as a genre of its own, different from other registers. Many special lexical and syntactic features of the language of law, even the punctuation, were already present in the first translations of EU directives.

The choice of Finnish variants rather than loanwords is also explained by the phonetic and morphological structure of the Finnish language. Finnish is a synthetic language: bound morphemes attached to free morphemes convey one or more grammatical meanings (such as case, tense, mood, etc.). Let us take as an example the verb form *käytettävissäni*, which incorporates the stem *käyte* of the verb *käyttää* 'to use', the passive marker *ttä*, the marker of a participle *v*, the plural marker *i*, the marker of an inessive case *ssä*, and the first-person singular possessive suffix *ni* (*käyte*+*ttä*+*v*+*i*+*ssä*+*ni*, 'at my disposal'). Another factor restricting the borrowing is the Finnish phonological structure, which does not include, for example, a combination of two consonants in an initial position.

Still, it was clear from the very beginning of Finland's EU membership that the Finnish-language versions of the European Union directives would differ from Finnish national laws both linguistically and textually. Since Finland joined the European Union, a debate has continued over the quality and impact of Finnish-language EU directives on Finnish legislation (for this discussion, see Piehl 2006). In 1998 and 2007, the Institute for the Languages of Finland conducted a survey among Finnish civil servants working on EU

affairs; the results showed that 80% of the respondents considered Finnish EU texts less readily intelligible than corresponding Finnish texts and that the intelligibility problem was rooted in, for example, excessively long and complex sentence structures, unfamiliar expressions, and unfamiliar phraseology (Piehl 2008: 2012).

The relationship between the language of EU directives and national legislation is clearly ambivalent. On the one hand, EU directives seek to represent the style of the legal text genre and use many genre features prevalent in national legal texts. On the other hand, EU directives contain features that are not part of the parameters of Finnish legal texts but rather represent the traditions and practices of those institutions that have produced these texts. Because these texts have also an impact on national laws, the drafters of national laws have to decide which linguistic features of EU directives may be considered appropriate in Finnish legislation.

At the Ministry of Justice, this falls within the remit of legislative inspection, which was first established as a body of its own in 1936. Legislative inspection was tasked with ensuring that the statutes conform to the standards and conventions of legal language. (Tynnilä 1984: 258–259) Today, almost all government bills to the Parliament are checked at the Ministry of Justice to make sure they are technically without fault and conform to the conventions of the special language required. It also follows from this that the impact of EU directives on the national language of law appears only when the applicability of phrases on legal language has been assessed as part of the legal inspection process.

The influence of EU directives on Finnish legal language has already been investigated through a small corpus (Piehl 2006): no significant changes were found in the language used in Finnish legislation, either in terms of sentence structure or in the use of grammatical words. Similar findings were reported concerning neccessive verb structures (Kanner 2011). While EU language does not seem to influence the language of national Finnish legislation, there are clear differences between the Finnish language of directives and the Finnish

language of implementing laws. This paper aims to investigate whether the larger Eurolect corpus will confirm these results and give more evidence to support the hypothesis of a Finnish Eurolect.

## 2. The data

The Finnish team joined the Eurolect project in 2014. At this point in time, the other teams had already compiled their corpora and were analysing the data. We had to catch up fast, and we could not use many of the methods used by other teams due to specific features of the Finnish language. However, we had previous experience of our own in working with corpus data. Therefore, our data and tools are not exactly the same as those used by the other language teams.

As with the rest of the Eurolect teams, the main data includes two subcorpora: the corpus of Finnish directives from the beginning of 1999 until the end of 2008 (hereafter corpus A) and the corpus of the measures of national implementation (hereafter corpus B).

Like the other teams, we collected the texts in corpus A from the old EU directives portal at <http://old.eur-lex.europa.eu/><sup>1</sup>. However, the number of documents found was 646, not 660. The most likely reason for this is that the search was performed later than the searches for the other language versions, and some of the directives may have been repealed or updated by other directives and therefore ceased to be in force.

The measures of national implementation (MNI) can be one or several of the following:

1. the adoption of new laws,
2. the amendment of existing laws,
3. the repeal of conflicting laws,

---

<sup>1</sup> Now the URL redirects to <http://eur-lex.europa.eu/>

4. no action if the practices suggested by the directive are already covered by existing national legislation, and the national laws do not contradict the directive.

In case 1, the search for MNI would yield one of a number of legislative acts implementing a directive, and these acts can be included in a corpus without any problems. In cases 3 and 4, no MNI will be found at all, and no corresponding material would be presented in corpus B. Case 2 is the most problematic one and the most used method of transposing directives. The MNI in this case would be certain articles added or changed in the existing act. For example, the Finnish Code of Judicial Procedure (Oikeudenkäymiskaari) was adopted in 1734 though it has been amended and changed several times since then. Some of its articles were subject to changes as a result of the implementation of three EU directives included in corpus A: 77/249/EEC, 2013/40/EU, and 2008/52/EC. The text of the code is quite long at 35,615 running words. Of the first six sections of the code, sections 1, 2, and 4 were adopted on 27 March 1987, section 3 was adopted in 1994, section 6 was adopted in 1734, and the title of the chapter was changed in the 1969. Section 5 was repealed in 1987.

To include the entire legislative acts affected by the directive in question therefore does not seem to be a suitable solution: the whole corpus would be skewed by very old texts. The inclusion of only the relevant articles would not cause distortions in the data, but neither would it provide enough comparable data from the national legislation. Therefore, the following strategy was adopted for compiling corpus B:

1. The corpus was compiled of whole texts, and included only acts of Parliament (fi *[eduskunta]laki*). Presidential decrees (fi *[tasavallan presidentin] asetus*), ministerial decrees (fi *[ministeriön] päätös*, *[ministeriön] asetus*), and other legislative acts not adopted by the Parliament were not included.

2. All laws adopted from 1993 onwards were included in the corpus as whole texts.<sup>2</sup>
3. From the laws that were adopted earlier, the sections written before 1993 were removed. The removed fragments were saved separately and can be used later as a comparable corpus of ‘pre-EU texts’.

The search for the measures of national implementation was performed in the FinLex databank (<http://www.finlex.fi/en>), which is a large online archive of Finnish legislative acts, which is owned by the Ministry of Justice of Finland. The codes (i.e. Celex numbers) of the directives were used as search keys.

The size of corpus A is 2,600,541 running words, consisting of 646 texts. Corpus B is smaller: it has 1,542,987 running words and includes 177 texts. Both corpora consist both of short and long texts. For example, directive 1999/47/EC is 23,326 running words in length whereas directive 1999/54/EC contains only 716 running words. Likewise, the Act on Pension Funds (Eläkesäätiölaki, 29-12-1995/1774) is 23,603 words long, but the Act on Consideration for the Energy and Environmental Impact of Vehicles in Public Procurement (Laki ajoneuvojen energia- ja ympäristövaikutusten huomioon ottamisesta julkisissa hankinnoissa, 29-12-2011/509) – despite its long name – is only 582 words long.

The study was performed by running searches on corpora A and B. However, we needed data that is more contrastive to verify the findings. For this purpose, we used the DGT Acquis Corpus, available at the website of the European Commission (<https://ec.europa.eu/jrc/en/language-technologies/dgt-acquis>). The English-Finnish data helps in checking the influence of the English language on the formation of certain terms and expressions.

---

<sup>2</sup> Finland began preparations to enter the EU and started to translate the legal documents of the European Union into Finnish in 1990 and by 1993 the influence of EU documents might have become visible

We used our own software tools for the analysis. TextHammer is a web-driven application that is currently being compiled by Mikhail Mikhailov and Juho Härme at the University of Tampere. This package of programs performs various search operations on corpora and subcorpora, including the generation of frequency lists, collocations, concordances, and so on. The texts of the corpus are lemmatised and annotated at morphological and syntactic levels with the Omorfi parser (see <https://github.com/flammie/omorfi> and <https://github.com/TurkuNLP/Finnish-dep-parser>). Grammar features can also be included in search queries.

To perform the cross-comparison of the corpora, we used the following routines:

- Frequency lists for types, lemmas, and morphological forms,
- Ngram search, i.e. the combinations of n words in a row. Studying the 500 top trigrams from corpora A and B helped to find terms, expressions, grammatical constructions, etc.
- Keyword search, i.e. comparing the frequency lists of the two corpora in order to find the words with statistically significant frequency differences. This function is available in the WordSmith Tools program (<http://www.lexically.net/wordsmith/>). However, it is also possible to compare lemmatised word lists with TextHammer, which is very important for Finnish because of the richness of its morphology.
- Collocations search. This involved checking the collocation patterns of high frequency words.
- Concordance search. Finally – yet importantly – the data obtained with the help of the previous function needed to be ‘touched’ and ‘felt’. The real examples tell more about the language and show the variability of the data. Some phenomena found on the basis of statistical data can only be explained after checking the actual contexts, through a qualitative analysis.

### 3. Lexical level

A long tradition of lexical purism is still evident in the Finnish language, and particularly in the language of law. Although the use of recognizable loanwords has increased in recent decades, they were ten years ago still less frequent in Finnish newspapers than in corresponding media of other Nordic countries, save Iceland (cf. Kváran 2007:186, Sandøy 2007:147-149). As Finnish is not an Indo-European language, Finnish phonology and grammar are very different from English, German, French, or Russian, which makes adapting foreign words difficult. Therefore, although borrowing via loanwords does take place (e.g. *skanneri* 'scanner', *skypettää* 'to skype', *innovaatio* 'innovation', etc.), the more common way of introducing new words is the use of calques (e.g. search engine – *hakukone* 'search + machine', printer – *tulostin* = agentive noun derived from *tulostaa* 'to print'). Many of loanwords are slang or professional colloquialisms, for example, *forwardoida* 'to forward' (a message), *hitti* (a musical) 'hit'. Therefore, one should not expect many loanwords to appear in the Finnish versions of EU directives, and even fewer in the Finnish laws.

The search for recent loanwords that are recognized as foreign by Finnish speakers was performed by studying frequency lists for corpora A and B. Indeed, the number of loanwords found was quite small: 284 words. The number is very modest compared to the sizes of the whole word lists: the lemmatised word list for corpus A is over 10,000 words and the size of the word list for corpus B is over 6,000 words (words with a frequency of 10 or greater were included).

Many of the loanwords with higher frequencies were borrowed a long time ago and have now been completely adopted into Finnish, even though they are still identified as loans (as



distinguished from very old loans like *Raamattu* ‘Bible’, *maanantai* ‘Monday’, *sisar* ‘sister’, which are not identified as foreign words by speakers of Finnish). These words are registered in normative dictionaries, and many of them do not have Finnish synonyms. Some examples include *teksti* ‘text’, *moottori* ‘engine’, *tekniikka* ‘technique, technology’, *tunneli* ‘tunnel’, *bakteeri* ‘bacterium’, and *signaali* ‘signal’. Among the loanwords are (parts of the) names of organisations or documents: *direktiivi* ‘directive’, *parlamentti* ‘parliament’, and *komissio* ‘commission’.

Recent loans are few, and they are usually special scientific or technical terms, for example, *testaus* ‘testing’, *homeopaattinen* ‘homeopathic’, *regeneroida* ‘to regenerate’, *koodata* ‘to encode’, *vektori* ‘vector’, and *fluoresoida* ‘to make fluorescent’.

Among the detected loanwords, there are some Anglicisms that are not technical terms, but rather stylistic markers of management or scientific discourse, for example, the adjective *validi* ‘valid’ or the noun *estimaatti* ‘estimate’. However, these words are quite rare: for example, *validi* occurs in corpus A only 15 times.

Most of the loanwords found in the corpora – with the exceptions of (parts of) proper names – are not Europeisms. Nevertheless, they demonstrate that EU texts are likely to have been translated or developed in a multilingual environment. The loanwords we found are usually more frequent in corpus A, and only about 40 loanwords from our lists are more frequent in corpus B.

As mentioned above, there exists a belief that borrowing by calques is more common for Finnish and therefore most of Europeisms should occur in the form of calques. However, it is not easy to check how common the calques are in the Eurolect data. The main reason for this is that unlike loanwords, calques are difficult to recognise. Although nowadays most calques in Finnish have an English origin, some borrowings from Latin, French, German, Swedish, and Russian can be found among earlier calques, which makes it necessary to compare words and phrases in different languages. The calques look like regular words of the Finnish language, and do not differ formally from non-borrowings, as loanwords do.

Therefore, it is difficult to detect them in a frequency list and they often have to be checked in dictionaries or corpora. In many cases, it remains unclear whether the word is a borrowing or was coined independently (e.g. *central bank* and *keskuspankki*, *postal address* and *postiosoite*). Finally, only around 270 calques have been found in the frequency list, thus the number of calques is basically the same as the number of loans. It is very likely, however, that a great number of calques passed unnoticed.

Here are some typical examples of the calques from the corpus A: *jäsenvaltio* ‘member state’, *siirtymäkausi* ‘transition period’, *sijoituspalveluyritys* ‘investment firm’, *elinkeinonharjoittaja* ‘own-account worker’, *päästökauppa* ‘emissions trading’, *verkonhaltija* ‘systems operator’.

In composite words, loan and calque methods are sometimes combined and the word contains both foreign and translated stems, for example, *likvidaatiomenettely* ‘winding-up proceedings’, *testauslaitos* ‘testing organisation’, *valmistusprosessi* ‘process of preparation’, and *talouskomissio* ‘economic commission’.

European acronyms are more common in corpus A than in corpus B, since the use of acronyms is discouraged in legislation, although some acronyms like *EU* and *EUVL* are used in legal references. In corpus A, only 121 different acronyms were found and only 34 acronyms were detected in the frequency list of corpus B. Acronyms are usually used in references and not in the texts of documents. Strangely, the acronym *EU* is more frequent in corpus B (A: 16, B: 332). Checking the contexts did not show much difference in usage: both in EU directives and Finnish laws, the acronym is used in references to directives and other documents of the European Union. Probably, the texts in corpus B always contain references to the directive being implemented by the national legislation (since all of them are implementations of the EU directives), while the directives do not necessarily contain references to other directives.

Latinisms are extremely rare due to specifics of the Finnish morphophonology. The search in parallel texts of the DGT corpus with the English-Finnish pair demonstrates that Latinisms disappear in the Finnish versions of the documents. For example, the phrases *de facto* and *de jure* are translated as *tosiasiallinen* ‘actual’ and *oikeudellinen* ‘legal’:

- (1) Lisäksi havaittiin tosiasiallinen ja oikeudellinen valikoivuus ...

Moreover, the existence of de facto and de jure selectivity was observed

(32014D0456)

Another example of avoiding direct loans from other languages is the term *acquis*, which is not used in the Finnish EU texts at all; a Finnish term, *yhteisön säännöstö* ‘community laws’ was coined to replace it. The only Latinisms found in our corpus were *in vivo* and *in vitro*, and these are not connected with EU discourse but rather with the natural sciences.

- (2) Tarkoituksena oli kehittää validi in vitro -vaihtoehto erilaisille käytössä oleville in vivo -testeille.

The in vitro <...> test was developed <...> to establish a valid in vitro alternative to the various in vivo tests in use.”

(2000/33/EC)

#### 4. Morphology

There are very few changes on the morphological level, be they EU-rooted or contact-induced phenomena (see Project Research Template in Introduction, this Volume). Some international prefixes – such as *bio-* in *biojäte* ‘biological waste’, *bioturvallisuus* ‘biological safety’, and *biopolttoaine* ‘biofuel’ – occur in our data. A few other words with loaned prefixes or prefixoids (which can also be considered compound words) were found in our data: *supernatantti*, *televiestintä*, *ultrapuhdas*. The use of international prefixes is not

frequent in Finnish because of strong lexical purism. Besides, prefixation is not typical of Finnish.

Still, some influence of EU documents can be traced here. For many frequent international prefixes, there exist Finnish equivalents, for example, *un-* = *epä-*, *sub-* = *ala-*, *super-* = *yli-*, *vice-* = *vara-*, *pseudo-* = *vale-*. Comparing the frequency of use of these prefixes in corpora A and B shows that prefixation is more common in corpus A. In corpus A, there are 19 words with the prefix *ala-* (*alakohta* ‘subsection’, *alaryhmä* ‘subgroup’, etc.) and only six in corpus B. The prefix *epä-* is also much more frequent in A, where there are 24 words with this prefix (e.g. *epäsuora* ‘indirect’, *epätavallinen* ‘unusual’, *epäsäännöllinen* ‘irregular’) compared to only nine words in corpus B.

Closed compound words (e.g. *jäsenvaltio* ‘Member State’, *asuinpaikkaoikeus* ‘right of abode’) are very common in Finnish; they are much more common than in English and even more common than in German. The data shows a significantly lower overall frequency of compound words in corpus A than in corpus B, and this can be explained by the influence of the English versions of the directives. The significance was checked with a log-likelihood test ( $p$ -value < 2.2e-16).

Structural calques, pre- or post-modifications, and other comparable forms were very unlikely to be found because many Indo-European syntactic structures are not applicable to Finnish. Changes in the use of verbal tenses and modal verbs in corpus A can be detected, however. Jenni Kettunen (2013) has found more extensive use of modal verbs in international treaties compared to bilateral treaties. Many modal verbs can be found in the keyword list generated from the lemmatised frequency list of corpus A vs. corpus B. However, it is not easy to check their frequencies because a number of Finnish modal verbs have non-modal meanings, for example, *saattaa* ‘may’/‘can’/‘to bring’/‘to place’ and *tulla* ‘must’/‘come’/‘arrive’. Therefore, only modal verbs that are generally used solely as modal verbs were checked. Most of these verbs have a positive keyness index (log-likelihood index), which means that they are more typical in corpus A, although two of them have a

negative keyness index, which means that they are more frequent in the national legislation. The results are presented in Table 1 (NB corpus B is smaller than corpus A, and therefore the lower absolute frequency in corpus B does not automatically imply that the keyness index is positive).

Table 1. Comparison of modal verb frequencies in the Eurolect data.

Verb	Corpus A	Corpus B	Log-likelihood index
<b>Positive keyness</b>			
<i>tarvita</i> ‘to need’	5,567	2,040	369.94
<i>täytyä</i> ‘must’	109	8	60.49
<i>pystyä</i> ‘to be able to’	312	94	38.03
<i>joutua</i> ‘to be forced into’	462	161	37.93
<b>Negative keyness</b>			
<i>voida</i> ‘can’	17,658	13,926	-632.29
<i>kannattaa</i> ‘it is worth of’	18	76	-79.53

Thus, the tendency to use more concrete modal verbs in the Finnish Eurolect can be observed: in corpus A, the general modal verb *voida* ‘can’ is used less compared to corpus B, while the more precise verbs *tarvita* ‘to need’ (this verb occurs in legal discourse only in negative forms, e.g. *ei tarvitse* ‘need not’), *täytyä* ‘must’, *pystyä* ‘to be able’, and *joutua* ‘to be forced into’ are used more frequently. It is worth noting that the verb *täytyä* ‘must’ is very rare in corpus B. It is also important to note that modal verbs are not the only way of expressing modality. In Finnish legal discourse, the necessive construction, *on tehtävä* ‘it

has to be done’, is very typical for expressing deontic modality (see, e.g. Kanner 2011; Rydzewska-Sięmiątkowska 2016). We discuss this construction later in this article in connection with the use of the passive voice.

Morphological tagging is available in our corpus, and this makes it possible to obtain frequency data on parts of speech and grammatical categories.

Table 2. Parts of speech statistics in the data.

<b>Parts of speech</b>	<b>Freq. A</b>	<b>% A</b>	<b>Freq. B</b>	<b>% B</b>
Adjective	259,458	9.98	109,370	7.09
Adposition	28,324	1.09	19,556	1.27
Adverb	101,071	3.88	57,264	3.71
Auxiliary	71,169	2.74	41,677	2.70
Conjunction	137,309	5.28	99,220	6.43
Noun	1,030,774	39.64	629,223	40.80
Pronoun	141,419	5.44	80,305	5.20
Proper noun	109,554	4.21	16,295	1.06
Subordinating conjunction	43,396	1.67	28,190	1.83
Verb	390,756	15.03	270,024	17.50

The parts of speech statistics (see Table 2) show that there are fewer verbs and more nominals (nouns + proper nouns + adjectives + pronouns) in corpus A. This means that the sentences in corpus A should be longer than in corpus B, which is confirmed by other statistics (see Section 5 of this article). There are also fewer conjunctions, which is most probably a signal of longer sentences: more words fit between conjunctions in corpus A than in corpus B. Another possible reason is that there are more lists and more conjunctionless constructions in the texts of corpus A.

Studying the morphology by grammar categories reveals more interesting tendencies. In the texts of corpus B, the use of moods other than the indicative is rare. The use of moods in corpus A is different: even a few occurrences of the potential mood were found. This mood, with the meaning of high probability, is highly infrequent because of its redundancy.

(3) Tuoreiden tietojen perusteella SVC-tauti ei täyttäne kaikkia edellytyksiä, jotta se voitaisiin sisällyttää direktiivin 2006/88/EC liitteessä IV olevassa II osassa esitettyyn muiden kuin eksoottisten tautien luetteloon.

On the basis of the recent information received, SVC appears not to meet all the criteria for inclusion in the list of non-exotic diseases in Part II of Annex IV to Directive 2006/88/EC.

(2008/53/EC)

The conditional is quite frequent in corpus A, with over 5% of all forms, while in corpus B it constitutes less than one percent. This mood is not used to convey legal rules in Finnish national legislation. In association with such rules, the conditional mood usually indicates an event that is possible but has not occurred. The usage in directives is otherwise similar (e.g. Kanner 2011: 55–57), but the conditional mood occurs frequently in preambles and annexes as the translation for expressions containing *should*. In national legislation, reasons for the government bill are given in a separate text and the indicative mood is also dominant in annexes.

Table 3. Moods in the Finnish Eurolect.

Mood	Freq. A	% A	Freq. B	% B
Conditional	9,833	5.18	1,044	0.84
Imperative	1,243	0.66	327	0.26
Indicative	178,583	94.15	122,605	98.89
Potential	19	0.01	0	0

The absence of a reference to the actor is a distinctive Finnish feature made possible by a range of grammatical means (see, e.g. Hakulinen 2001: 1987), one of them being the passive voice. The strong position the passive voice occupies in the Finnish language can also be seen in the findings from the corpora. Passive forms are not over-represented in corpus A; although they seem to be more frequent, the difference between corpora A and B is not very large. The finding is statistically significant (log-likelihood test,  $p$ -value  $< 2.2e-16$ ).

Clauses in the passive voice are extremely common in Finnish legal and administrative language: predicate verbs in the passive voice are twice as common in the corpus of the Institute for the Languages of Finland on legislative texts as they are in a corpus of general language (Piehl 2010: 169). In Finnish, passive predicates can only be created of verbs implying a human actor. The Finnish passive voice is agentless; rather, the actor in a passive clause is implicit – the actor is either anybody in general or deducible from the context. The obligations of a statute typically apply to everyone, which makes the passive voice a useful means of expression in legal language.

Data with specification by the finite/infinitive/participle shows that the finite passives and infinitive passives have almost the same frequencies, and the main difference is in the use of participles. The finding is statistically significant (log-likelihood test,  $p$ -value  $< 2.2e-16$ ).

Table 4. The active and passive in the Finnish Eurolect.

<b>Verb form</b>	<b>Freq. A</b>	<b>% A</b>	<b>Freq. B</b>	<b>% B</b>
Finite Active	148,689	33.11	96,879	32.12
Finite Passive	47,231	10.52	32,476	10.77
Infinitive Active	53,763	11.97	41,626	13.8
Infinitive Passive	4,680	1.04	2,908	0.96
Participle Active	64,017	14.26	45,881	15.21
Participle Passive	130,661	29.1	81,816	27.13



The Ngrams search demonstrates that in corpus A, constructions with the present participle passive – such as *on varmistettava* ‘have to be guaranteed’, *on säädettävä* ‘have to be regulated’, and *on toimitettava* ‘have to be delivered’ – are very typical. The top ten trigrams in corpus A include seven present participle passives. The list of trigrams contains 16 patterns with the participle passive. In the trigram list for corpus B, the first trigram containing the present participle passive (*on otettava huomioon* ‘has to be considered’) occupies only 27<sup>th</sup> place. The total number of trigrams containing passive participles is also much smaller: there are nine patterns in the corpus. The construction with the verb *olla*, ‘to be’, in the third-person singular plus the past participle is called the *nessiivirakenne* and is used to express deontic modality. The form is so common in corpus A because it is typically used as an equivalent for English constructions containing *shall*. The difference between the corpora can also be explained by the growing tendency in national legislation to pose obligations without any grammatical markers of deontic modality (Kanner 2011: 37).

(4) Osapuolten on erityisesti varmistettava, että vaihdettavien tietojen paikkansapitävyys tarkastetaan säännöllisesti.

The Parties shall in particular ensure that the accuracy of data exchanged is regularly reviewed.

(22014A0604(01))

Table 5. Case forms in the Finnish Eurolect.

Case	Freq. A	% A	Freq. B	% B
Abessive	2,481	0.15	3,105	0.33
Ablative	12,654	0.77	7,918	0.83

Accusative <sup>3</sup>	51	0	206	0.02
Adessive	49,163	3	32,700	3.45
Allative	34,764	2.12	24,224	2.55
Comitative	196	0.01	474	0.05
<b>Elative</b>	<b>68,766</b>	<b>4.2</b>	<b>72,504</b>	<b>7.64</b>
Essive	30,734	1.88	13,817	1.46
<b>Genitive</b>	<b>456,076</b>	<b>27.83</b>	<b>277,619</b>	<b>29.26</b>
Illative	77,724	4.74	56,603	5.97
<b>Inessive</b>	<b>107,219</b>	<b>6.54</b>	<b>77,655</b>	<b>8.18</b>
Instructive	10,939	0.67	6,863	0.72
<b>Nominative</b>	<b>584,511</b>	<b>35.67</b>	<b>243,425</b>	<b>25.65</b>
<b>Partitive</b>	<b>184,808</b>	<b>11.28</b>	<b>118,032</b>	<b>12.44</b>
Translative	18,680	1.14	13,770	1.45

Finnish has very rich declension, with 15 cases. Despite the fact that some of the cases – such as the instructive case (by means of X, *paljain silmin* ‘with naked eyes’) or comitative case (with X, *perheineen* ‘with his family’) – are very rare and are used mostly in idiomatic expressions, the number of actively used cases is still greater than in the Indo-European languages, even in those with rich inflection like Slavonic or the Baltic languages. The statistics on the use of cases in the research data shows an interesting tendency. Corpus A shows a dramatic increase in use of the nominative case: 35.67% in corpus A compared to 25.65% in corpus B. At the same time, a decrease in use of all other cases is detected. A change over 1% is observed in the use of the most frequent cases: the genitive, partitive, inessive (‘in X’), and elative (‘from X’).

<sup>3</sup> Only for pronouns. The accusative case for nouns and adjectives is homonymous to genitive and is not differentiated by parsers.

No evident explanation can be found for this phenomenon. The reason may be the abundance of lists and tables in directives as such elements are fewer in number in national legislation.

## 5. Syntactic level

More differences seem to exist on the syntactic level. Some contact-induced differences can be discerned in the word order between the corpora, although in rather marginal cases. The neutral word order typical of the Finnish language in the active voice is SVO or SVX, where X may be not only an object but also some other complement of a verb, such as an adverbial or a predicative. Words can, however, appear in almost any order within a sentence, because the constituents can be identified from their case. In non-neutral word order (VSO or OVS), the meaning remains the same, but the tone will change, becoming argumentative, for example, in sentences without context. In practice, word order is largely dependent on the information structure of the sentence.

Qualitative analysis of the corpora suggests that the most common types of word order are identical in the two corpora. The three most common categories are SVO; the neutral order of a passive clause with a transitive verb, OVA; and another order of a passive clause, AVO. After these three most common categories, the types of word order appear to differ between the corpora, but conclusions are hard to draw from the small numbers involved in the qualitative analysis. Qualitative scrutiny nevertheless implies that corpus A contains more verb-initial clauses, while there are more cases of verb-final word order in corpus B. Corpus queries show that national legislation has few verb-initial clauses and that this is a distinguishing feature in the language use of the directives and national legislation. Verb-initial passive clauses are used as standard phrases in particular when amending directives,

as in *Muutetaan x artikla* ‘Amending article x’ or *Korvataan x artiklan y kohta* ‘In Article x, paragraph y is replaced’. This formulation does not occur in national legislation to indicate what is to be amended. The corpora include very few verb-initial clauses in the active voice. First-person and second-person verb forms commonly begin a clause in Finnish, but third-person verb forms, which typically occur in legal language, are found in an initial position only in some types of zero-subject clause.

What the corpus-based findings do show is that clauses ending with a verb are indeed more common in national legislation. They are, however, reasonably common also in the directives. An especially common type of verb-final clause is the passive subordinate clause that refers to another statute or provision. Ending a clause with a verb is a feature of Finnish that distinguishes legal language from general language use. For example, the subordinate clause in the following sentence *Väliaikainen kiello on voimassa, kunnes asia lopullisesti ratkaistaan* ‘A temporary injunction will stay in force until the matter is ultimately settled’ would probably be phrased as *kunnes asia ratkaistaan lopullisesti* in general Finnish use. This type is found especially in subordinate clauses. A verb in the final position of a sentence does not as a rule represent the most emphatic constituent of a clause even in legal language, but it is apparently considered an integral stylistic feature, as drafters of legislation have shown an unwillingness to abandon it regardless of the needs of the information structure (Kankaanpää et al. 2012: 404).

The use of pronouns is another issue that can reflect the syntactic structure of texts. The frequencies of pronouns in Table 2 show little difference on the level of parts of speech. However, the more detailed statistics with pronoun types (Table 6) show an obvious difference in frequencies of personal, interrogative, and relative pronouns. The decrease in frequencies of personal pronouns in corpus A can be explained by the use of passives. The pronoun *hän* ‘he’/‘she’ appears less frequently in legal language than in regular language use, but there is also a marked variation in its use between the corpora. This may be

because the directives aim more to obligate the member states, while the provisions of national law are targeted directly at the citizens of the member states.

Table 6. Statistics on pronouns.

<b>Pronoun type</b>	<b>Corpus A</b>	<b>% A</b>	<b>Corpus B</b>	<b>% B</b>
Demonstrative	69,872	49.49	35,798	44.66
Indefinite	26,397	18.70	15,503	19.34
Interrogative	1,170	0.83	3,657	4.56
Personal	2,240	1.59	4,941	6.16
Reciprocal	1,334	0.94	1,393	1.74
Relative	40,182	28.46	18,871	23.54

Finnish abounds with syntactic calques. The 600 years of shared history between Finland and Sweden means that many of these loans come from Swedish, which in turn has been influenced by German and Latin. Since the latter half of the twentieth century, influences on the Finnish language have mainly come from English. It is therefore these English-like structures that have made their presence felt in the Finnish-language versions of EU directives, but this variation is mainly intra-linguistic.

The directives make frequent use of the expression ‘Member States shall/should/may ensure that...’, which manifests the relation between the directives and national legislation. Such a function does not exist in national legislation. It is therefore understandable that the

standard Finnish equivalent chosen for the verb, *varmistaa*, is significantly more common in corpus A. On occasion, one also finds *huolehtia* ‘take care that’ as the equivalent, which appears more frequently in corpus B. In both corpora, these verbs occur in structures that are typically Finnish and are used to place an obligation on somebody. The actors differ to a certain extent between the corpora. In corpus A, the subject of the structure *on varmistettava* ‘shall ensure’ is commonly *jäsenvaltio* ‘Member State’, and on occasion also *komissio* ‘the Commission’. The actors in corpus B are various authorities or other national bodies.

An obvious syntactic loan, albeit with a long history in the Finnish language, is the structure *niin* + adverbial + *kuin* (on) *mahdollista* ‘as + adverbial + as (is) possible’, which is twice more common in corpus A than in corpus B. An example of this construction is *Yrityksen on niin pian kuin mahdollista tiedotettava yleisölle kaikesta toimintaansa koskevasta merkittävästä kehityksestä* – ‘The undertaking must inform the public as soon as possible of any major new developments in its sphere of activity –’ (79/279/EEC). In the place of an elliptical or non-elliptical clause, (on) *mahdollista* may also be another subordinate clause, such as *niin* + adverbial + *kuin* + subordinate clause, as in *Lupakirja pysyy voimassa niin kauan kuin 16 artiklan 1 kohdassa säädetyt ehdot täyttyvät* ‘The licence shall remain valid, provided that the conditions in Article 16(1) remain fulfilled’ (2007/59/EC). The adverbial position can be filled with a range of expressions, including *usein* ‘often’, *pitkälle* ‘far’, *vähän* ‘little’, *pian* ‘soon’, *suuressa määrin* ‘to a large extent’, *myöhäisessä vaiheessa* ‘at a late stage’, and so on. The prevalence of this construction likely stems from the fact that the structure ‘as + adverbial + as possible (or as + subordinate clause)’ is commonly used in the source texts.

Another syntactic calque that contains the adverb *niin* ‘as’ is the coordinating structure *niin x kuin y* ‘x as well as y’/‘both x and y’. This rhetorical device, typical of the Eurolect, does not occur in national legislation at all, although it is found in other Finnish text genres. An

example from a directive is provided by the following sentence, which uses the structure to equate groups of various actors:

- (5) Näihin järjestelyihin voi kuulua kuhunkin toimivaltaiseen viranomaiseen perustettavia neuvoa-antavia komiteoita, joissa olisi oltava edustettuina mahdollisimman hyvin eri markkinatoimijat, niin liikkeeseenlaskijat, rahoituspalvelujen tarjoajat kuin kuluttajat. These arrangements may include consultative committees within each competent authority, the membership of which should reflect as far as possible the diversity of market participants, be they issuers, providers of financial services or consumers.
- (2003/6/EC)

A comparison of sentence complexity reveals intralinguistic syntactical variation between the two corpora. One frequently used marker of complexity is the average length of sentences. It is also considered an important feature for the text type, genre, or certain author. On account of a previous study (Piehl, 2006), we expected to find differences in sentence lengths between corpus A and corpus B.

Calculating sentence lengths is not as easy as might be expected. Punctuation and division in paragraphs in legal texts deviate from general use, especially in the case of directives. Therefore, calculations based on punctuation would not be very precise. Instead, we used morphological tagging to evaluate the average length of clauses. The average clause length in a corpus is the ratio of the corpus size to the number of finite verb forms in the corpus. The mean clause length in corpus A is 13.27 words, while for corpus B it is 11.93 words. The findings are statistically significant (log-likelihood test,  $p$ -value  $< 2.2e-16$ ). This confirms the previous findings, which also indicate that sentences in the directives are clearly longer than sentences in national legislation.

The complexity of sentence structure is also indicated by the number of clauses and phrases. There are both similarities and differences in how the two corpora convey subordination and coordination. As already mentioned, both kinds of conjunctions are more

frequent in corpus B (see Table 2). In both corpora, a few conjunctions account for most of the occurrences. There are, however, differences between the corpora in which conjunctions are the most common and how large the differences are between the most common conjunctions.

Coordination is obviously more common than subordination in the two corpora: both clauses and phrases can be coordinated, whereas only clauses can be subordinate to one another. Abundant coordination is a typical feature of legal language, which tends to present lists of equal or alternative characteristics.

Table 7. The three most common coordinating conjunctions.

<b>Conjunction</b>	<b>Freq. A</b>	<b>% A</b>	<b>Freq. B</b>	<b>% B (rank of occurrence)</b>
<i>ja</i> ‘and’	83,292	3.2	47,960	3.11 (1 <sup>st</sup> )
<i>tai</i> ‘or’	40,775	1.57	37,249	2.42 (2 <sup>nd</sup> )
<i>sekä</i> ‘and, as well as’	8,477	0.33	8,890	0.58 (3 <sup>rd</sup> )

It seems that national legislation rather presents its case in terms of different alternatives. Whether one uses an additive or a disjunctive conjunction may also be influenced by the differing uses of the conjunctions in the source language and target language. In addition, English-language directives often use the combination ‘and/or’, which in national legal texts is altered to one of the two conjunctions, the choice depending on each case (see Piehl 2013).



The fourth most common coordinating conjunction is not the same in the two corpora. It is *joko* ‘either’ in corpus A and *taikka* ‘or’ in corpus B. *Taikka* is a peculiarity of Finnish legal language; it is a synonym in regular language of the conjunction *tai*. Legal language uses this conjunction in a way which differs from most of regular language use, namely to structure a list of options. In legal language, it separates items more clearly than *tai*, indicating the scope of the phrase. (Lainkirjoittajan opas, 2013) *Taikka* is clearly more uncommon in corpus A, for the source text does not encourage a similar distinction between conjunctions, and the text cannot necessarily be organised to follow the principles of Finnish legal language. This distinction is not commonly known. There is also a third conjunction in the group of disjunctive conjunctions: *tahi*. Among the three conjunctions, *tahi* is the one that has – in legal language – the strongest separating power. It has an archaic tone in regular language use and is already rare even in corpus B. It does not appear at all in corpus A.

Correspondingly, the additive conjunction *sekä* separates the constituent items of a list more distinctly than does *ja* in legal language. This is also evident in regular language use, even if the use is more systematic in legal language. The special nature of *sekä* in national legal language is probably seen in the fact that this conjunction is more common in corpus B than in corpus A. The correlative conjunctions *joko-tai* ‘either-or’ and *sekä-että* ‘both-and’ are more common in corpus A, which may suggest that the source text has also used correlative conjunctions and that this rhetoric device is more common in the directives than in national legislation. Coordinating conjunctions indicating other kinds of relationship are rare in both corpora, although *mutta* ‘but’ is slightly more common in corpus A. In addition, the choice of conjunctions is more varied in corpus A: it features two conjunctions (*sillä* ‘since, because’ and *eli* ‘or’) that are not found in corpus B at all. This also tallies with the findings of the previous comparative corpus study (Piehl 2006).

Table 8. The four most common subordinating conjunctions.

Conjunction	Freq. A	% A	Freq. B	% B (rank of occurrence)
<i>että</i> ('that')	<b>13,994</b>	<b>0.54</b>	5,553	0.36 (3 <sup>rd</sup> )
<i>jos</i> ('if')	11,251	0.43	<b>13,668</b>	<b>0.89 (1<sup>st</sup>)</b>
<i>kuin</i> ('as')	8,166	0.31	5,567	0.36 (2 <sup>nd</sup> )
<i>kun</i> ('when')	5,618	0.22	2,628	0.18 (4 <sup>th</sup> )

There are far fewer subordinating conjunctions in Finnish than there are coordinating conjunctions. In corpus B, *jos* 'if' clearly holds a dominant position. This suggests that the conditions to legal rules are more uniformly expressed in Finnish legal language than in the EU directives, which use a wider range of means. Still, 'if' clauses are not the only means for national legislation to express conditionality. For example, relative pronouns and temporal conjunctions are used, too.

A typical case of subordination is also the relationship that relative clauses and interrogative clauses have to the main clause. Clauses beginning with the relative pronoun *joka* 'who'/'which' are the most frequent type of subordinate clause in both corpora; they are much more common than 'if' clauses (*jos*) and 'that' clauses (*että*), but clauses beginning with *joka* are significantly more common in corpus A.

The second most common pronoun beginning a subordinate clause is *mikä* 'which'. This may refer to either a relative clause or an interrogative clause, but interrogative subordinate clauses are considerably rarer in legal language in particular (Virtaniemi 1992: 102). The

pronoun *mikä* occurs much more rarely in corpus A than in corpus B, even if it is part of some recurring sentence types. There is also a difference in the frequency of occurrence of the relative adverb *jolloin* ‘when, whereupon’ and the interrogative adverb *milloin* ‘when’, which both indicate temporal relations. In corpus A, *jolloin* is more common than *milloin*, while the opposite is the case in corpus B. In both corpora, *milloin* occurs in the initial position in an interrogative subordinate clause, but in corpus B in particular, it also expresses a condition in a manner similar to the temporal conjunction *kun* ‘when’. Such usage is typical of legal language, and it is also found in corpus A to a certain extent: this mode of expression is, however, already somewhat archaic also in legal language.

As a rule, legal texts contain few causal clauses (see, e.g. Virtaniemi 1992: 100), but the two corpora are somewhat different in this respect, for the causal conjunction *koska* ‘as’ /‘because’/‘since’ is more common in corpus A, and the conjunction *sillä* ‘for’ does not occur in corpus B at all. Finnish legal language refers to cause also by the combination of the adverb *siksi* and the conjunction *että* (which together translate as ‘because, for the reason that’). *Siksi* ‘therefore’ on its own is similarly more common in corpus A. Legal texts do not typically express consequence, either: the relevant adverb *joten* ‘so’ does not occur in corpus B at all, but there are some examples of it in corpus A.

Legislation commonly indicates a condition, because the application of legal rules typically depends on the fulfilment of certain requirements. It has already been noted that conditional ‘if’ clauses (with the conjunction *jos*) are common in both corpora and clearly more common in corpus B. A second conditional conjunction, *mikäli* ‘if’/‘provided’, occurs to a degree in corpus A, but not once in corpus B. It is not part of the national legal register, although this conjunction by no means represents informal or everyday language, quite the opposite.

The if-then construction is not found in corpus B, even though clause-initial ‘if’ clauses (beginning with the conjunction *jos*) are highly typical of Finnish legal language. In

Finnish, the particle *niin* 'then' begins the apodosis in regular language use and speech. It is stylistically somewhat ordinary, so the construction is extremely rare in national legislation and only marginally more common in the directives.

In Finnish, grammatical and semantic relations can be expressed not only by grammatical cases but also by adpositions, some of which may be syntactic calques. The majority of adpositions are postpositions. There is a clear difference between the two corpora: the directives make use of a much wider range of adpositions than the national legislation. Corpus A contains 35 adpositions that are not found at all in corpus B, while corpus B has only three such adpositions that do not occur in corpus A. This may be contact-induced, as the source languages typically use prepositions.

The Finnish legal drafter's manual, *Lainkirjoittajan opas*, states that legal texts must refer to statutes in the passive rather than in active forms. The manual consequently advises drafters to write *laissa säädetään* 'it is provided in the law' (local construction + finite verb in passive form) rather than *laki säättää* 'the law provides' (subject + finite verb). Nouns such as *laki* 'law', *direktiivi* 'directive', *asetus* 'decree', and *sopimus* 'treaty' are not used as instrumental subjects with the verbs *säätää* 'to provide'/'lay down' and *määrätä* 'to determine'/'provide' in either of the corpora. The manual makes a similar declaration in favour of passive participles, which abound in legal texts. According to the manual, one should write *laissa tarkoitettu* 'which is intended in law' (local construction + passive participle) instead of *lain tarkoittama* 'intended by law', a so-called agent participle, where the genitive of the first word refers to the agent of the participle. Such constructions with an instrumental subject appeared in Finnish legislation already decades before Finnish membership of the EU.

Regardless of the advice of the manual, there are some cases in corpus B of an agent participle, *tarkoittama*, coupled with a genitive, where the law or an equivalent word denoting a statute appears as an active subject, as in:

(6) Jos ehdotetaan säästöpankin purkamista muussa kuin 2 momentin tarkoittamassa tapauksessa, kokouskutsu on kuitenkin toimitettava viimeistään kuukautta ennen kokousta.

If the proposal relates to the dissolution of the savings bank in a case other than that referred to in paragraph 2, the notice to convene shall, however, be submitted at the latest one month before the meeting.

(Savings Bank Act 1502/2001, section 102)

Overall, however, the guidelines of the manual have probably been effective, for there are only 43 such occurrences in corpus B. Corpus A has even fewer instrumental subjects constructed with an agent participle, only 14. The passive participle recommended in the manual, *laissa tarkoitettu*, is vastly more common in both corpora; it appears around 7,000 times in corpus A, and there are just under 8,000 occurrences in corpus B. Agent participles of other verbs do not occur together with an instrumental subject in either corpus. The few instrumental subjects do not appear to be contact-induced: where the Finnish-language version of a directive has an agent participle with a subject, the English-language version does not use an instrumental subject, but rather employs varied phrases, such as ‘covered by’, ‘referred to in’, ‘within the meaning of’.

## 6. Textual level

Since legal language aims at precision and accuracy, special attention is paid to coherent terminology and clear co-references. As a result, Finnish legal language favours repetition (Mattila 2013: 88), not synonymy or paraphrases. Nevertheless, pronouns are among the most common words in both corpora, as they also are in regular language, where the pronouns *se* ‘it’, *hän* ‘he’/‘she’, and *joka* ‘who’ are among the twenty most common words.

Both corpora use the pronoun *tämä* ‘this’ most often to refer to the statute in which it is incorporated: *tässä direktiivissä* ‘in this directive’, *tämän lain nojalla* ‘under this law’.

Table 9. Cohesive pronouns.

<b>Pronouns</b>	<b>Rank in freq. of all words A</b>	<b>Freq. A</b>	<b>% A</b>	<b>Rank in freq. of all words B</b>	<b>Freq. B</b>	<b>% B</b>
<i>se</i> ‘it’	4 <sup>th</sup>	40,847	1.57	5 <sup>th</sup>	22,607	1.47
<i>joka</i> ‘who/which’	6 <sup>th</sup>	39,160	1.51	6 <sup>th</sup>	17,733	1.15
<i>tämä</i> ‘this’	9 <sup>th</sup>	28,966	1.11	11 <sup>th</sup>	13,081	0.85
<i>hän</i> ‘(s)he’	134 <sup>th</sup>	2,094	0.08	28 <sup>th</sup>	4,889	0.32

It is not easy to track repetition by frequency lists or concordance searching, but these methods also make it possible if one uses such words as the pronoun *tämä* ‘this’, the proadjective *tällainen* ‘such’/‘as this’/‘like this’, and other precision-expressions. Especially in the past, repetition alone was not deemed to be an adequate means of ensuring the precision of legal texts. The so-called ‘juridical pronouns’ were then used to make cohesion stronger (Mattila 2013: 88). These expressions consist of a participle and a preceding qualifier, and they are used as a preceding qualifier of a word or word element. Such expressions include *sanottu* ‘said’, *edellä/yllä/allajäljempänä mainittu* ‘above-mentioned’/‘mentioned below’, *puheena/kyseessä oleva* ‘present’/‘in question’, and the adjective *kyseinen* ‘in question’. Their use is a typical feature of legal and official language in particular.

In an earlier legal drafter's manual (Lainlaatijan opas 1996: 65), juridical pronouns are mostly considered superfluous, and modern legal language is seen to avoid them. However, they still occur in Finnish legal texts, and they are not absent from the Finnish versions of directives, either. Compared to grammatical pronouns, however, they are used much less; they are slightly more common in corpus A. This is likely the influence of the source language: if the text contains the word 'said' or 'above-mentioned', it is not likely that the Finnish version would lack an equivalent word.

In addition, the proadjective *tällainen* 'such as'/'like this' and to some extent the demonstrative pronoun *tämä* 'this' occur when a word or word element is repeated. These two are similarly more common in the directives than in the national legislation. These devices do not provide direct information about the extent of and variation in the use of repetition; after all, texts do include repetition that has not been marked in any way.

The use of inference markers shows intralinguistic variation between the directives and national legislation. Expressions indicating how a writer perceives relations between things are not among the most common phrases in legal texts. More often, one finds phrases that define factual relations, such as *viimeistään* 'at the latest', *enintään* 'at most', and *vähintään* 'at least'. Inference markers that indicate a causal connection are relatively rare in legal texts, as are causal conjunctions and adpositions, and corpus B has fewer of them than corpus A. There are, for example, a few hundred phrases such as *näin ollen* 'hence'/'consequently', *tämän/sen vuoksi* 'to this end'/'this is why', and *siksi* 'therefore' in corpus A, while there are only a few dozen or none indicating a causal relation in corpus B.

Inference markers are more common in corpus A, probably because it is seldom possible not to translate the source language words. In addition, the material in corpus A includes directive preambles, which differ genre-wise somewhat from the actual legal text, and there is no corresponding text in the national statutes.

There are clear differences between the inference markers in the two corpora. The assortment of inference markers is larger in corpus A – there are more than 500 of them – which makes their use more varied. Corpus B contains just over 300 inference markers.

Table 10. Inference markers.

Inference markers	Freq. A	% A	Freq. B	% B (rank of occurrence)
<i>erityisesti</i> ‘in particular’	3,621	0.14	293	0.02 (9 <sup>th</sup> )
<i>myös</i> ‘also’	3,374	0.13	<b>3,321</b>	<b>0.22 (1<sup>st</sup>)</b>
<i>kuitenkin</i> ‘however’	2,606	0.10	2,631	0.17 (3 <sup>rd</sup> )
<i>ainoastaan</i> ‘only, solely’	1,813	0.07	375	0.02 (6 <sup>th</sup> )
<i>siten</i> ‘as, so’	1,719	0.07	1924	0.13 (2 <sup>nd</sup> )
<i>vain</i> ‘only’	1,012	0.04	1323	0.10 (4 <sup>th</sup> )
<i>lisäksi</i> ‘further’/‘moreover’	988	0.04	1526	0.10 (5 <sup>th</sup> )
<i>mahdollisesti</i> ‘possibly’	768	0.03	207	0.01 (11 <sup>th</sup> )



There is a distinct difference in the frequency of use of the inference marker *mukaan lukien* ‘including’, which contains the adverb *mukaan* ‘along’ and the infinitive *lukien* ‘counting’. It appears clearly more often in the directives than in the national legislation, serving as the Finnish equivalent of the preposition *including*.

Expressions qualifying legal rules may refer to the legal basis of rules or their relation to other legal rules. Some phrases indicate that the item to be observed appears elsewhere in legislation. Other phrases show that the legal rule in question does not prevent the application of another legal rule or provision but gives way to this rule, or vice versa.

There is a distinction between the directives and national legislation in the use of these markers. In corpus A, a legal rule or provision that applies to a case is usually referred to by the adposition *mukaisesti* ‘according to’. It is usually the equivalent of the phrase ‘in accordance with’. It also appears in corpus B, albeit less dominantly. Constructions typical of corpus B that refer to another legal rule are subordinate clauses that have the passive *säädetään* ‘is provided’ of the verb *säätää*. These include relative clauses such as *noudatetaan/on voimassa, mitä - - säädetään* ‘shall observe what is provided’ or conjunctive clauses, such as *niin kuin - - säädetään* ‘as is provided’. These clause types are less frequent in corpus A, so much so that there are no occurrences of the conjunctive constructive with the verb *säätää*. The verb itself is clearly more common in corpus B.

There are also differences in how legal rules and provisions construct conflict-resolving expressions that signal the prioritisation of one legal rule over another. The most common phrase for this purpose in corpus A is the infinitive *rajoittamatta* (literally: ‘without restricting’). The phrase usually appears as *sanotun kuitenkaan rajoittamatta* (literally: ‘without being restricted by what has been said’) and as the equivalent of the phrase *without prejudice to*. This phrase appears twice in corpus B, and only four times in the national legislation as a whole, and even then only in legislation enacted since Finland’s

membership of the EU. On its own, the infinitive *rajoittamatta* is typically found in a clause-initial position in the directives, as is *without prejudice to*, whereas the few occurrences in the national legislation tend to come mid-sentence.

National legislation prefers the infinitive *estämättä* ‘notwithstanding’ (literally: ‘without preventing’) to indicate in which cases the provision at hand comes first in relation to the provision mentioned in the infinitive phrase. The infinitive *estämättä* has either an initial qualifier (as in *salassapitosäännösten estämättä* ‘notwithstanding provisions relating to secrecy’) or it occurs with a subordinate clause that contains the passive *säädetään* (e.g. *sen estämättä, mitä 2 momentissa säädetään* ‘notwithstanding what is provided in paragraph 2’). *Estämättä* also appears in legislation predating Finnish EU membership. The phrase *sen estämättä* is used to a certain degree in the directives as the equivalent of *notwithstanding*, and like its equivalent, it appears in a sentence-initial position in corpus A. In corpus B, too, this infinitive occurs at the beginning of the sentence.

There are also other ways to indicate that the provision at hand is secondary to another provision, such as including a conditional clause with the adverb *muualla* (literally: ‘elsewhere’), as in *Tätä lakia sovelletaan, jollei muualla laissa toisin säädetä* ‘This Act shall be applied, unless otherwise provided by law’, and the use of a conditional clause with the pronoun *muu* ‘other’, which is the case in the following example:

(7) Jäsenvaltiot voivat sallia vitamiinien ja kivennäisaineiden lisäämisen liitteessä I olevassa I osassa määriteltyihin tuotteisiin, jollei direktiivistä 90/496/EEC muuta johdu.

Subject to Council Directive 90/496/EEC of 24 September 1990 on nutrition labelling for foodstuffs, the addition of vitamins to the products defined by this Directive is permitted in some Member States.

(2001/112/EC)

The latter kinds of sentence also appear in corpus A, but this corpus does not contain any constructions corresponding to *muualla laissa*.

The legal base is commonly referred to in the Finnish versions of the directives by the infinitive phrase *ottaen huomioon*, an equivalent of *having regard to*. The Finnish *ottaen huomioon* consists of an infinitive and a noun. The phrase was in use before Finland's membership of the EU, but it appears much more frequently in corpus A than in corpus B. The corresponding position in the national legislation, preceding the actual text containing the provisions, is occupied by the sentence *Eduskunnan päätöksen mukaisesti säädetään* 'In accordance with decision of Parliament'. If the provision stems from a directive, this is indicated briefly at the end of the act, as in: *Euroopan parlamentin ja neuvoston direktiivi 2006/115/EC (32006L0115); EUVL L 376, 27.12.2006, s. 28* 'Directive 2006/115/EC of the European Parliament and of the Council (32006L0115); EUVL L 376, 27.12.2006, p. 28'.

## 7. Conclusions

The study of the Finnish versions of the EU directives and their comparison with the MNI shows that the influence of the source-language versions on the Finnish-language versions of EU directives and also the influence of the latter on Finnish legislation is limited to only certain features of the Eurolect Research Template.

The main reason is that Finnish is a non-Indo-European language with different phonology, morphology, and syntax. Lexical borrowing is very much restricted, and some kinds of morphological interference are impossible. While there are restrictions on the borrowing of morphemes and syntactic structures, the expressive needs created by language contacts are met by using the resources available in the Finnish language. This kind of influence is mainly manifested as intralinguistic variation. We found few features that could be counted as EU-rooted or contact-induced. They may be difficult to detect.<sup>4</sup>

---

<sup>4</sup> It might be necessary to note, that in Sweden the quality of EU Swedish and its influence on Swedish language in law has been discussed along very similar lines to Finland. There are no corresponding corpus-based studies to compare with, but other

Most of our findings demonstrate the indirect influence of the structure and style of English on the Finnish-language versions. Also an influence of French can be seen, especially in the older directives. The position of French in the EU was still very strong in the beginning of the 1990-ies, and thus the *acquis* was translated into Finnish from this language, not from English. The effect of source languages can be detected in changes in frequencies, the overuse of certain constructions, and the underuse of others. This was evident in many features of language: in corpus A, we found the overuse of loanwords, proper names, prefixes, adjectives, modal verbs, adjectives, and the nominative case, and the underuse of other grammatical cases that are so distinctive to the Finnish language.

The hypothesis that translations are simplified versions of the original texts in the same language, as suggested in many works on translation studies (see, e.g. Baker 1996; Olohan 2004: 91–104), does not seem to apply to the language of law. On the contrary, the range of tenses and moods on the morphological level and the assortment of adpositions and inference markers on the syntactic level, for example, seem more varied in the Finnish-language versions of the EU directives than in the national legislation. The explanation may be, however, that national legal texts are stylistically more constricted and that it may be possible to use phrases and constructions in the directives that in a national legal text would be altered to comply with established practice. However, there are findings that point to simplification, such as the underuse of most cases.

Overall, our findings reveal that many of the features that distinguish legal language from other genres appear alike in the directives and national legislation, among them the overuse of the passive voice, the indicative mood, long clauses, demonstrative and juridical pronouns, and the peculiarities of word order. On the other hand, we demonstrate that when

---

studies show that differences between the language of directives and that of national legislation are evident also in Swedish. The sentence structure in directives is more complex with markedly longer sentences and more layered hypotaxis (Hofman 2015: 96-97). Also, differences in choice of prepositions and use of words and phrases instead of regular prepositions have been reported (Edgren 2001: 72). Interestingly, Hofman's findings indicate that the more archaic structures peculiar to Swedish language of law are lacking in directives; a result resembling ours.

it comes to choosing between the alternative linguistic expressions available, the writers of the directives and the writers of Finnish national legislation often make different choices. They do not always resort to the same adpositions, difference markers, or phrases to qualify the legal rules, and so on. The root of these differences is most probably interference from the English-language versions, which trigger the choice of an expression as close as possible to the source language. The differences discovered may well be distinctive enough to enable us to say that a Finnish Eurolect does exist.

## References

- Baker, Mona. 1996. Corpus-based translation studies: the challenges that lie ahead. In *Terminology, LSP and Translation. Studies in Language Engineering, In Honour of Juan C. Sager*, H. Somers (Ed.), 175-185. Amsterdam/Philadelphia: John Benjamins.
- Edgren, Håkan. 2001. Svenskan i EU och EU på svenska. Svenskans beskrivning 24. In *Förhandlingar vid Tjugofjärde sammankomsten för svenskans beskrivning Linköping, 22–23 oktober 1999*. L. Jönsson, V. Adelswärd, A. Cederberg, P. A. Pettersson, C. Kelly (eds.), 69-80. *Studies in Language and Culture*, No. 2. Linköping Electronic Conference Proceedings, No. 6. Linköping: Linköping University Electronic Press.
- Hakulinen, Auli. 2001. Persoonaviittauksen välttäminen suomessa. In *Lukemisto. Kirjoituksia kolmelta vuosikymmeneltä*, L. Laitinen, P. Nuolijärvi, M.-L. Sorjonen, & M. Vilkuna (eds.), 208-218. Helsinki: Suomalaisen Kirjallisuuden Seura. (Finnish translation of Hakulinen 1987).

Hakulinen, Auli. 1987. Avoiding personal reference in Finnish. In *The pragmatic perspective. Selected papers from the 1985 International Pragmatics Conference*, J. Verschueren & M. Bertucelli-Papi (eds.), 141-153. Amsterdam: John Benjamins.

Hofman, Geertrui. 2002–2003. Har det hänt något på klarspråksfronten? En synkron och diakron jämförelse av svenska EU-direktiv. Verhandeling voorgelegd aan de Faculteit Letteren en Wijsbegeerte voor het verkrijgen van de graad van licentiaat in de taal- en letterkunde: Germaanse talen.

Kanner, Antti. 2011. Kansallisen lainsäädännön ja EU-säädösten kieli nesessiivisten rakenteiden valossa. Pro gradu -tutkielma. Helsingin yliopisto, Suomen kielen, suomalais-ugrialaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos, suomen kieli.

Kankaanpää, Salli et al. 2012. Kuinka tutkia lakitekstin syntyä ja tekstilajin rajoja? Tekstianalyysia, muutosehdotuksia ja osallistuvaa havainnointia. In *Genreanalyysi: Tekstilajitutkimuksen käytäntöä* V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiililä, & M. Lounela (eds.), 129-156. Kotimaisten kielten keskus. <<http://scripta.kotus.fi/www/verkkojulkaisut/julk29/Genreanalyysi.pdf>>.

Kettunen, Jenni. 2013. Pakon ja mahdollisuuden ilmaisemiskeinot suomen- ja venäjänkielisissä kansainvälisissä sopimuksissa. Pro gradu -tutkielma. Tampereen yliopisto, Keli-, käännös- ja kirjallisuustieteiden yksikkö, Käännöstiede. <http://urn.fi/urn:nbn:fi:uta-1-23796>

Kváran, Guðrún. 2007. Brug af afløsningsord i de nordiske sprog. Sammenligning og konklusioner. In *Udenlandske eller hjemlige ord? En undersøgelse af sprogene i Norden*, G. Kváran (ed.) 169-187. Moderne importord i språka i Norden VI. Oslo: Novus.

*Lainkirjoittajan opas* 2013. Helsinki: Oikeusministeriö.

*Lainlaatijan opas* 1996. Helsinki: Oikeusministeriö.

Mattila, Heikki E.S. 2010. Suomalaisen oikeuskielen ominaispiirteet. In *Oikeuskieli ja säädöstieto – rättspråk och författningsinformation* H. Mattila, A. Piehl & S. Pajunen (eds.), 181-213. Suomalaisen Lakimiesyhdistyksen julkaisu, C-sarja 4. Suomalainen Lakimiesyhdistys.

Mattila, Heikki E.S. 2013. *Comparative Legal Linguistics. Language of Law, Latin and Modern Lingua Francas* (2<sup>nd</sup> edition). Farnham: Ashgate.

Olohan, Maeve. (2004). *Introducing Corpora in Translation Studies*. London and New York: Routledge.

Piehl, Aino. 2013. Fine tuning style and precision: Adapting directive citations to Finnish statutes. *Linguistica Antverpiensia. New Series – Themes in Translation Studies* No 12. Research models and methods in legal translation. <<https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/313>>.

Piehl, Aino. 2012. Les fonctionnaires dans la préparation des actes législatifs de l'UE: L'équilibre entre leur propre langue et celles de travail. In *L'Europe des 27 et ses langues. Collection Europe(s) No 3*, Herreras, J. C. & Ziegler K. (eds.), 685–698. Presses Universitaires de Valenciennes.

Piehl, Aino. 2010. Suomalaisen oikeuskielen kehittäminen ja huolto. In *Oikeuskieli ja säädöstieto – rättspråk och författningsinformation*, H. Mattila, A. Piehl & S. Pajunen (eds.), 147-179. Suomalaisen Lakimiesyhdistyksen julkaisu, C-sarja 4. Suomalainen Lakimiesyhdistys.

Piehl, Aino. 2008. Virkamiehet EU:n säädösvalmistelussa: Tasapainoilua oman kielen ja työkielten välissä. In *Kielen ja oikeuden kohtaamisia: Heikki Mattilan juhlakirja*, R. Foley, T. Salmi-Tolonen, I. Tukiainen, & B. Vehmas (eds.), 207–227. Helsinki: Talentum.

Piehl, Aino. 2006. The Influence of EU Legislation on Finnish Legal Discourse. In *Linguistic Insights. Studies in Language and Communication*, Maurizio Gotti and Davide S. Giannoni (eds.) Vol. 44 (pp. 183-194). Bern: Peter Lang.

Rydzewska-Siemiątkowska, Joanna. 2016. Linguistic analysis of necessity expressions in Finnish and Polish legal text in terms of deontic strength. *Comparative Legilinguistics. International Journal for Legal Communication*. Vol 26. <<http://dx.doi.org/10.14746/cl.2016.26.03>>.

Sandøy, Helge. 2007. Avisspråket i Norden – ei jamføring. In *Fire dagar i nordiske aviser. Ei jamføring av påverknaden i ordforrådet i sju språksamfunn*. B. Selbak, H. Sandøy (eds.), 127-155. Moderne importord i språka i Norden III. Oslo: Novus.

Tyynilä, Markku. 1984. *Lainvalmistelukunta 1884–1964*. Suomalaisen Lakimiesyhdistyksen julkaisuja, A-sarja 166. Suomalainen Lakimiesyhdistys.

Virtaniemi, Asta. 1992. Säädoskielen virkerakenne. Pro gradu -tutkielma. Jyväskylän yliopisto.