
Contrastive Analyses, Translation, and Speaker Involvement: the Case of *Puisque* and *Aangezien*

LIESBETH DEGAND[□]

1 Introduction

The goal of this article is threefold. First, I would like to demonstrate the strength of corpus studies combining both comparable and parallel corpora for contrastive linguistic analyses. My second aim is to bring some additional evidence to the linguistic validity of the Speaker Involvement Scale (Degand and Pander Maat, in press; Pander Maat and Degand, 2001), and finally, I would like to illustrate how the interaction between corpus analyses and the use of the Speaker Involvement Scale can bring us closer to a thorough description of the supposed French-Dutch translation pair *puisque-aangezien* ('since').

Contrastive linguistics and translation studies have shared a common domain of investigation for many years. The aim of contrastive linguistics is to describe similarities and divergences between two or more languages through systematic comparisons. The goal of these comparisons can be both theoretical and applied. Translation studies proceed in a similar way, but here the goal is to systematically compare original texts with their

[□] The author is a research fellow at the Belgian National Fund for Scientific Research (F.N.R.S.). This work is supported by grant FRFC 2.4551.03. I would like to thank Sylviane Granger (University of Louvain, Louvain-la-Neuve) for comments on an earlier version of this article.

translations with the aim of discovering translation schemes or even universals of translation (Baker, 1993). Despite these similarities in methods and goals, a convergence between the two disciplines has arisen only fairly recently with the appearance of vast corpora (Granger 2003). Access to these data does indeed enable the researcher to test and quantify his or her intuitions about linguistic contrasts in two or more languages.

In contrastive linguistics, two types of corpora can be used:

- Comparable corpora consisting of original texts in two or more languages, matched by criteria such as the time of composition, text category, intended audience, and so on. (Johansson and Hasselgård 1999).
- Parallel corpora that contain a series of source texts aligned with their corresponding translations (Granger 2003).

Comparable corpora display the advantage of presenting original data produced by native speakers of the languages under investigation. Hence these data should be free of any crosslinguistic influence and thus represent natural linguistic data in context. In addition, they are fairly easy to collect (cf. literary texts and written press). In sum, comparable corpora make it possible to investigate the context and constraints of use of a given linguistic structure in one or more languages, but as such they do not permit one to establish (translation) equivalences between language pairs. This latter task is coped with in parallel corpora that vehicle a (quasi) identical semantic content in at least two different languages. Their main disadvantage, however, is that they very often contain traces of the source language and can thus not count as totally reliable data, specifically in terms of occurrence frequencies (Granger 2003). Another disadvantage is that they are difficult to collect because on the one hand, not all text types are translated (letters, e-mail), and on the other hand, not all languages are translated with the same frequency (English texts are translated more often than Dutch ones, for instance).

2 *Puisque—Aangezien*: Translation Equivalents?

In the following sections, I would like to illustrate through a concrete example how systematic analyses of parallel and comparable corpora in the light of a given linguistic theory are an indispensable tool to explore and uncover the (adequate or inadequate) usage of given linguistic structures. To this end, I will concentrate on the use of the French-Dutch supposed translation pair *puisque-aangezien* ('since'). This pair is interesting for several reasons.

In the first place, it is a rather stable translation pair in bilingual dictionaries. Table 1 shows that in traditional translation dictionaries

(Boogaards, 1998; Larousse 1983; Coenders 1993) the lexical entry for *puisque* always proposes *aangezien* as one of its translation alternatives (although not always as the first one), and *puisque* is nearly always one of the alternatives given for *aangezien*.

	puisque	aangezien
Van Dale (Boogaards 1998)	aangezien (since, as, given that), daar ... immers (as indeed), nu ... toch (as it happens), want (because/for)	puisque, étant donné que, attendu que
Larousse (1983)	aangezien, daar (as)	puisque, comme, vu que, attendu que, étant donné que, du moment où, dès lors que, dès l'instant que, d'autant que
Wolters (Herckenrath and Dory 1985)	aangezien, daar (immers) (as indeed), nu ... toch (as it happens)	vu que, attendu que, puisque
Kramers (van Kooten 1982)	daar (as), aangezien, naardien (considering that); nu (now that), immers (after all, indeed)	attendu que, vu que, puisque
Grootaers (1979)	daar (as), aangezien	vu que, attendu que

Table 1. Translation alternatives for *puisque* and *aangezien* in traditional bilingual dictionaries

Nevertheless, as Table 1 also shows, the meaning conveyed by *puisque* and *aangezien* can be expressed by other linguistic constructions as well. This becomes even more apparent from a search in the CD-ROM version of the Dutch-French bilingual Van Dale (1997), in which *aangezien* also appears as a translation alternative for the conjunctions *attendu que*, *comme*, *du moment que*, *parce que*, *pendant que*, and *vu que*.¹ Similarly, *puisque* is given as a possible translation for the conjunctions *daar*, *dewijl*, *doordat*, *doordien*, *immers*, *naardien*, *nou*, *nu*, *omdat*, *toch*, *vermits*, and *waar*.² While dictionaries are very sparse in giving semantic and contextual information which would guide the language user in picking out the most ade-

¹ Considering that, as, as long as, because, while, seeing that.

² As, while, because of the fact that, owing to the fact that, after all, considering that, now that, now that, because, after all, since (arch.), where.

quate alternative, this wide variety of alternatives could be a hint that the semantics of the *puisque-aangezien* pair is very complex.

In this context, it is noteworthy that the linguistic literature seems to implicitly consider the two connectives as equivalent. Van Belle (1989) basing his description of *aangezien* on that of *puisque* by Groupe \square -1 (1975) makes use of the same syntactic, semantic, and pragmatic parameters, although his description is not explicitly contrastive in nature. While monolingual descriptions of *puisque* have been fairly frequent (Martin 1973 \square Groupe \square -1 1975 \square Hensen 1976 \square Delbey 1988 \square Iordanskaja 1993 \square Franken 1995 \square Olsen 2001), this is not the case for *aangezien* (Van Belle 1989), nor for contrastive studies between the two connectives. Their usage in Dutch and French comparable corpora is, however, rather divergent. In a recent contrastive study (Degand and Pander Maat, in press), we observed very divergent frequencies of the two connectives in written press corpora³: 266 per million words for *puisque* vs. 35 per million words for *aangezien*. In addition, the syntactic behavior of the two connectives is also very different, *aangezien* being used in preposed position (Aangezien Y, X) in 46% of the cases in the examined written press corpus, against only 4% for *puisque*. In order to investigate whether these divergences in frequency and syntactic behavior do, in fact, reflect semantic differences between the two connectives, we did a systematic analysis of the distribution and semantic nature of the two connectives in these written press corpora (Degand and Pander Maat, in press). In the following section, I will briefly describe the methodology used and the results we obtained. I will then proceed in a similar way with the parallel *puisque-aangezien* data.

3 Speaker Involvement, Causal Relations and Connectives

The linguistic framework used for this research is the Speaker Involvement Scale (Degand and Pander Maat, in press; Pander Maat and Degand 2001). This scale is an alternative proposal to classify the use and meaning of connectives in discourse. Going beyond dichotomous and trichotomous classifications (Halliday and Hasan 1976; Martin 1992; Moeschler 1989; Sweetser 1990, Van Dijk 1977), our proposal is to represent (causal) coherence relations and connectives in a scalar way. This scalar representation reflects the fact that connectives are not strictly domain specific, but that they neverthe-

³ The corpora we used were electronic versions of two daily newspapers: *NRC Handelsblad* (year 1994), a Dutch national newspaper, distributed electronically by INL (Dutch Institute for Lexicology) counting approximately 27 million words; and *Le Soir* (year 1997), a Belgian francophone newspaper distributed on CD-ROM, counting approximately 26.8 million words.

less impose constraints on the contexts in which they can occur, with some contexts being more ‘natural’ than others. In addition, a number of causal connectives seem to take an intermediate position between the traditional categories. According to Degand and Pander Maat, this situation is an indication of the need for a scalar perspective on the spectrum reaching from non-volitional causality in the content domain to epistemic and speech act causality. The scale we have developed is one of *speaker involvement* (SI), on which the inherent expressive power of connectives can be represented. Our hypothesis is that the different causal relations can be ordered along a scale from minimal to maximal speaker involvement. SI refers to *the degree to which the present speaker is implicitly involved in the construal of the causal relation*. SI increases with the degree to which both the causal relation and the related segments convey assumptions and actions of the present speaker. Four characteristics of coherence relations may enhance the SI level of the relation: the involvement of a conscious protagonist, a lack of isomorphism between the relation and states of affairs in the real world, proximity of the relation to present speaker and speaking time, and the implicit vs. explicit realization of the protagonist (cf. *infra*). The different causal relations we distinguish are, in order of increasing SI: nonvolitional (1) and volitional content relations (2); causal (3-4) and noncausal epistemic (5) relations, and speech-act relations (6-7).

- (1) Snow White died (because) she ate a poisoned apple.
- (2) Cinderella ran away at midnight (because) she knew the spell was about to be broken.
- (3) Snow White had to die (because) the queen was jealous.
- (4) Pinocchio was lying (because) his nose was growing.
- (5) The queen had to be punished, (because) this is a fairy tale.
- (6) Do you want me to read the Bambi story? (Because) it’s time to go to bed.
- (7) Walt Disney’s cartoon empire—[’cause] that’s what it is—has to face some serious rivals.

3.1 Scaling Causal Relations

As mentioned above, the ordering of these causal relations on the SI scale is the result of an interplay between four characteristics which will now be briefly illustrated (but see Pander Maat and Degand 2001 for a more thorough development):

1 Involvement of a conscious participant

The more the causal relation implies the subjective involvement of a conscious participant, the higher the SI. Nonvolitional relations (8) do not involve such a participant, while volitional relations (9) do. Since conscious participants may entertain assumptions and may be identified with by the speaker/hearer, volitional relations carry a higher degree of SI than nonvolitional ones—as a speaker/hearer, I can identify myself with Peter and agree with his reason for leaving early, or I might disagree. Epistemic relations (10) are still higher on the scale because they involve beliefs of a concluding protagonist with whom the speaker must share a number of assumptions in order to understand the causal relation. Finally, maximal SI is reached with speech-act relations (11) in which the protagonist is identical to the Speaker by definition.

- (8) The whole area flooded because it had been raining nonstop for ten days.
- (9) Peter left the party early because he had to get up at 6 a.m.
- (10) Joan thinks the children should be punished because they have been mean.
- (11) What's your address, please, 'cause I have to fill in this form.

2 A lack of isomorphism between the relation and the states of affairs in the real world

The less isomorphic a relation is with real world causality, the higher the SI. The volitional relation in (9) represents a real world causal relation. The same causal situation is present in (12), but this time it is no longer the primary causal relation expressed: a reason-consequence relation in the real world is transposed into the mental domain of making inferences; in other words, it is used by the speaker to base an argumentation upon. This is why the epistemic relation in (12) carries a higher degree of SI than the volitional relation in (9). In (13), then, real world causality is not represented, but it continues to impose constraints on the propositional content of the related segments, the real world causal relation is 'turned around' to yield a pattern of abductive inference. This is no longer the case for noncausal epistemic relations (14), based on analogy reasoning, for instance, and speech act relations (11), which show maximal detachment from real world causality.

- (12) Peter must have left the party early because he had to get up at 6 a.m.
- (13) Peter certainly has to get up early tomorrow because he's already left the party.
- (14) Leo is tall because his brother is tall too.

3 Proximity of the speaker

The closer the relation is to the present speaker and moment of speaking, the higher the SI. Hence, (15), (16), and (17) illustrate increasing SI.

- (15) He felt tired. He left.
- (16) I felt tired. I left.
- (17) I feel tired. I'm going home.

4 Implicitness of participant

The more implicit the realization of the conscious participant in the discourse, the higher the SI. Examples (18) and (19) contain some explicit elements referring to the speaker's perspective. In Example (19) the modal adverb *probably* invokes the speaker as the source for the probability judgment. This is even more explicit in (18) where the speaker's perspective *I think* is more or less 'objectified'. This objectification appears from the fact that it may be referred to by demonstrative pronouns in subsequent utterances. E.g. *that is not true* would refer to the *I think* clause, not to its complement. At this sentence level, (19) has a higher SI than (18). Factual statements like (20) can also occur in a discourse context that heavily suggests an interpretation as a speaker conclusion. In this case the speaker's inferential activity is part of the relational interpretation, but at the same time it is maximally implicit. Hence this time it has a maximal degree of SI, higher than (19) and still higher than (18).

- (18) I think he is Hungarian.
- (19) He is probably Hungarian.
- (20) He is Hungarian ('cause he can only be Czech or Hungarian, and I know he's not Czech).

3.2 SI and Connectives

Causal relations can be marked explicitly by a connective. We claim that the Speaker Involvement Scale is not only a way to classify coherence relations but that connectives—as prototypical markers of coherence relations—can be analyzed in terms of the scale as well. In particular, the appropriateness of a connective in a given context can be measured in terms of SI, i.e., every connective *encodes a certain SI level*, which it contributes to the interpretation of its discourse environment. When this level is too low or too high to fit in the environing discourse, the use of the connective is inappropriate, as illustrated in (21-22).

- (21) Snow White died, ??*since/because* she ate a poisoned apple.
 (22) The queen had to be punished, *since/because* this is a fairy tale.

To uncover the semantic profile of a given connective as well as its interaction with the surrounding discourse, we make use of systematic corpus analyses combining distributional data and semantic intuitions. Thus in previous work (Degand and Pander Maat 1999, in press; Pander Maat and Degand 2001) we have demonstrated that the Speaker Involvement Scale can be used to analyze and classify forward and backward causal connectives in Dutch and French. With respect to the connectives *puisque* and *aangezien* we have shown that *puisque* occupies a higher SI area on the scale than *aangezien* (see also Degand, in press). By operationalizing the four SI characteristics mentioned above, we have shown that:

- there is no difference between the two connectives with respect to the involvement of a conscious participant in the causal relation;
- there is a significant tendency for *aangezien* to express causal relations that are iconic with respect to real world causality (nonvolitional, volitional, and causal epistemic relations), while *puisque* can also express relations which are noniconic (noncausal epistemic and speech-act relations). Hence *aangezien* expresses lower SI than *puisque* ($\chi^2 = 12.72$, $df = 5$, $p < 0.05$);
- there is a significant tendency for *puisque* to occur in relations expressing proximity with the present speaker (first-person participants), while *aangezien* does not show this proximity. Hence *puisque* expresses higher SI than *aangezien* ($\chi^2 = 8.491$, $df = 1$, $p < 0.01$);
- there is no difference between the two connectives with respect to the degree of implicitness of the participants involved in the causal relations.

According to Degand and Pander Maat, these findings on the basis of comparable (newspaper) corpora allow us to conclude that *puisque* is used to express causal relations with a higher level of SI than *aangezien*, even if the two connectives do also partially operate within the same SI area. These distributional divergences in comparable data could be an indication that the two connectives do, in fact, not form a ‘translation pair’, or only a partial one. To get insight in this latter matter further investigation of parallel data is needed. This would allow us to answer the question whether translators do consider *puisque* and *aangezien* to be translation equivalents, and if so whether this leads to under-, over- or misuse of one of the connectives in

given contexts. If the SI hypothesis is correct, translations of these connectives should respect their inherent SI profile, thus making use of connectives (or other discourse structures) respecting this SI level. Hence if the connectives display slightly diverging SI profiles as seems to be the case for *puisque* and *aangezien*, they should not always be translated by their ‘official counterpart’, but rather by linguistic structures which respect their SI level.

4 *Puisque* and *Aangezien* in Parallel Data

For the aim of this translation study I collected two parallel samples. The first one, PUISQUE-FR, contained fifty original occurrences of *puisque* in context, with its aligned Dutch translations. The second one, AANGEZIEN-NL, contained fifty original occurrences of *aangezien*, again with its corresponding translations. In building the samples I have tried to keep them as comparable as possible, however, some divergences in genre and register could not be avoided. The basis of the two samples were all occurrences of *puisque*, respectively *aangezien*, in the ‘Parallel aligned Namur Corpus’ (Paulussen 1999), a parallel French-Dutch-English corpus of European parliamentary debates.⁴ Because this did not result in sufficient occurrences, I supplemented the samples with bilingual data collected from the internet, i.e. information provided by the European Community on the rights and duties of Belgians emigrating to France or to the Netherlands,⁵ and information provided by the Belgian federal departments.⁶ To supplement the lack of *aangezien* occurrences I consulted Belgian federal press releases as an additional source of information.⁷ Table 2 gives an overview of the sources used and the number of occurrences of each connective drawn from the different sources.

⁴ I would like to sincerely thank Hans Paulussen for giving me access to these data and helping me to collect the parallel data for a number of Dutch and French causal connectives.

⁵ Since this information is provided by French and Dutch authorities respectively, I have considered that the French version was the original one in the first case, and Dutch was the original version in the second case. The URL of the internet site is <http://citizens.eu.int/>.

⁶ Information provided by francophone departments (for instance, at the time of collection, the departments of finance, transport, internal affairs, and so forth) have been considered as originally written in French and translated into Dutch, while information provided by Dutch-speaking departments (prime minister, social affairs, justice, and so forth) have been considered as originally written in Dutch and translated into French (see http://www.belgium.be/fr_index.htm).

⁷ Again, only press releases from Dutch-speaking departments have been used (http://feiten.fgov.be/cgi/sga_cnct/sfi?LANG=FR).

PUISQUE-FR (n=50)	AANGEZIEN-NL (n=50)
European parliamentary debates (n=32)	European parliamentary debates (n=9)
Information for Belgians in France (n=2)	Information for Belgians in the Netherlands (n=8)
Information from Francophone federal departments (n=16)	Information from Dutch-speaking federal departments (n=16)
	Federal press releases (n=17)

Table 2. Sources used to build the parallel samples

The next step consisted in comparing the semantic profile of the connectives in the newspaper corpora with that of the connectives in the parallel data, in order to see whether the observations concerning *puisque* and *aangezien* could be generalized. The analysis of the original data in the two samples (*puisque* in PUISQUE-FR, and *aangezien* in AANGEZIEN-NL) shows the same significant divergences between the two connectives on the different SI features: identity of the causal participant ($\chi^2 = 8.122$, $df = 3$, $p < 0.05$), iconicity with real world causality ($\chi^2 = 19.539$, $df = 6$, $p < 0.005$), and linguistic realization of the causal participant. These analyses confirm that *puisque* expresses a higher level of SI than *aangezien* in these two different genres.

4.1 Translation Equivalence Between *Puisque* and *Aangezien*

Do the parallel data provide us with evidence that the connectives *puisque* and *aangezien* are a translation pair? In order to answer this question, I collected the translations of the two connectives in the two samples PUISQUE-FR and AANGEZIEN-NL. Results are given in Table 3.

Puisque (n = 50)		Aangezien (n = 50)	
<u>aangezien</u>	<u>24</u>	étant donné que (given that)	14
omdat (because)	8	vu que (seeing that)	8
immers (after all, indeed, for)	5	present perfect	8
want (because/for)	3	comme (as)	6
daar (as, seeing that - formal)	2	<u>puisque</u>	<u>4</u>
dus (so)	1	en effet (indeed)	3
namelijk (you see, as it happens)	1	vu + NP (considering + NP)	2
causal NP	1	parce que (because)	1
		dès lors que (seeing that)	1
		du fait que (because of the fact that)	1
not translated	<u>5</u>	not translated	2

Table 3. Translations of *puisque* and *aangezien* in the parallel samples

What jumps out at the viewer in Table 3 is that the *aangezien-puisque* pair can NOT be considered as a translation pair in the AANGEZIEN-NL sample. Only 8% of the occurrences of *aangezien* are translated by *puisque*, the preferred translations being the conjunctive expressions *étant donné que* (28%) and *vu que* (16%),⁸ and (causal) present perfect constructions (16%). This is exemplified in (23-25):

- (23) L'Union Européenne ne possède pas de vraies ressources propres, **étant donné qu'**elle est pour la plus grande part financée par des contributions basées sur le PNB. (Parliamentary debates, translation from Dutch original *aangezien*.)
The European Union does not possess real proper resources [étant donné que] it is financed in the majority by contributions based on the BNP.
- (24) Ce montant doit encore être adapté, **vu qu'**un certain nombre de dossiers seront encore corrigés ou refusés. (Federal departments, translation from Dutch original *aangezien*.)
This amount needs to be adapted [vu que], a certain number of files still have to be corrected or refused.

⁸ As was pointed out to me by an anonymous reviewer, 'gezien' in 'aangezien' means 'vu' ('seen') as in 'vu que'. This iconicity in form and meaning could well explain this translation alternative.

- (25) La cour de cassation **étant** saisie d'un plus grand nombre d'affaires en langue néerlandaise, les arrêts en cette langue sont plus nombreux. (Federal departments, translation from Dutch original *aangezien*.)
The Supreme Court of Appeal being referred to more frequently for matters in the Dutch language, the judgments in that language are more numerous.

Things appear to be different for the *puisque-aangezien* pair, since in the PUISQUE-FR sample, *puisque* is translated in the first place by *aangezien* (48%), followed by the conjunction *omdat* (because) (16%), the adverb *immers* (after all, indeed) (10%), or an implicit construction (10%) (see Examples (26-29)).

- (26) Tot slot kunnen ook de markten verstoord raken, zowel binnen de Gemeenschap als over de gehele wereld, **aangezien** bepaalde bijproducten belangrijker dreigen te worden dan de hoofdprodukten. (Parliamentary debates, translation from French original *puisque*)
Finally markets could get disturbed, within the community and all over the world, [aangezien] a number of secondary products threaten to become more important than primary products.
- (27) Het hoger onderwijs in Luxemburg is nogal uitzonderlijk **omdat** het momenteel slechts een beperkt aantal richtingen omvat en in een klein aantal onderwijsinstellingen wordt gedoceerd. (Federal departments, translation from French original *puisque*)
Higher education in Luxemburg is rather exceptional [omdat] insofar as it comprises only a restricted number of trainings and teaching finds place in a small number of institutions.
- (28) Bovendien gaat het niet om kleine, maar om zeer grote bedragen. De structuurfondsen vertegenwoordigen **immers** een groot gedeelte van onze begroting. (Parliamentary debates, translation from French original *puisque*)
In addition this does not concern small but very big amounts. The structural funds represent [immers] a large part of our budget.
- (29) Indien dat het geval zou zijn, waartoe zou het internet dan dienen als men er niets mee mag aanvangen? (Federal departments, translation from French original *puisque*)
If this were the case, what would be the use of the internet if you're not allowed to use any of it?

Thus, it is as if *aangezien* were the translation equivalent of *puisque*, whereas *puisque* would not be the counterpart of *aangezien*. In order to investigate this asymmetry, I did an additional (independent) SI analysis of the translations of each of the fifty original occurrences of *puisque* and *aangezien*. This analysis should tell us whether the SI level of the original discourse environment is preserved during the translation process. Taking into account the findings on the original data, we know that *puisque* expresses a higher level of SI than *aangezien*. The matter at stake here is what happens with these two connectives in translation?

4.2 SI of Original and Translated Discourse Environments

Figures 1 and 2 display the relational context in which *puisque* and *aangezien* occur, along with their translations. The comparison of the SI level of the original and translated segments shows very clearly that the level of SI is preserved during translation.⁹ The other aspects intervening in calculating the SI level (identity of the participants, explicitness of the participant) do all equally result in nonsignificant differences. Of course, Figures 1 and 2 also show that there is not a perfect one-to-one convergence between the SI level of the original segments and the SI level of the translated segments, but from a global point of view there are no significant differences.

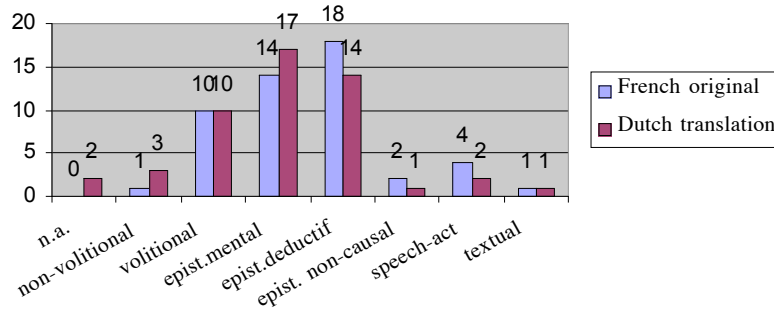


Figure 1. Relational discourse context of fifty occurrences of *puisque* and its Dutch translations ($\chi^2 = 4.790$, $df = 7$, $p = 0.686$ (N.S.))

⁹ Two categories need a word of explanation. The category n.a. refers to those cases where the translation has made it impossible to keep a causal relation between two distinct segments. The category epistemic mental is the result of a further evolution of our work on causal interpretation.

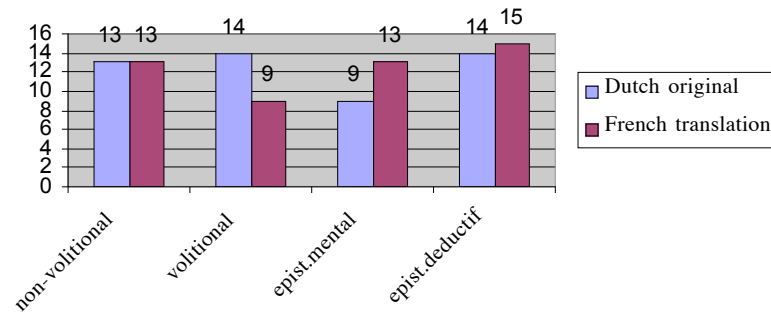


Figure 2. Relational discourse context of fifty occurrences of *aangezien* and its French translations ($\chi^2 = 1.849$, $df = 3$, $p = 0.604$ (N.S.))

While these figures seem to show that the level of SI is preserved during translation, the question remains whether the connectives themselves are used in a similar way in the original and translated contexts. To discover this, I would like to compare the discursive contexts in which the original and translated connectives are used. Since my data comprise only four occurrences of *puisque* in translation, I will restrict my investigation here to the fifty original occurrences of *aangezien* which will be contrasted with the twenty-four occurrences found in the translations of the sample PUISQUE-FR.

4.3 Semantic Analysis of Original and Translated *Aangezien*

Contrasting the type of discourse environment in which original and translated *aangezien* occurs leads to the conclusion that this connective is not used in a similar way in translation. Original *aangezien* shows a preference for low SI contexts, in particular nonvolitional and volitional relations (cf. Figure 2). This is not the case for *aangezien* in translation ($\chi^2 = 10.031$, $df = 4$, $p < 0.05$) (Figure 3).

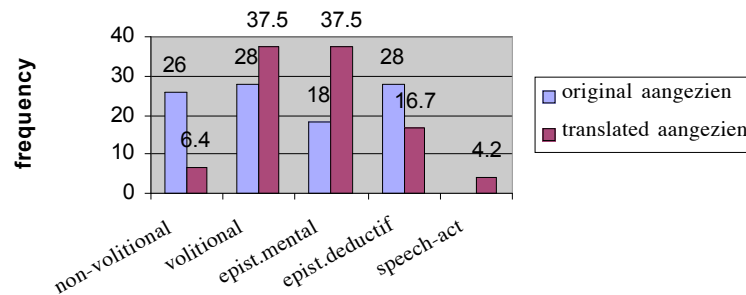


Figure 3. Relational context of original and translated *aangezien*

Apparently translated *aangezien* is used in a higher SI context. A plausible explanation for this phenomenon can be found when looking at the original French data. It has indeed been shown that original *puisque* is used in rather higher SI contexts (Figure 1), which is not the case for *aangezien*. Nevertheless, when *puisque* is translated into Dutch, the translator's choice goes in the first place to the «counterpart» *aangezien*, which results in an **overuse** of *aangezien* in these higher level contexts. This overuse can be explained in terms of a transfer from the discursive context of *puisque* to that of *aangezien*. In terms of the SI scale, it is as if the translation process «pushes» *aangezien* to the right side of the scale, i.e. towards the more subjective areas of the scale. This is illustrated in Examples (30) and (31).

- (30) French (original) Le graphique montre que le niveau global moyen de formation s'améliore, **puisque** le niveau d'instruction des classes d'âge plus jeunes est supérieur à celui de l'ensemble de la population. (Federal departments)
 Dutch (translation): Uit de grafiek blijkt duidelijk dat het gemiddeld opleidingspeil verbetert, **#aangezien** het scholingsniveau van jongere leeftijdscategorieën beter is dan dat van de totale bevolking.
 The graph shows that globally the mean level of education is improving [puisque/#aangezien] the education level of younger people is higher than that of the whole population.
- (31) French (original): Je voudrais donc savoir, M. le Président, ce que vous pensez faire, **puisque** je vous rappelle qu'à l'article 117 de notre règlement, il est précisé que le vote est individuel et personnel ... (Parliamentary debates)

Dutch (translation): Mijnheer de Voorzitter, ik zou dus willen weten wat u van plan bent te doen, #**aangezien** ik u in herinnering breng dat artikel 117 van ons Reglement bepaalt dat het stemrecht een persoonlijk recht is ...

Mister President, I would like to know what are your plans [puisque/#aangezien]. I would like to remind you that article 117 from our regulation stipulates that voting is an individual right ...

With respect to the other features of SI (identity of the participant, linguistic realization, and so forth) no significant divergences could be established between the original and the translated data. Actually, while one could speak of an overuse of translated *aangezien*, Figure 3 shows that one should rather speak of an **underuse** of translated *aangezien* in low SI contexts, in particular in nonvolitional relations. This phenomenon of underuse finds its straightforward explanation in the original French data: *puisque* is only very rarely used to express lower SI relations. A better candidate to express non-volitional and volitional (backward) causal relations in French is the connective *parce que* ('because') (Degand and Pander Maat, in press); on the basis of the translations of *aangezien* used in a low SI context (54% according to Figure 3), *étant donné que* is another plausible candidate for the expression of objective contexts, beside the use of present perfect constructions and the connective *comme* as is shown in Table 4. To verify this hypothesis, I analyzed fifty original occurrences of *étant donné que* in a newspaper corpus. The relational distribution contrasted with that of *aangezien* is given in Figure 4.

French translations	original non-vol. and vol. <i>aangezien</i>
<i>present perfect</i>	7 (25.9%)
<i>étant donné que</i> ('given the fact that')	6 (22.2%)
<i>comme</i> ('as')	5 (18.5%)
<i>en effet</i> ('indeed, in effect')	3 (11.1%)
<i>parce que, dès lors que</i> ('because, as soon as')	1

Table 4. French translations of non-volitional and volitional *aangezien*

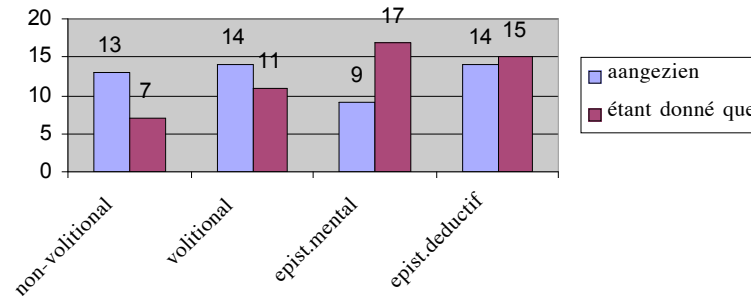


Figure 4. Relational discourse context of fifty occurrences of *aangezien* and *étant donné que* ($\chi^2 = 4.65$, $df = 3$, $p = 0.20$ (N.S.))

Figure 4 shows that *étant donné que* is indeed a better translation alternative for *aangezien* than *puisque*. The two connectives do not significantly diverge from one another with respect to their relational context. In other words, the two connectives tend to express rather low SI relations. Other alternatives exist, however, in French to express low SI contexts. In particular, Table 4 seems to suggest that present perfect constructions are frequently used to present information in an objectified way. Further research should, however, confirm this.

5 Conclusions

One of the aims of this study was to illustrate how combining comparable and parallel corpus analyses could benefit contrastive linguistics. The comparable data on *puisque* and *aangezien* made it possible to uncover semantic convergences and divergences between Dutch and French with respect to the use of these two connectives. The parallel data gave rise to a set of plausible translation alternatives for the two connectives, thus partially correcting existing descriptions. At the same time, these data made it possible to discover cases of over- or underuse of given structures in given contexts.

In particular, the study of the *puisque-aangezien* 'pair' in terms of Speaker Involvement has shown that this 'pair' is not a translation pair, contrary to what translation dictionaries would predict, but in line with what the comparable data had predicted. Both in newspaper and parliamentary debates corpora, the two connectives *puisque* and *aangezien* have been found to diverge drastically in their semantic SI profile. This led to my hypothesis that the two connectives would not be translated by one another in all dis-

course contexts. This appeared to be the case in particular for *aangezien* which was translated by *puisque* in only 8% of the cases. The translator seems to respect in the first place the level of Speaker Involvement of the translated segments, thus having recourse to a wide variety of translation alternatives. Among these translation alternatives, *étant donné que* ('given the fact that') appeared to be a more plausible translation equivalent of *aangezien* than *puisque*, since the two connectives seem to express similar relations, with a fairly low level of SI. These findings should, however, be confirmed by further analyses both on comparable and parallel data, since present perfect constructions also seem to express these lower level SI contexts in French. Access to a larger French-Dutch parallel corpus should also permit us to investigate what is happening with *puisque* in translation. This connective is used in original French texts to express relations with a fairly high level of SI. From earlier work, we know that this kind of relation is expressed in Dutch by the connective *want* ('for/because') (Degand and Pander Maat, in press). If we could have access to more extensive parallel data, we could systematically investigate how high and low SI constructions are translated, which could lead to 'contextualized' or discourse-bound translation pairs. Working with parallel data is, however, not free of problems. Cases of over- and underuse show that the influence of the source language can be hiding everywhere.

References

- Baker, M. 1993. Corpus Linguistics and Translation Studies. Implications and Applications. *Text and Technology*, eds. M. Baker, G. Francis, E. Tognini-Bonelli, 233-50. Amsterdam/Philadelphia: John Benjamins.
- Bogaards, P. et al. 1998. *Van Dale Groot woordenboek: Frans-Nederlands/Nederlands-Frans* (Third printing). Utrecht/Antwerpen: Van Dale Lexicografie.
- Coenders J. et al. 1993. *Standaard Handwoordenboek Frans-Nederlands / Nederlands-Frans*. Antwerpen: Standaard Uitgeverij.
- Degand, L., 1998. Het ideationele gebruik van *want* en *omdat*: een geval van vrije variatie? *Nederlandse Taalkunde* 3: 309-26.
- Degand, L. 2000. Contextual constraints on causal sequencing in informational texts, *Functions of Language* 7: 173-201.
- Degand, L. in press. De l'analyse contrastive à la traduction: le cas de la paire *puisque* – *aangezien*. *Actes des 2èmes journées de la linguistique de corpus*, ed. G. Williams. Reims: P.U.R.
- Degand, L., and H. Pander Maat. 1999. Scaling causal relations in terms of Speaker Involvement. *Levels of Representation in Discourse, Working Notes*

- of the *International Workshop on Text Representation*. Edinburgh University. 45-54.
- Degand L., and H. Pander Maat. In press. A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. *Usage-Based Approaches to Dutch*, eds. A. Verhagen, and J. M. van de Weijer, LOT International Series. LOT: Utrecht.
- Degand, L., and T. Sanders. 1999. Causal connectives in language use. Theoretical and methodological aspects of the classification of coherence relations and connectives. *Levels of Representation in Discourse, Working Notes of the International Workshop on Text Representation*, Edinburgh University, 3-12.
- Delbey, A. 1988. Les connecteurs car- puis – puisque et la justification en ancien français. *Revue de linguistique romane* 52: 397-419.
- Dijk, T.A. van. 1979. Pragmatic connectives. *Journal of Pragmatics* 3: 447-56.
- Franken, N. Pour une nouvelle description de puisque. *Revue Romane* 31: 3-18.
- Granger, S. 2003. The corpus approach – a common way forward for contrastive linguistics and translation studies. *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, eds. S. Granger, J. Lerot, S. Petch-Tyson. Amsterdam/Atlanta: Rodopi.
- Grootaers, L. 1979. *Nieuw Nederlands Frans woordenboek*. Leuven, Brussel: Vander Uitgever.
- Groupe -1, 1975. Car, parce que, puisque. *Revue Romane* 10: 248-80.
- Halliday, M. A. K., and R. Hasan. 1976. *Cohesion in English*. London: Longman.
- Hensen, I. 1976. Quelques réflexions sur le mot puisque et la notion d'adverbe de phrase. *Studia Neophilologica* 28: 152-54.
- Herckenrath, C. R. C., and A. Dory. 1985. *Wolters' Woorden boek Nederlands/Frans; Frans/Nederlands* (2 volumes), Groningen: Wolters-Noordhoff.
- Iordanskaja, L. 1993. Pour une description lexicographique des conjonctions du français contemporain. *Le Français moderne* 2: 159-90.
- Johansson, S., and J. Hasselgård. 1999. Corpora and cross-linguistic research in the Nordic countries. *Contrastive Linguistics and Translation, Special Issue of Le Langage et l'Homme* 34 (1), eds. S. Granger, L. Beheydt, J-P. Colson, 145-62. Leuven: Peeters.
- Larousse. 1983. *Groot Nederlands Larousse Woordenboek Frans-Nederlands/Nederlands-Frans*, Hasselt/Utrecht: HeideLand-Orbis and Schel-tens-Giltay.
- Martin, J. R. 1992. *English Texts: system and structure*. Amsterdam/Philadelphia: John Benjamins.
- Martin, R. 1973. Le mot puisque – Notions d'adverbe de phrase et de présupposition sémantique. *Studia Neophilologica* 45: 104-14.

- Moeschler, J. 1989. Pragmatic connectives, argumentative coherence and relevance. *Argumentation* 3: 321-39.
- Olsen, M. 2001. Puisque: Syllogisme caché. *Revue Romane* 36 : 41-58.
- Pander Maat, H., and L. Degand. 2001. Scaling causal relations and connectives in terms of Speaker Involvement. *Cognitive Linguistics* 12: 211-45.
- Pander Maat, H., and T. Sanders. 1995. Nederlandse causale connectieven en het onderscheid tussen inhoudelijke en epistemische coherentie-relaties. *Leuvense Bijdragen* 84: 349-74.
- Paulussen, H. 1999. A corpus-based contrastive analysis of English 'on/up', Dutch 'op' and French 'sur' within a cognitive framework, Doctoral dissertation, Universiteit Gent.
- Sweetser, E. 1990. *From Etymology to Pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: C.U.P.
- Van Belle, W. 1989. Want, omdat, en aangezien: een argumentatieve analyse. *Leuvense Bijdragen* 78: 435-556.
- Van Dale, 1997. *Woordenboek Nederlands-Frans/Frans-Nederlands op CD-ROM*. Leuven: Van Dale.
- van Kooten, G., ed. 1982. *Kramers' Frans-Nederlands/Nederlands-Frans woordenboek*, Amsterdam/Brussel: Elsevier.