

Adversative discourse markers in contrast

The need for a combined corpus approach*

Liesbeth Mortier and Liesbeth Degand

Université catholique de Louvain (UCL)

This paper deals with the semantics of two discourse markers, viz. French *en fait* (“in fact”) and Dutch *eigenlijk* (“actually”), commonly associated with the expression of “opposition” and “reformulation”. A special focus lies on methodological issues in the description of such markers, since their non-propositional meanings seem to require what is called a ‘combined corpus approach’, including written and spoken comparable data as well as translation corpora. It is argued that *eigenlijk* and *en fait* are best described as adversatives, at the intersection of “opposition” and “reformulation” which constitute their basic meanings, and from which other meanings such as “causality”, “counterexpectation”, “enhancement” and “attenuation” can be inferred. Evidence from all sets of corpora moreover suggests that it is the semantic underspecification of *en fait* and (especially) *eigenlijk* which ultimately accounts for their high level of polysemy.

Keywords: discourse markers, adversativity, translation corpora, comparable corpora

1. Introduction

Bringing the semantics of discourse markers to light is a complex task requiring both subtle linguistic intuitions and a sound methodology in support of these intuitions.¹ The aim of the present study is twofold: on the one hand, we want to uncover the semantics of a set of adversative discourse markers; on the other hand, we will pay special attention to the advantages of combining different types of corpora in the study of discourse markers. Both objectives are inevitably intertwined, since a well thought-out methodology will facilitate semantic description, whereas the resulting semantic profile will reflect the richness of the methodologies used.

A first focus will be on the French marker *en fait* (“in fact”) and what will be shown to be its most salient equivalent in Dutch, viz. *eigenlijk* (“actually”), both

endowed with a complex semantic profile and taking up a particular place within the range of adversative markers. The need for such an account is legitimated by the lack of substantial research into their semantics, especially so for Dutch *eigenlijk*, which remains undiscussed even by the most important reference grammar of Dutch (Haeseryn et al. 1997) in spite of its high frequency.

How one establishes such a semantic profile through corpus analysis is a second topic of interest. Methodology thus represents a research objective in itself, starting from the idea that a combination of different types of corpus material as well as a diversified set of exploitation techniques may benefit linguistic description. As such, we will explore the (dis)advantages of combining a “classical” analysis of written comparable data, i.e. data similar in genre and quantity of words over more than one language, with a discussion of comparable spoken data as well as of translation corpus data. The combination of these three data types allows for interesting perspectives: not only can one contrast languages, either in translation or in texts of a comparable nature and size, but deviations between genres and between the written and spoken register are expected to surface as well. Depending on the nature and size of the data available, a variety of approaches is conceivable. Here, we will use the written comparable data for a micro-analysis of semantic features, with parameters being equal or similar in both languages, whereas translation data allow for a mirror analysis, i.e. the creation, through back-and-forth translations (cf. Aijmer & Simon-Vandenberg 2004, also Section 3), of a formal map of markers in the semantic realm of the marker studied. Finally, our choice to take large-scale, comparable spoken data into account, is motivated by their potential to shed light on the frequency and vitality of tendencies and features observed in the analyses of translation and comparable written data.

The outline of this paper is as follows: in Section 1, we will take a closer look at previous research on the markers focused on, however scarce, by discussing the main values associated with them. Section 2 will account for our methodological choices, especially the interest taken in translation corpora as a device for retrieving information on semantics and formal diversity. The empirical study “proper” will be dealt with in Section 3 of this paper and will aim at establishing semantic profiles for *en fait* and *eigenlijk* by combining data from all three sets of corpora.

2. *En fait* an adversative?

French *en fait* has been described in the literature as both a connective and what is called an “embrayeur d’intervention” by Danjou-Flaux (1982) and Rossari (1992), i.e. a marker by means of which the speaker signals his intervention in a discursive sequence, or, in short, a discourse marker. Many different semantic functions

have been associated with *en fait*, among which the expression of “opposition”, of “counterexpectation”, and of “reformulation” are most salient (Danjou-Flaux 1980, 1982; Rossari 1992). Two main issues are found to be relevant in the discussion of the semantics of *en fait*:

- i. Does *en fait* express “contrast” or “concession”, and what is the contribution of the term ‘adversativity’? (Section 2.1)
- ii. Does the semantics of *en fait* extend to other domains, such as “additivity” and “causality”? (Section 2.2)

For the larger part, the current literature on these matters involves not French *en fait*, but its English counterpart *in fact*, usually dealt with in discussions of subjectivity as a feature of discourse markers. Although the etymological affinity between *en fait* and *in fact* is not a sufficient reason to make generalizations about their common behavior, it is likely that some trends observed for *in fact* apply to *en fait* as well, or constitute at the very least a starting point for our own analysis of the French marker. On a genealogical level, both markers are undoubtedly related: *fact* is derived from Indo-European **dhe-* (do), was borrowed from Latin FACTUM in the 16th century and started functioning as an adverbial with the preposition *in* in the 17th century (Traugott & Dasher 2002: 165); *fait* was borrowed from Latin around 1268, but probably took on its use as unintegrated adverbial (*en fait*) only in the 17th or 18th century (TLF-i).²

2.1 Contrast vs. concession vs. adversativity

English *in fact* has been described extensively in the literature (e.g. Traugott 1999, Schwenter & Traugott 2000, Traugott & Dasher 2002, Oh 2000, Aijmer & Simon-Vandenberg 2004), usually as pertaining to the class of pragmatic markers. *In fact* figures in descriptions of both discourse markers and metadiscursive markers, suggesting a complex polysemy and an ambiguous status in the language. Hyland (1998: 442), for instance, considers *in fact* an emphatic marker of interpersonal metadiscourse, i.e. a marker expressing the speaker’s stance towards the discourse, in this case the emphatic force or the speaker/writer’s level of certainty in a message. Other research focuses on concepts such as ‘(counter-)expectation’ and ‘adversativity’ in relation to *in fact*, based on the idea that speakers, for various reasons, explicitly position themselves and their utterances vis-à-vis expectations raised by preceding discourse or by background knowledge (Aijmer & Simon-Vandenberg 2004).

The matter of how to label the core meaning of *in fact*, and by extension that of *en fait*, is a complex one, on which no clear consensus can be found. ‘Adversativity’, defined by Schwenter (2000: 259) as “contrast between different points of view as

these are constructed in language use”, is a recurring concept in most discussions, but is itself classified as belonging to different semantic domains. Schwenter (ibid.), for example, links up “adversativity” with “contrast”, whereas Malchukov (2004) sees a connection with “concession”, of which adversatives would be a subclass.

Let us first have a look at what the semantic domains of “contrast” and “concession” are about. König (1988: 146), in his cross-linguistic analysis of concessives, claims that, in producing utterances of the sort exemplified in (1), the speaker commits himself to the truth of the two propositions linked by *although*:

- (1) Although it is raining, I am going out for a walk.
“although q, p”

More precisely, the speaker validates both propositions against the background assumption that what is asserted in p and q is in fact incompatible. In terms of ‘mental spaces’, the use of concessive markers instructs the hearer to set up two distinct mental spaces, one in which the causal inference “p therefore q” is valid and one in which the truth of q is denied so that the causal inference is not valid (Verhagen 2000). Crucially, then, “concession” is characteristic of situations in which “someone acknowledges that in highly similar circumstances a mind very similar to one’s own draws a valid causal inference, while this inference is actually not valid” (ibid.: 367).

“Contrast” is of a different nature. According to Schwenter (2000: 260), “contrast”, and specifically “adversativity”, forces an interpretation in the hearer in which the speaker explicitly asserts (supposed) incompatibility between p and q and indicates his viewpoint as the only relevant. In (2), taken from Schwenter (2000: 260), the use of *but* forces the hearer to consider perceived incompatibility between shortness and basketball, at the same time urging him to disregard the implications of that physical feature in the activity:

- (2) John is short, but he’s a good basketball player.

In the case of “contrast”, thus, the speaker’s viewpoint is presented in the context of a competing viewpoint which, via inference, is excluded from further consideration. Unlike “concession”, “contrast” does not presuppose the validity of “p therefore q” in a distinct mental space, and infers an assumption which is in contrast with the two statements being presented side by side (Lang 2000: 244).

It seems that, in order to determine what place to attribute to adversatives (in occurrence, *in fact* and *en fait*) in this field, another concept should be taken into account, viz. that of ‘counterexpectation’, “used for propositions that present counters to normative viewpoints characterizable as ‘people say/think that X’” (Traugott & Dasher 2002: 157). The concept of ‘counterexpectation’ is sometimes used to refer to the “novelty factor” which is so fundamental to most discourse and

relates to the concepts of ‘(ad)mirativity’, i.e. the expression of unexpected or new information to the speaker, ‘evidentiality’, i.e. information obtained by inferencing, hearsay or which is marked for source of evidence, and ‘modality’, i.e. information the speaker is completely confident about (DeLancey 2001). “Counterexpectation” is often mentioned as a core feature of *in fact* (e.g. Traugott 1999, Schwenter & Traugott 2000, Traugott & Dasher 2002), more precisely as yielding a contrast between the speaker’s viewpoint and an external, normative viewpoint. It seems to be compatible with the domains of both “contrast” and “concession”, although its assessment is relatively different in both cases:

- (3) Although measures are being put into place, the situation continues to deteriorate.
- (4) Measures are being put into place, but the situation continues to deteriorate.

In (3), the authoritative power of “although q” is acknowledged, while the speaker observes that the inference “measures thus improvement” is invalid. In (4), the inference of “counterexpectation” is stronger, with the speaker focusing on the invalidity of the expectation raised by p, which consequently needs to be further disregarded. It seems obvious, then, to allocate adversatives with a strong sense of “counterexpectation” to the category of “contrast”; however, it has been noted for *en fait* that the core meaning is not one of strict “opposition”, but rather one of a deviation between p and q, i.e. accommodating for a sense of vagueness in the oppositive force (Rossari 1992).

One may tentatively conclude that *in fact*, and *en fait*, bridge the domains of “concession” and “contrast”, provided that “counterexpectation” is a (subtle) part of their interpretation. Potential meanings pertaining to the domain of “contrast/concession”, put forward by previous research and to be verified in our own research, thus include:

- **Strict opposition:** equaling simple “contrast” (Fraser 2006: 76–81), “contradiction” or the mutual exclusion of two possibilities, and the idea of “contrary” with situations being opposites (Lewis 2006).
- **Deviation:** or « écart », cf. Rossari (1992: 159).
- **Counterexpectation:** (Traugott 1999, Schwenter & Traugott 2000, Traugott & Dasher 2002).

In the remainder of this paper, we will refer to *en fait* and its equivalents as ‘adversatives’, with the understanding that their semantics provides for more subtle values than a strictly oppositive one. ‘Adversativity’ here refers, then, to the common ground found for the meanings of “contrast” and “concession”, and not merely to “adversativity” as a subclass of “contrast” (as suggested by Schwenter 2000: 260).

2.2 Reformulation, additivity, causality

Another argument prompting us to put the oppositive strength of *en fait* and its equivalents into perspective is their ability to express additive meanings of “reformulation”. As Traugott & Dasher (2002: 157) note, *in fact* may be used to signal that the upcoming proposition is a more precise formulation than the previous one. In (5), the *in fact* proposition is additive rather than adversative:

- (5) Humanity is not often present. *In fact*, it is usually absent.

“Reformulation”, be it a more precise, more detailed or more correct formulation of something stated previously, relates in itself to “deviation”: pending the corrective power of the reformulation, q will provide a more or less important shift away from p, without there being strict opposition *per se*.

We hypothesize here that in some cases there may be a causal aspect to this meaning of “reformulation”: when something is stated in more detail, one will frequently find a causal connection, i.e. either an explanation (“cause”) for p or a consequence of p, between two propositions. In (6), for example, the *en fait* proposition does more than adding information to the p proposition; it helps explain the novelty factor of the two forms of recreation in Belgium:

- (6) Ces deux formes de récréation sont relativement neuves en Belgique et ont **en fait** été propagées par les Pays-Bas.
« These two forms of recreation are relatively new in Belgium and have *en fait* been propagated by the Netherlands.”

Values relating to “additivity” which will be taken into account from previous research and for our present research thus relate to:

- (more precise, detailed, correct) **reformulation** of (part of) p or q
- **Causal relations** (cause, consequence) between p and q

When emphasis lies on the reformulative meanings in the remainder of this paper, we will refer to *en fait* and its equivalents as ‘reformulatives’.

3. A combined corpus approach

One of our major objectives in this paper is to show that the study of discourse markers benefits most from a combined corpus approach, i.e. the taking into account of different types of corpora, which each have their strengths and may help shed light on the semantic/pragmatic profile of the marker(s) as a whole. While such a cross-corpus approach is not new in itself, we aim to show the specific

benefits of different corpus types and of the ways to approach them, with special reference to the markers under investigation. Moreover, most recent contributions to corpus linguistics do not explicitly combine different types of corpus material, and restrict themselves either to translation corpora or to comparable data (e.g. Granger et al. 2003, Johansson & Oksefjell 1998, Lindquist & Mair 2003).

Apart from the analysis of written data (Section 3.1) and collocations in spoken data (Section 3.2) in Dutch and French, this study will take up the innovative trend in present-day corpus linguistics to include translation corpora (Section 3.3) in the analysis as well. Issues relating to data collection (Section 3.4) further determine what we would like to call here a 'combined corpus approach'.

3.1 A comparable corpus analysis of written data

The main part of the analysis is based on a more or less traditional study of self-compiled comparable data, i.e. corpora in two languages, namely French and Dutch, with a similar distribution over genres. A total of 50 occurrences per marker were subjected to a thorough semantic analysis, the aim of which was to determine the nature of *p* and *q*, respectively, and of the relationship between *p* and *q*, with and without the presence of the marker focused on. It was assumed that *q* contains the marker, while *p* is generally the proposition immediately preceding the one containing *q*, although this turned out to vary depending on the particular context. Parameters taken into account relate primarily to semantic features, such as the fact-opinion-action nature of *p* and *q*, their ideational relation and the effect of the marker on this relation, as will be discussed in more detail below (cf. Section 4.2.2).

3.2 Collocations in comparable spoken data

Spoken comparable data were taken into account specifically to test and validate certain tendencies observed in the analysis of written comparable data. We opted for large existing corpora of spoken data, focusing on collocations commonly found for *en fait* and its equivalents.

The inclusion of spoken data is particularly useful in the study of discourse markers such as *en fait* and *eigenlijk*. "Orality" is indeed one of the recognized properties of discourse markers (Schourup 1999), so that studying them in their most "natural" context comes as a prerequisite. Speakers' intentionality can only be traced back to the context of communication itself, i.e. in the interaction between speakers who invariably negotiate meaning. This explains why discourse markers naturally and abundantly turn up in spontaneous conversation, where they function as clues to the hearer on how a message is to be understood. Even

if, as a defining feature, discourse markers often have little or no propositional content (e.g. Brinton 1996; Fraser 2006; Hansen 2005), their meaning of use can be uncovered in this context of communication (cf. Section 4.2.4).

However, the analysis proposed in Section 4 will not deal with spoken and written corpora on an equal level of importance. While the written data will be subjected to a fine-grained, qualitative analysis in terms of a set of fixed parameters, the oral data are merely taken into account with regard to what they tell us about frequent collocation patterns for *eigenlijk* and *en fait*. It is thus only fair to state that the corpus study to be presented is primarily interested in written data; spoken data are relevant only insofar as they (in)validate certain tendencies in the written register.

3.3 Translation in linguistics

The choice to include translation data was motivated by the recent interest invested in translation as a heuristic for discovering contextual dimensions and for laying bare semantic fields, i.e. “a large, potential ‘sense’ which is not necessarily the sense of one sign, but rather the joint ‘sense’ of a set of semantically related signs” (Dyvik 1998: 60). In the context of discourse markers, Aijmer et al. (2006: 111–112) have argued that such semantic fields can be established by checking back and forth, i.e. from source language to target language and the other way around, the functional equivalents which can be assigned a correspondence value. For example, if X in language A is translated by Y and Z in language B, then one can, by using B as a source language, check for the translation of Y and Z in A, which thus becomes the target language. Such an analysis would “allow us to show how the pragmatic marker X is related to other pragmatic markers, or to other linguistic items such as modal particles or response words, in the same language” (ibid.: 112).

Problems surrounding the use of translation corpora are predictably manifold: not only is there a problem of context and typological differences, one should also be careful not to generalize individual instances of language use. For example, one should do well to take into account the potentiality of a mismatch between the writer’s intention and the reader’s interpretation, especially when writers, readers and translators constitute an intricate web of communicators. However, in the area of lexical semantics, it has been argued that it is precisely the impossibility of perfect, one-to-one translations which makes translation so interesting for the study of meaning. As Dyvik (1998: 7) states, “translators have no theoretic concern in mind, evaluate the interpretational possibilities of linguistic expressions (...), and then try to recreate the same interpretational possibilities in a target text serving a comparable purpose in another language”. Jaszczolt (2003), who is concerned with defining the semantics/pragmatics interface, deems translation an exponent

of communication, for which no other model than the general code model of communication (cf. Gutt 1998) is needed. Ultimately, a translation approach seems to qualify for most of the demands of contemporary linguistics: it is corpus-based, it is contrastive and thus has typological relevance, but most of all, it is task-based, in as much as it treats translation data as a collection of informants' judgments about the meanings of the linguistic forms in the source text (cf. Noël 2003:759).

Although more than one application of translation is conceivable in linguistic research, we will restrict its scope to what we would like to call a "mirror analysis", which takes "back-and-forth" translations as a way of establishing the field of formal equivalents in one language or across languages. This will help us not only determine what is the most adequate Dutch equivalent for French *en fait*, but also what semantic values can be associated with each marker.

3.4 Data collection

An important aspect of doing corpus linguistics, especially when different types of material are used, involves the way in which data are collected. Size, genre issues and quality of data (especially in translation) as well as of data transcription (for spoken material) are parameters to be taken into account during data collection. Working with more than one language further complicates the constraints imposed on their assemblage, since homogeneity in genre and size is required from one language to another as well.

For the written comparable data, we collected texts from newspapers (56% of all data), literary novels (30%) and essays on literature and writing (14%), leading up to a corpus of 1,730,000 words on average for each language.³ One can readily infer that the distribution of data over genres is not equal in the written corpus; however, their proportion is more or less respected for both languages, so that genres can be compared individually for French and Dutch.

For the spoken comparable data, we drew from the *CGN (Corpus Gesproken Nederlands)*, for Dutch, and from the *Valibel (Variétés Linguistiques du français de Belgique)* database, for French.⁴ With the focus set on spontaneous face-to-face conversation, only part of the available data in these well-established corpora were put to use: the *CGN* holds 8,940,098 words, of which 1,431,545 were compiled into a subcorpus of spontaneous conversations; the *Valibel* database provided only 50,668 words (or 3,373 turns) of this genre, in spite of its overall size (3,483,131 words). Given the discrepancy between the two subcorpora with regard to size, the complete *Valibel* corpus was taken into account to back-up some observations made on the basis of the smaller subcorpus.

The translation corpus, finally, was self-compiled and set up to contain translations of newspaper articles and literary novels, with an equal distribution over

genres (257,985 words for literary novels; 289,588 words for newspaper articles), but with an unequal distribution over languages (187,763 words for French as source language; 359,810 words for Dutch as source language).⁵

Each corpus thus has strengths and weaknesses, which makes a combined corpus approach even more desirable: perceived tendencies for one type of material can be invalidated or further enhanced by other material which does not present the flaws inherent to a different data source.

In all, the following sets of data were taken into account (see Table 1).

Table 1. Overview of corpus data used

		words
translation corpus	Source Dutch	359,810
	Source French	187,763
comparable corpus	Dutch	1,634,082
	French	1,833,481
spoken corpus	CGN	1,431,545
	Valibel	50,668 (3,483,131)

4. Empirical analysis

The study to be presented in the remainder of this paper will be both qualitative and quantitative in nature, and corpus-based rather than corpus-driven. The tripartite corpus analysis was formalized into features apt for processing in SPSS, allowing us to make empirically validated claims about, successively, the most “adequate” equivalent of French *en fait* in Dutch (Section 4.1) and their semantic profiles (Section 4.2).

4.1 *En fait* and its Dutch equivalents

An initial check of *en fait* in translation corpora with French as both source and target language yielded 65 occurrences which were translated by *eigenlijk* (“actually”) in 29.2% of all cases, and by *in feite* (“in fact”) in 30.8% of the cases (see Table 2).

Table 2. Initial mirror analysis for *en fait*

	Frequency	Percent
<i>eigenlijk</i>	19	29.2
<i>in feite</i>	20	30.8
<i>feitelijk</i>	3	4.6
<i>in werkelijkheid</i>	5	7.7
<i>in wezen, in principe</i>	2	3.1
residuals	9	13.8
no translation	7	10.8
Total	65	100.0

Translations of minor importance include *feitelijk* (“factually”), *in werkelijkheid* (“in reality”) or *in wezen* (“in essence”). In nearly 14% of the cases, there was a lexical reformulation of the meaning of *en fait*, and for close to 10.8% of the cases, *en fait* was not translated in the target text.

This translation distribution leads us to believe that both *eigenlijk* and *in feite* are most appropriate as translation equivalents of *en fait*. In order to determine which of both is the most salient equivalent, we carried out a cross-check analysis. The introduction of *eigenlijk* and *in feite*, respectively, in the translation data with Dutch as source language provided the following picture:

- *Eigenlijk* has a very high level of non-translations (47.7%), while translations with *en fait* cover only 13.6% of the cases, which nevertheless appears to be the preferred translation among the diverse alternatives (see Table 3).

Table 3. Cross-check analysis of *eigenlijk* (Dutch as source language)

	Frequency	Percent
<i>en fait</i>	12	13.6
<i>au fond, au juste, au fait</i> (“actually”)	7	8.0
<i>enfin, en fin de compte, finalement, à la fin, après tout</i> (“finally”)	8	9.1
<i>en réalité, dans la réalité, vraiment, réellement, à vrai dire</i> (“in reality”)	8	9.1
<i>plutôt, même, peut-être</i> (“rather”)	8	9.1
residuals	3	3.4
no translation	42	47.7
Total	88	100.0

- *In feite*, on the other hand, is translated in nearly half of the cases by *en fait*; however, with only 7 occurrences in the translation data, its low frequency hinders further investigation (see Table 4).

Table 4. Cross-check analysis of *in feite* (Dutch as source language)

	Frequency	Percent
<i>en fait</i>	3	42.9
<i>effectif</i> (“effectively”)	1	14.3
<i>en réalité</i> (“in reality”)	1	14.3
no translation	2	28.6
Total	7	100.0

The infrequency of *in feite* in Dutch is confirmed when one takes the overall frequencies of the three markers in speech (spoken corpus) and writing (comparable corpus) into account (see Table 5):⁶

Table 5. Frequency of *en fait*, *eigenlijk*, and *in feite* in the written and spoken comparable data

	writing		speech	
	abs. frequency	freq. / 10,000 w.	abs. frequency	freq. / 10,000 w.
<i>en fait</i>	220	1.20	43	8.49
<i>in feite</i>	97	0.59	11	0.08
<i>eigenlijk</i>	502	3.07	3712	25.93

The frequencies in Table 5 show that *in feite* is rather infrequent in writing (five times less frequent than *eigenlijk*, two times less frequent than *en fait*) and most strikingly infrequent in speech (324 times less frequent than *eigenlijk* and 106 times less frequent than *en fait*). This low frequency prompted us to discard *in feite* as the most salient equivalent of *en fait* and to consider *eigenlijk* instead. Evidently, *eigenlijk* cannot be considered as a strict equivalent of *en fait* (should such a strict equivalence exist in the first place): their frequencies diverge (*eigenlijk* being more frequent than *en fait*, especially in speech) and the translation relation is asymmetric (*en fait* being translated more often by *eigenlijk* than vice versa). Nevertheless, in this study, we will consider *eigenlijk* as a more likely, and certainly more interesting, equivalent than *in feite*.

4.2 Semantic profiles of *eigenlijk* and *en fait*

Our purpose is to establish the main semantic values for *en fait* and *eigenlijk*, by using a combined corpus approach. Initial clues are provided by the translation

data, for which we carried out a mirror analysis, indicating the main semantic fields in which to locate the values of *en fait* and *eigenlijk* (Section 4.2.1). In an in-depth semantic analysis of written comparable data, we focused on 50 occurrences of each marker, which are expected to confirm as well as extend the results found for the mirror analysis (Section 4.2.2). Finally, we checked the tendencies observed first against the results of the mirror analysis and secondly against frequent collocations in spoken comparable corpora (Section 4.2.3).

4.2.1 Clues from the mirror analysis

A mirror analysis consists of back-and-forth translations of a given item from the source language to the target language, and from the target language back to the source language. For *en fait* and *eigenlijk*, this procedure provided us with the model displayed in Figure 1:

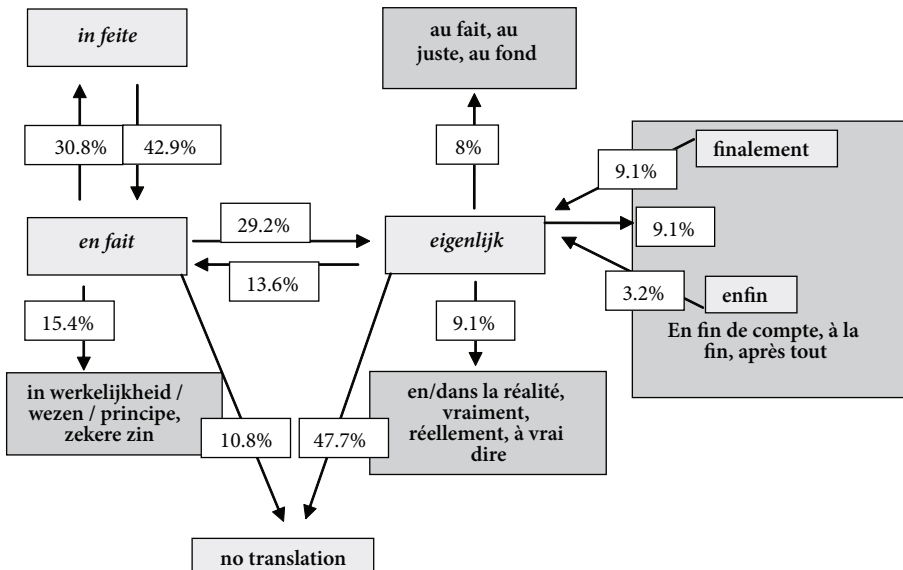


Figure 1. Translation network for *en fait* and *eigenlijk*

The striped fields represent the basic set of correlations conceivable for *en fait* and *eigenlijk*, i.e. which have a two-way relation with at least one marker and/or which represent a substantial part of the data. The gray fields express a one-way relation with *en fait* or *eigenlijk* and usually represent a more modest subset of the data.

What can we derive from the mirror analysis? Firstly, it seems that *eigenlijk* has a higher level of polysemy, with a wider range of possible translations with which it interacts in a generally weaker way. Secondly, *eigenlijk* also stands out by its elusive nature: with an extremely high number of non-translations (47.7%), it expresses meanings which are not always picked up on by the translator, or more

generally, which are pragmatic in nature and are not invariably inferred by the reader (and by extension by the hearer).

From the combined set of equivalents we can deduce a set of diversified meanings, including:

- **Factuality:** suggested by *en fait* (“in fact”), *au fait* (“for that matter”), *in feite* (“in fact”)
- **Truth-conditionality:** appearance ↔ reality, suggested by *in werkelijkheid* (“in reality”), *en/dans la vérité* (“in truth”), *vraiment* (“truthfully”), *réellement* (“really”), *à vrai dire* (“to tell (you) the truth”)⁷
- **Precise or deeper meaning:** *au juste* (“precisely”), *au fond* (“in depth”), *in wezen* (“fundamentally”), *in zekere zin* (“in a certain way”)
- **Temporality or conclusion:** *finalement* (“finally”), *enfin* (“after all”), *en fin de compte* (“after all”), *à la fin* (“in the end”), *après tout* (“after all”)

We grouped these meanings together in larger semantic categories, comprising (i) *factuality*, (ii) *opposition* (~ *truth-conditionality*) and (iii) *reformulation* (~ *search for precision, in-depth meaning or explanation, conclusion*) to be used in the comparable analysis.⁸

4.2.2 Comparable data analysis

“Factuality”, “opposition” and “reformulation” will function as marks for the outline of the analysis of written comparable data to be presented here. It is our aim to determine the general semantic structure of *en fait* and *eigenlijk*, by investigating the nature of the propositions p and q, p usually preceding the proposition q containing the marker, as well as their interrelation, discursive effects and formal components.

We set out to look for all occurrences of our two markers, viz. *en fait* (220 cases) and *eigenlijk* (502 cases). Our comparable data analysis was realized for 50 random occurrences per marker as they were found in two monolingual corpora comprising literary, essayistic and journalistic texts. Initial questions include:

- i. Does *en fait* include “factuality” in its semantics, as suggested by its etymological relation with “facts”? If so, is this feature present in *eigenlijk* as well, given the semantic equivalence with *en fait*?
- ii. How (strong) is “opposition” expressed in *en fait* and *eigenlijk*? Does “counter-expectation” play a part?
- iii. How (strong) is “reformulation” expressed in *en fait* and *eigenlijk*?

4.2.2.1 Factuality. On pure etymological grounds, one could consider “factuality” to be the core meaning of *en fait* “in fact” and, given the translation equivalence with *en fait*, of that of *eigenlijk* “actually” as well. Whether this hypothesis has any

validity can be established by looking into the nature of the proposition in which we found the marker. Such an analysis rests on the following hypothesis: if the marker commonly occurs in a proposition “(en fait) q” expressing a fact, then the marker is likely to be more factual in nature. If, however, it is more frequently found in a proposition expressing something other than a fact (viz. an opinion or an action, cf. Pander Maat & Degand 2001), then the marker is easier to associate with a non-factual, subjective meaning.

Specifically, we examined the proportion of facts, actions and opinions (i) over p, (ii) over q, (iii) comparing p with q, and (iv) comparing *eigenlijk* with *en fait* (see Figure 2):

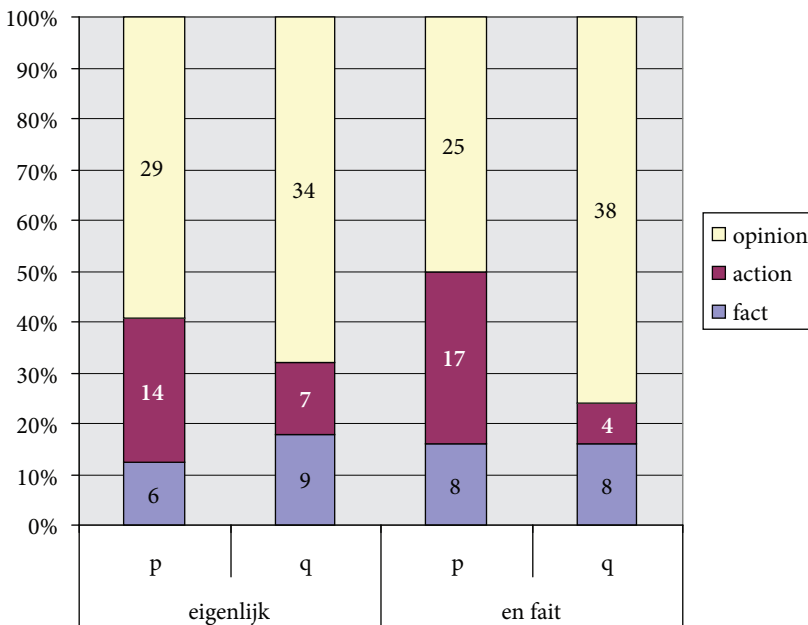


Figure 2. Nature of p and q with *eigenlijk* and *en fait*

Figure 2 suggests a clear preference for opinions, both for *en fait* and *eigenlijk*, and in p as well as in q. This is confirmed statistically: the distribution of facts, opinions, and actions over p does not diverge between *eigenlijk* and *en fait* ($X^2(2) = 0.86$; $p = 0.65$ (NS)); nor for the nature of q ($X^2(2) = 1.1$; $p = 0.577$ (NS)). A closer view at the two markers separately indicates a similar distribution over p and q for *eigenlijk* (with a slight rise of opinions between p and q from 29.3% to 34.3%, and a drop in actions from 14.1% in p to 7.1% in q, but this remains non significant: $X^2(2) = 3.32$; $p = 0.19$). For *en fait*, opinions rise from 25% in p to 38% in q, probably caused by the dramatic fall of actions, representing 17% of the p propositions and only 4% of q propositions ($X^2(2) = 10.73$; $p < .01$).

These tendencies suggest that neither *eigenlijk* nor *en fait* introduce factual propositions; instead, they introduce an opinion relegating a previously stated fact, opinion or action to the domain of appearances. However, the association with “factuality” is not coincidental: *en fait*, and to a lesser extent *eigenlijk*, do yield a factual interpretation, but then *as perceived by the speaker*. In other terms, while proposition *q* in itself may be an opinion, it is in fact the speaker who, by using a marker, qualifies “his” *q* proposition as more real, more true or more correct than what is stated in *p*. Our two-fold hypothesis, according to which the nature of proposition *q* determines the nature of the marker which it contains, is thus also put into perspective: what is perceived by an objective observant as an opinion is actually originally intended by the speaker to express something which is more real than the opinion, fact or action previously expressed. *En fait*, and *eigenlijk*, may be the pivotal agents in this respect, since it often seems to be their function to express that a seemingly subjective statement (“*q*”) is the objective truth, at the very least in the speaker’s mind. It can be readily concluded that *eigenlijk*, and especially *en fait*, are strongly argumentative and integrate “factuality” as a complex and contradictory value in their semantics.

4.2.2.2 Opposition. Concepts such as ‘more real’, ‘more true’ and ‘more correct’ naturally evoke the image of a dichotomy between what is perceived as pertaining to the domain of appearances (something stated in *p*) and what is perceived as being the actual truth or the more objective state of affairs (something stated in *q*). Appearance ↔ reality is an instantiation of “opposition” as a semantic field, although one cannot deny that oppositions of this sort, viz. paraphrasable in terms of ‘more true/correct/real’, are not absolute, but rather more or less strong deviations from an original proposal, and thus relate to adversativity as defined in 2.1.

Using the written comparable data, we looked into the kind of relation between propositions *p* and *q* as it is expressed on an *ideational* level, i.e. without taking into account the pragmatic influence of *eigenlijk* or *en fait* on their relation. To this effect, we delimited propositions *p* and *q* and established how they would relate to each other if *en fait* and *eigenlijk* were absent. Such an analysis is useful to determine not only what types of relations our markers occur most frequently in, but also, in a later stage, what their individual effect is on the original, ideational relation between *p* and *q*. The analysis focused on the semantic relation expressed by *q* with respect to *p* and provided us with four case scenarios. Either *q* expressed an opposition to or a reformulation of *p*, or *q* had a pragmatic function (i.e. discourse marking or unclear ideational association with *p*) (see Table 6).⁹

Table 6. Meanings expressed between p and q on an ideational level for *eigenlijk* and *en fait*

	<i>eigenlijk</i>	<i>en fait</i>
opposition to p	17 (34.7%)	28 (57.1%)
reformulation of p	24 (49%)	15 (30.6%)
discourse marking	4 (8.2%)	absent
no clear association	4 (8.2%)	6 (12.2%)
Total	49	49

As one can observe, “opposition” accounts for 57.1% of all relations in which *en fait* occurs, and for 34.7% of those featuring *eigenlijk*. The manner in which this oppositional meaning between p and q is expressed or suggested is not uniform, however. It is precisely the underspecification of strict opposition in both markers which accounts for the differences in interpretation of the oppositional meaning. This becomes particularly clear when one analyses the effect of *eigenlijk* and *en fait* on the relation between p and q. In some cases, *eigenlijk* and *en fait* are the only markers yielding a linguistically expressed oppositional interpretation to an otherwise ideational opposition between p and q; in (7), for example, *en fait* makes explicit the opposition between what is commonly associated with “democracy” and what is expressed as the actual truth in this case:

- (7) La fameuse démocratie, rempart supposé de toutes les injustices, reposait **en fait** sur les rapports de force des différents lobbies.
 « The celebrated democracy, alleged stronghold against all injustice, was *en fait* based on balances of power between the different lobbies.”

In other cases, another linguistic marker is used to express “opposition” (the prototypical one having a meaning of “but”), in which case *en fait* and *eigenlijk* merely have an enhancing function:

- (8) Op mijn zwerftochten door de bossen, altijd vergezeld van een rossig hondje, ontmoet ik een jongen die zich Jan noemt **maar eigenlijk** David heet.
 “In my wanderings through the woods, always accompanied by a ruddy little dog, I met a boy who called himself Jan, **but** was *eigenlijk* called David.”

A third recurrent, somewhat less obvious, type of contrast often found for *eigenlijk* and *en fait* is that of “opposition to a normative viewpoint”, which recalls the notion of “counterexpectation” and “adversativity” presented in Section 2. In such cases, the use of *en fait* or *eigenlijk* is legitimated against the background assumption that some hearers might not agree with proposition q, based on existing norms, beliefs and expectations in society. This inference is difficult to identify unambiguously, and often seems to supplement other types of opposition, thus operating as a secondary oppositional value:

- (9) Le président Clinton, interpellé sur l'affaire, a tenté de calmer les passions. Je pense, **en fait**, que les Chinois vont faire l'impossible pour que tout se déroule impeccablement, affirme Bill Clinton.
 « President Clinton, when asked about the issue, has attempted to calm people's feelings. I think, *en fait*, that the Chinese are going to do the impossible to make sure that everything goes perfect, affirms Bill Clinton.»

The frequent co-occurrence of “counterexpectation”, as well as the merely enhancing effect of the marker on an otherwise opposite relation between p and q are indications that opposition is indeed not to be interpreted in absolute terms for the two markers, and thus best thought of in terms of ‘adversativity’.

It should be noted that, counterintuitive as it may seem, adversative *en fait* and *eigenlijk* do not merely occur in adversative relations: in some cases, p and q are not connected through an adversative relation on an ideational level, but *en fait* or *eigenlijk* yield an adversative note to the interpretation nonetheless. In (10), q functions as an explanation for part of p, viz. une composante étrangère “a foreign component”; *en fait*, however, seems to express conflict with previous expectations on the part of the researchers (i.e. it is not a foreign component, but rather light reflected by a planet):

- (10) C'est en analysant le spectre lumineux de l'étoile Tau Bootis, annonce le magazine “Nature”, que les chercheurs auraient découvert une “composante étrangère”, qui proviendrait **en fait** de la lumière réfléchiée par la planète en question.
 « It is during the analysis of the luminous spectrum of the star Tau Bootis, announces the journal « Nature », that researchers would have discovered a « foreign component », which would *en fait* come from the light reflected by the planet in question.»

In terms of frequency, “enhancement of an explicit opposition” seems to be the predominant adversative value in both *en fait* (40% of all adversative relations) and *eigenlijk* (35.3%), followed at a distance by “explicitation of an implicit opposition” (20% for *en fait*; 23.5% for *eigenlijk*). The importance of supplementation with “opposition to a normative viewpoint” is rather modest, although it often seems to be possible as a secondary interpretation. Non-oppositional effects in a primary oppositional relation account for 20% of all cases of *en fait* (these effects being “attenuation” and “enhancement” of a part of p or q); for *eigenlijk*, these uses explain 29.4% of the cases, with “causality” in addition to “enhancement” and “attenuation”. These details are summarized in Table 7.

Table 7. Proportions of oppositive relations for *en fait* and *eigenlijk*

<i>EN FAIT</i>	<i>EIGENLIJK</i>
opposition (57.1%)	opposition (34.7%)
a. enhancement of an explicit opposition to p (40%)	a. enhancement of an explicit opposition to p (35.3%)
b. explicitation of an implicit opposition to p (20%)	b. explicitation of an implicit opposition to p (23.5%)
c. causality (13.3%)	c. attenuation of p/q/part of p or q (17.6%)
d. attenuation of p/q/part of p or q (13.3%)	d. enhancement of p/q/part of p or q (5.9%)
e. enhancement of p/q/part of p or q (6.7%)	e. causality (5.9%)
f. supplementation with an opposition to a normative viewpoint (6.7%)	f. supplementation with an opposition to a normative viewpoint (5.9%)

4.2.2.3 Reformulation. As stated before, the idea of “more true”, “more real” or “more correct” is often expressed very subtly, not so much as opposition, but rather as a (slight) deviation from an original proposal. The function of such “deviating” propositions is to add information deemed necessary to understand a proposition p, at the same time introducing a more or less important shift with respect to that proposition p. We opted for the concept of ‘reformulation’ to denote such cases, since it emphasizes the fact that something is formulated again, but in a somewhat different manner.

Within the realm of “reformulation”, different case scenarios were found. The most important one is that of “precision”, i.e. a more precise, detailed or correct formulation of a previous statement. Scope is important here, since the precision can apply to different chunks of texts, varying in size and salience, among other things the entire proposition p, part of it or part of the proposition q. A frequently found instance of such a reformulative function is the precision of a concept which the speaker deems not accurate enough for his proposal to be fully understood. In (11), for example, *lichaam* “body” is corrected by a more precise formulation, viz. *lichaamszwaartepunt* “bodily center of gravity”:

- (11) Wanneer de last niet dichterbij het lichaam kan gebracht worden, moet je trachten het lichaam, *eigenlijk* het lichaamszwaartepunt, dichterbij de last te brengen.

“When the burden cannot be brought any closer to the body, one should try to bring the body, *eigenlijk* the bodily center of gravity, closer to the burden.”

A second type of “reformulation” is a natural extension of the idea of a “more in-depth” formulation of a previous statement, pushing the interpretation beyond the limits of pure addition or deviation and into the field of causality; in providing more context for or information on p, the speaker easily introduces cause and ef-

fect, i.e. an explanation of p or of why p is legitimate. In (6), reproduced here as (12), the *en fait* proposition has explanatory potential for what is stated in p:

- (12) Ces deux formes de récréation sont relativement neuves en Belgique et ont **en fait** été propagées par les Pays-Bas.
 « These two forms of recreation are relatively new in Belgium and have *en fait* been propagated by the Netherlands. »

A third type is only weakly associated with « reformulation » and involves an emphatic focus, either as enhancement or as attenuation of p, part of p or part of q. In such cases, *en fait* and *eigenlijk* emphasize something which is deemed important or which one wants to attune to the reality as perceived. “Enhancement” and “attenuation” relate to “reformulation” in that they implicitly shift the focus on the information from less to more important, or from more to less important. In (13), the merits of women are extended to a higher level in q as compared to p:

- (13) Vrouwen hebben “aan de wieg van onze beschaving” gestaan. Vrouwen hebben *eigenlijk* alles van betekenis uitgevonden.
 “Women have always been at the birthplace of our civilization. Women have *eigenlijk* invented everything of consequence.”

In (14), the shape in which the force manifests itself is expressed more carefully or more thoughtfully in q (“dissipated”) than in p (“destroyed”):

- (14) Lorsque deux corps mous se heurtent et s’immobilisent, leur force est-elle détruite, ou bien s’agit-il seulement d’une apparence, la force s’étant *en fait* dissipée dans les parties menues des corps ?
 « When two soft bodies collide with each other and come to a hold, is their strength destroyed, or could this merely be an impression, and could their strength *en fait* have disappeared in the miniscule parts of the body?»

A last case scenario has an even weaker association with linguistically expressed reformulation, since it resides entirely in the discursive mind of the speaker. The introduction of a ‘metaperspective’, as we would call it, involves the linguistic expression of a disconnected thought, i.e. a proposition q which has no strong ideational association with p, a mental “leap”, or of a meta-comment, i.e. a comment made by the speaker on a thought which is not or only implicitly expressed in the preceding context. They are related to “reformulation” insofar as the speaker introduces a correction, precision or evaluation, but only weakly so, as the object to which they apply is not (explicitly) expressed in discourse. An example is (15), in which one can visualize the mental processing going on in Liana, leaping from “feeling good” to “claiming that Jíí is a decent boy”, and enhanced by the time adverbial *eindelijk* “in the end”:

- (15) “Voelen jullie je ook zo zalog?” vraagt Liana en ze drukt precies uit wat wij allemaal voelen. “Die Jii is *eigenlijk* best een geschikte jongen,” zegt ze eindelijk.
 “‘Do all of you feel as deliciously as I do?’ asks Liana and she expresses exactly what we all feel. ‘That Jii boy is *eigenlijk* a rather good lad,’ she says in the end.”

Again, as was the case for “opposition”, reformulative *en fait* and *eigenlijk* can also be used in non-reformulative, i.e. oppositive, contexts. In those cases, p and q have an ideational relation which is non-reformulative in nature, but *en fait* and *eigenlijk* attribute a reformulative interpretation to their relation. In (16), for example, the relation between p and q is clearly oppositive (if only by the presence of *maar* “but”); however, *eigenlijk* modifies *veel* “much” in p by attenuating its strength, thus introducing a more precise or correct reformulation of (part of) p:

- (16) De tijd mag verstrijken, behalve wij. Je bent wel veel veranderd, maar **eigenlijk** ook niet, zegt ze.
 “Time may go by, except for us. You changed a lot, but *eigenlijk* not really, she says.”

As one can infer from Table 8, frequencies indicate that “precision” is the predominant type of reformulation, both for *eigenlijk* (29.2% of all reformulative relations) and *en fait* (50%). “Causality” is the first runner up, with equal distributions over *eigenlijk* and *en fait* (both 25% of all reformulations). “Enhancement” is an important value for *eigenlijk* (20.8%), but not so for *en fait* (7.1%); “attenuation”, on the other hand, is virtually absent for *eigenlijk* and does not exceed 3.6% for *en fait*. Reformulative uses account for 33.3% of all oppositive relations in *en fait*, for 23.5% in *eigenlijk*.

Table 8. Proportions of reformulative relations for *en fait* and *eigenlijk*

<i>EN FAIT</i>	<i>EIGENLIJK</i>
reformulation (30.6%)	reformulation (49%)
a. precision of p/q/part of p or q (50%)	a. precision of p/q/part of p or q (29.2%)
b. causality (25%)	b. causality (25%)
c. explicitation of an implicit opposition to p (7.1%)	c. enhancement of p/q/part of p or q (20.8%)
d. supplementation with an opposition to a normative viewpoint (7.1%)	d. supplementation with an opposition to a normative viewpoint (12.5%)
e. enhancement of p/q/part of p or q (7.1%)	e. explicitation of an implicit opposition to p (8.3%)
f. attenuation of p/q/part of p or q (3.6%)	f. metaperspective (4.2%)

4.2.3 *Converging evidence from the mirror analysis*

The analysis of the translation equivalents for *eigenlijk* and *en fait* suggested four meanings for the two markers, viz. “factuality”, “truth-conditionality”, “precision” and “causality”. All of these meanings were accounted for in the comparable corpus analysis, which conversely provided us with a set of meanings not originally identified in the translation analysis, viz. “counterexpectation”, “enhancement/attenuation” and “metaperspective”.

The most salient one is that of supplementation with an “opposition to a normative viewpoint”, or “counterexpectation”, referring to those cases in which *eigenlijk* or *en fait* is used to counter an expectation that the hearer, or even the speaker, might have regarding the interpretation of p, or of its relation with q. This pragmatic value can be found both in oppositive and reformulative relations and adds up to 18.4% for *eigenlijk* and 13.8% for *en fait*.

Related to this value is another meaning which pairs up with the introduction of a “metaperspective”, i.e. a disconnected thought from p or a meta-comment on discourse. It was found almost exclusively with *eigenlijk* (4.2%), suggesting that *eigenlijk* is more prone to expressing elusive, discourse-oriented values than *en fait*, for which no meta-function could be identified.

Two more values not found in the mirror analysis relate to “enhancement” or “attenuation” of p, of q or of part of p or q. The expressive enhancement or attenuation of a contextual element was found especially for *eigenlijk*: “attenuation” was most prominent in oppositive relations, in which *eigenlijk* was thus used to soften the opposition or strength of the arguments; “enhancement” could be found with a relatively high frequency in reformulative relations (20.8% for *eigenlijk*, against 7.1% for *en fait*).

Generally speaking, the set of meanings found for both markers suggests that *eigenlijk* is indeed, as hypothesized in the mirror analysis, more strongly polysemous than *en fait*. It occurs predominantly in reformulative relations and can express a variety of meanings which are either absent in *en fait* (viz. “metaperspective”) or (very) infrequent (viz. “attenuation”, “enhancement” and “counterexpectation”).

4.2.4 *Converging evidence from collocations*

While the main objective of the mirror analysis was to give clues on the major semantic fields in which to situate the semantics of *en fait* and *eigenlijk*, an analysis of collocations may give us more information on the distribution and strength of these meanings. We hypothesize here that two meanings in particular could benefit from such an approach, viz. “enhancement of an explicit opposition” and “causality”. “Enhancement of an explicit opposition”, which was important in both *en fait* (40%) and *eigenlijk* (35.3%), was found exclusively for cases in which *en fait*

or *eigenlijk* collocates with another discourse marker already yielding an oppositional meaning (prototypically, “but”), and which *en fait* or *eigenlijk* comes to enhance. “Causality”, which was defined as indirectly related to “reformulation”, was accounted for in *en fait* and *eigenlijk* with equal distributions (each 25% of all reformulative relations).

Using PhraseContext, we outlined the most important collocations found with *en fait* and *eigenlijk* in the spoken data.¹⁰ We checked for collocations in a span of five positions to the left and to the right of the marker (see Tables 9 and 10).¹¹

Table 9. Frequent collocations with *en fait* in spoken French (*Valibel*, spontaneous face-to-face conversations)

<i>EN FAIT</i>	1st left	TOTAL left	1st right	TOTAL right	TOTAL overall
long or short pause	50	144	39	105	249
hesitation marker	6	33	38	64	97
<i>et</i> (“and”)	12	29	4	19	48
<i>mais</i> (“but”)	15	27	–	7	34
<i>donc</i> (“so”)	11	18	4	12	30
<i>parce que</i> (“because”)	2	8	2	11	19

Table 10. Frequent collocations with *eigenlijk* in spoken Dutch (*CGN*, spontaneous face-to-face conversations)

EIGENLIJK	1st left	TOTAL left	1st right	TOTAL right	TOTAL overall
<i>ja</i> (“yes”)	83	506	31	401	907
hesitation marker	37	282	81	441	723
<i>en</i> (“and”)	35	369	16	134	503
<i>maar</i> (“but”)	62	312	45	221	533
<i>dus</i> (“so”)	163	335	9	43	378
<i>want</i> (“because”)	21	111	15	114	225
<i>toch</i> (“nevertheless”)	39	79	34	71	150
<i>willen</i> (“want”)	50	106	8	23	129
<i>kunnen</i> (“can”)	15	49	15	78	127
<i>zouden</i> (“should”)	27	64	33	48	112
<i>toen</i> (“then”)	15	64	–	7	71
<i>mogen</i> (“may”)	6	6	–	9	15

These data suggest first of all that hesitation markers, such as *euh* or *uh(m)*, as well as short and long pauses, are found most frequently in the vicinity of both markers. These usually indicate an attempt on behalf of the speaker to find the right words to express his thoughts; the presence of *eigenlijk* and *en fait* amidst these linguistic means could suggest that they, too, have a “stalling” function, i.e. try to slow down the pace of conversation to allow the speaker to express himself in a coherent way.

With the exception of the connective *en* “and” in Dutch, it seems that the next frequent markers with which *en fait* and *eigenlijk* are found express meanings of “opposition” (*mais* “but” for French, *maar* “but” for Dutch). The collocation *mais en fait* is most frequently found, with 15 occurrences or 0.04/10,000 words, and, taking a wider context of 5 positions to the left and 5 to the right into account, (*en fait...*) *mais (...en fait)* occurs 30 times (or 0.1/10,000 words). Collocations of *maar* with *eigenlijk* are much more frequent, with 162 occurrences or 0.43/10,000 words; taking into account the wider context, *maar* is even more frequent (3.72/10,000 words). This seems to suggest that *eigenlijk* “needs” *maar* to express the relation between p and q to an extent that *en fait* does not and that its intrinsic oppositive force thus is considerably weaker.

A similar observation can be made for a third frequent type of collocation, i.e. that featuring a causal marker meaning “so”, expressed in French by *donc* and by *dus* in Dutch. Our analysis indicates that *donc en fait* is much less frequent, with 11 occurrences or 0.03/10,000 words, than *dus eigenlijk* (163 occurrences or 1.14/10,000 words). Again, this seems to suggest that *eigenlijk* in itself does not express overt meanings of “causality”, and benefits from the presence of a causal marker such as *dus* “so” to do so.

5. Discussion

The fact that *en fait*, and especially *eigenlijk*, frequently constitute collocations with other discourse markers of “causality” (*dus/donc* “so”) and “opposition” (*maar/mais* “but”) raises questions as to the strength of the meanings identified as part of their semantics. Other types of analysis point in the same direction. The analysis of written comparable data, for example, showed that in many cases, the two markers have a merely expressive function, enhancing or attenuating (a part) of p or q, or an existing opposition. The translation analysis moreover showed that especially *eigenlijk* has a feeble meaning of its own, since it is not translated in nearly half of all its occurrences (47.7% to be precise). All these observations are in line with the hypothesis that *eigenlijk* and *en fait* express non-absolute meanings, and that in particular their meaning of “opposition” should be conceptualized rather as

“deviation” than as strict contrast. It seems that it is precisely this “vague” interpretation which may account for the high level of polysemy characteristic, again especially, of *eigenlijk*, allowing this marker to occur in a multitude of contexts. What’s more, its underspecification seems to make it an excellent “device” for speakers to manipulate in discourse, as it easily assumes meanings of, for example, “counterexpectation” as well.

A further hypothesis could be that it is the meaning of deviation which underlies both of the conceptually quite different meanings of “opposition” and “reformulation”. “Deviation” is in fact the common denominator for the two meanings as identified in our analyses: “opposition” often takes on the form of an “écart” (Rossari 1992: 159, cf. 2.1) between appearance and reality; “reformulation” seems to involve in a number of cases the “more precise”, “more detailed” or “more correct” formulation of an original proposal. “Causality”, then, is a natural outcome for those cases of reformulation in which more information is provided for (part of) p, viz. in the shape of an explanation or a pointing out of consequences. The semantics of *en fait* and *eigenlijk* could thus be presented as follows:

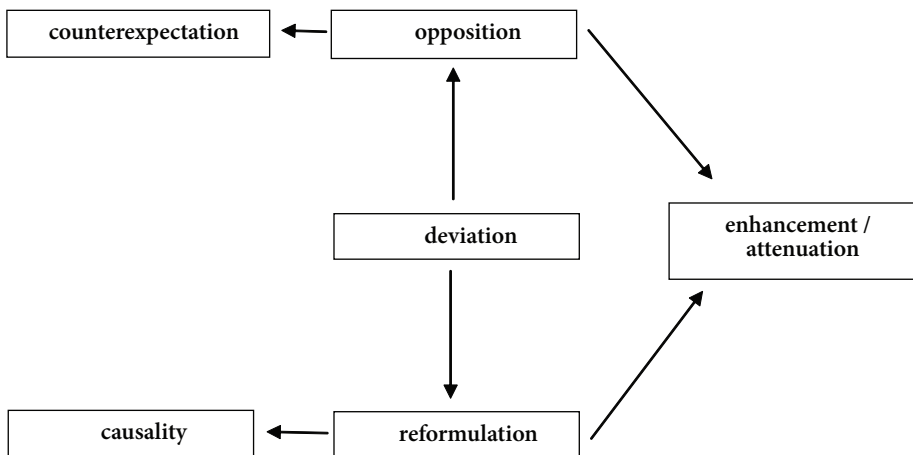


Figure 3. Relational semantic field for *eigenlijk* and *en fait* (hypothesis)

In Figure 3 we find further hypothetical correlations between “opposition” and “counterexpectation”, based on the idea that opposition between p and q may go hand in hand with (countering of) speaker’s expectations concerning the natural ordering of relations in the real world. More “expressive” meanings such as “enhancement” and “attenuation” were found for both “reformulation” and “opposition”, which could be the result of the semantic underspecification of *eigenlijk* and *en fait*, as well as of their potential to operate on a discourse marking level.

Further research should indicate whether this model has any relevance on a diachronic level of analysis, since the study of these markers in different time

periods may shed a different light on the origins of their semantic values as identified in present-day speech and writing.

Another topic worthy of further investigation concerns the particular usage constraints (or rather lack thereof) on *eigenlijk* and *en fait* in the spoken register, with special reference to spontaneous face-to-face conversations (ongoing work). Overall frequencies (cf. Table 5) show a major discrepancy with regard to the occurrence of *eigenlijk* and *en fait* in the spoken corpora (8.49 / 10,000 words in *Valibel*; 25.93 / 10,000 words in *CGN*), on the one hand, and their frequency in the written data (1.20 / 10,000 words in *Valibel*; 3.07 / 10,000 words in *CGN*), on the other. Merely on the basis of these figures, one can hypothesize that *eigenlijk* as well as *en fait* have a different status in spontaneous conversation, which may very well illustrate an ongoing loosening of constraints in informal speech. Evidence supporting such a hypothesis comes, among other things, from the frequent collocations with hesitation markers as found in spontaneous conversation, which seem to transfer to *eigenlijk* and *en fait* a sort of stalling function. Further research should establish the specific functions of these markers in spontaneous conversation and pinpoint potential divergences with respect to the written register.

In this paper, we hope to have shown the benefits of a combined corpus approach, involving translation data as well as written and spoken comparable corpora. Using mirror analyses on translation data, one is able to do a first exploration of the semantic and formal field of linguistic forms, in our case discourse markers. Initial inferences made on the basis of translation equivalents, specifically in terms of ‘semantic domains’ relevant in their description, can provide a rather accurate idea of general tendencies. Our analyses did show, however, that a more “traditional” type of corpus analysis is required to gain full insights into their semantics. The study of collocations, finally, can be considered a valuable asset to determine the frequency and actual vitality of certain observations and tendencies. In all, we hope to have contributed to the study of discourse markers from a contrastive point of view, and to have shown that a variety of corpus approaches offers the best chances of gaining insight into the semantics of these markers.

Notes

* The second author is research associate with the Belgian Science Foundation (FRS-FNRS). This research was supported by a Interuniversity Attraction Pole Project on “Grammaticalization and (Inter)Subjectification” (P6/44) funded by the Belgian Federal Government.

1. It should be clear from the outset that we do not make a clear distinction between semantic and pragmatic meanings for the markers to be discussed in this paper. While some are more

clearly pragmatic in nature, we believe that pragmatic as well as semantic meanings are located in a continuum, with no obvious distinction between them.

2. The electronic version of the *Trésor de la langue française* can be consulted on <http://atilf.atilf.fr/> (accessed May 2009). The Dutch marker *eigenlijk*, which will be treated in this paper as the most important equivalent of French *en fait*, obviously has a different etymological history. It is most likely the result of a fusion between the adjective *eigen* “one’s own” and the suffix *-lijk* “like”. Its etymology remains undiscussed by most dictionaries of Middle Dutch, which can be explained by its rather late appearance in Dutch (an exploratory study suggests that *eigenlijk* does not appear before 1560 and becomes frequent only in the last quarter of the 17th century).
3. The written comparable corpus was assembled by the authors and contains extracts from 8 novels, extracts from 5 different newspapers and 4 essays for the Dutch part. The French material comprises 8 (incomplete) novels, extracts from 3 newspapers and 2 economical texts, and 5 essays.
4. The *Corpus Gesproken Nederlands* was created and financed by the Nederlandse Taalunie and comprises some 9 million words. It contains private and public dialogues and monologues, with genres ranging from face-to-face conversations, interviews and phone dialogues to news bulletins, lectures and debates (cf. http://lands.let.kun.nl/cgn/doc_Dutch/topics/design/design.htm#intro — accessed May 2009). The *Valibel* corpus was developed at the Université catholique de Louvain and currently has over 4 million words worth of data (Francard et al. 2002), consisting of transcriptions of interviews on sociolinguistic matters (42%), of informal conversations between friends (30%), and of a variety of specific communication situations (radio interviews, debates, work reunions, classes, reading tasks; 28%) (cf. <http://www.uclouvain.be/valibel> — accessed May 2009).
5. The translation corpus was assembled by Liesbeth Mortier and contains for Dutch 4 novels (Dutch translated into French), and extracts from 1 newspaper (*Industrie*). The French data comprises 4 novels (French translated into Dutch), and extracts from 2 newspapers (*Industrie*, *Le Monde-De Morgen*). All translations were made by professionals and it was assured that all extracts and novels were translated by as many different translators as possible.
6. The 50,668 words *Valibel* subcorpus provides 8.49 occ / 10,000 words; given the small size of this corpus compared to the data contained in the CGN, we also checked the overall *Valibel* corpus. This provided 1454 occurrences or 4.17 occurrences per 10,000 words, which makes *en fait* almost twice less frequent than in spontaneous face-to-face conversation. This could point to a significant tendency for *en fait* to occur primarily in truly spontaneous communication.
7. The symbol ↔ here has the meaning of “opposes”.
8. The symbol ~ here has the meaning of “relates to”.
9. For both *eigenlijk* and *en fait*, there was 1 occurrence which was not relevant for the discussion, hence 49 occurrences for each marker.
10. *PhraseContext* is a tool for text analysis, writing, collocation analysis, concordancing, text and XML output. Cf. <http://www.hjkm.dk/PhraseContext/> (accessed May 2009).
11. In the Dutch data, there were no specific indications pointing to long or short pauses.

References

- Aijmer, K. & Simon-Vandenbergen, A.-M. 2004. "A model and a methodology for the study of pragmatic markers: The semantic field of expectation". *Journal of Pragmatics*, 36 (10), 1781–1805.
- Aijmer, K., Foolen, A. & Simon-Vandenbergen, A.-M. 2006. "Pragmatic markers in translation: A methodological proposal". In K. Fischer (Ed.), *Approaches to Discourse Particles*. Amsterdam: Elsevier, 101–114.
- Brinton, L. 1996. *Pragmatic Markers in English. Grammaticalization and Discourse Functions*. Berlin/New York: Mouton de Gruyter.
- Danjou-Flaux, N. 1980. "A propos de *de fait, en fait, en effet et effectivement*". *Le Français Moderne*, 48 (2), 110–139.
- Danjou-Flaux, N. 1982. « *Réellement et en réalité: données lexicographiques et description sémanitique* ». *Lexique*, 1 (1), 105–150.
- DeLancey, S. 2001. « The mirative and evidentiality ». *Journal of Pragmatics*, 33 (2), 369–382.
- Dyvik, H. 1998. "A translational basis for semantics". In S. Johansson & S. Oksefjell (Eds.), *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*. Amsterdam: Rodopi, 51–86.
- Francard, M., Geron, G. & Wilmet, R. 2002. "La banque de données VALIBEL: Des ressources textuelles orales pour l'étude du français en Wallonie et à Bruxelles". In C. Pusch & W. Raible (Eds.), *Romanistische Korpuslinguistik — Korpora und Gesprochene Sprache / Romance Corpora Linguistics — Corpora and Spoken Language*, Vol. II. Tübingen: Gunter Narr, 71–80.
- Fraser, B. 2006. "On the universality of discourse markers". In K. Aijmer & A.-M. Simon-Vandenbergen (Eds.), *Pragmatic Markers in Contrast*. Amsterdam: Elsevier, 73–92.
- Granger, S., Lerot, J. & Petch-Tyson, S. (Eds.) 2003. *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam/Atlanta: Rodopi.
- Gutt, E.-A. 1998. "Pragmatic aspects of translation: Some relevance-theory observations". In L. Hickey (Ed.), *The Pragmatics of Translation*. Clevedon: Multilingual Matters, 41–53.
- Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J. & van der Toorn, M. (Eds.) 1997. *Algemene Nederlandse Spraakkunst*, 2 volumes. Groningen: Martinus Nijhoff.
- Hansen, M.-B. 2005. "A comparative study of the semantics and pragmatics of *enfin* and *finale-ment*, in synchrony and diachrony". *French Language Studies*, 15 (1), 153–171.
- Hyland, K. 1998. *Hedging in Scientific Research Articles*. Amsterdam/Philadelphia: John Benjamins.
- Jaszczolt, K. 2003. *Meaning Through Language Contrast*. Amsterdam/Philadelphia: John Benjamins.
- Johansson, S. & Oksefjell, S. (Eds.) 1998. *Corpora and Cross-Linguistic Research. Theory, Method and Case Studies*. Amsterdam/Atlanta: Rodopi.
- König, E. 1988. "Concessive connectives and concessive sentences: Cross linguistic regularities and pragmatic principles". In J. Hawkins (Ed.), *Explaining Language Universals*. Oxford: Basil Blackwell, 145–185.
- Lang, E. 2000. "Adversative connectors on distinct levels of discourse: A re-examination of Eve Sweetser's three-level approach". In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause-Condition-Concession-Contrast. Cognitive and Discourse Perspectives*. Berlin: Mouton de Gruyter, 235–256..

- Lewis, D. 2006. « Contrastive analysis of adversative relational markers, using comparable corpora ». In K. Aijmer & A.-M. Simon-Vandenberg (Eds.), *Pragmatic Markers in Contrast*. Amsterdam: Elsevier, 139–153.
- Lindquist, H. & Mair, C. (Eds.) 2003. *Corpus Approaches to Grammaticalization in English*. Amsterdam/Philadelphia: John Benjamins.
- Malchukov, A. 2004. “Towards a semantic typology of adversative and contrast marking”. *Journal of Semantics*, 21 (2), 177–198.
- Noël, D. 2003. “Translations as evidence for semantics: An illustration”. *Linguistics*, 41 (4), 757–785.
- Oh, S.-Y. 2000. “Actually and in fact in American English: A data-based analysis”. *English Language and Linguistics*, 4 (2), 243–268.
- Pander Maat, H. & Degand, L. 2001. “Scaling causal relations and connectives in terms of Speaker Involvement”. *Cognitive Linguistics*, 12 (3), 211–245.
- Rossari, C. 1992. “De fait, en fait, en réalité: Trois marqueurs aux emplois inclusifs ». *Verbum*, 14 (3), 139–161.
- Schourup, L. 1999. “Discourse Markers”. *Lingua*, 107 (3–4), 227–265.
- Schwenter, S. 2000. “Viewpoints and polysemy: Linking adversative and causal meanings of discourse markers”. In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause–Condition–Concession–Contrast*. *Cognitive and Discourse Perspectives*. Berlin: Mouton de Gruyter, 257–282.
- Schwenter, S. & Traugott, E. 2000. “Invoking scalarity: The development of in fact”. *Journal of Historical Pragmatics*, 1 (1), 7–25.
- Traugott, E. 1999. “The rhetoric of counter-expectation in semantic change: A study in subjectification”. In A. Blank & P. Koch (Eds.), *Historical Semantics and Cognition*. Berlin: Mouton de Gruyter, 177–213.
- Traugott, E. & Dasher, R. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Verhagen, A. 2000. “Concession implies causality, though in some other space”. In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause–Condition–Concession–Contrast: Cognitive and Discourse Perspectives*. Berlin: Mouton de Gruyter, 361–380.

Authors' addresses

Liesbeth Mortier
 Université catholique de Louvain (UCL)
 Place B. Pascal 1
 1348 Louvain-la-Neuve
 Belgium

Liesbeth.Mortier@uclouvain.be

Liesbeth Degand
 Université catholique de Louvain (UCL)
 Place B. Pascal 1
 1348 Louvain-la-Neuve
 Belgium

Liesbeth.Degand@uclouvain.be