

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/20679> holds various files of this Leiden University dissertation.

Author: Koier, Elizabeth

Title: Interpreting particles in dead and living languages : a construction grammar approach to the semantics of Dutch *ergens* and Ancient Greek *που*

Issue Date: 2013-03-28

CHAPTER 4

A corpus study of *ergens*

4.1 Introduction

In the previous chapter, we found that in most cases there was a trigger in the direct context of *ergens* which seemed to point the addressee in a specific direction. In some cases, this trigger was found outside the direct context. This raises the question whether collocations and triggers generally are the same thing. In order to find that out, we will compare the collocations of *ergens* in three corpora, which will be described below, with the triggers that were found above. This will also give us the chance to see whether the conclusions that were drawn above about the contextual requirements for a specific interpretation hold for a larger set of examples.

4.1.1 The corpora

The synchronic corpora that were used in this study are the Corpus Gesproken Nederlands (CGN) (Corpus of Spoken Dutch, 1998-2004), the Eindhoven corpus (a corpus from the sixties and early seventies) and a number of literary novels¹ all but one published in the past ten years (A. Grunberg *De Asielzoeker* (2003), G. Reve *Op weg naar het einde* (1963), H. Haasse *Sleuteloog* (2002), R. Dorrestein *Het duister dat ons scheidt* (2003), H. Bouazza *Paravion* (2003)).

The Corpus Gesproken Nederlands is a 8,916,272 word corpus and in total it contains 2757 instances of *ergens*. The corpus has been built out of the components given

¹With thanks to Suzanne Fagel, who was so kind as to let me use her digital versions of these novels.

in table 4.1 below.² These components consist of various settings in which spoken language is used. The frequency scores given in table 4.1 are the number of instances of *ergens* in that component from the Netherlands and Belgium respectively divided by the total number of words from the Dutch or Belgian part of the corpus times 10.000.

Component	frequency ergens BE	frequency ergens NL
a. Spontaneous conversations ('face-to-face')	4.6	3.8
b. Interviews with teachers of Dutch	5.1	2.7
c. Spontaneous telephone dialogues 1	4.4	3.6
d. Spontaneous telephone dialogues 2	4.5	3.2
e. Simulated business negotiations	n/a	(2.0) ³
f. Interviews, discussions, debates (broadcast)	1.9	2.5
g. (political) Discussions, debates, meetings (non-broadcast)	1.0	0.9
h. Lessons recorded in the classroom	2.4	2.6
i. Live (sports) commentaries (broadcast)	1.5	0.2
j. Newsreports, reportages (broadcast)	1.1	2.2
k. News (broadcast)	0.4	0.3
l. Commentaries, columns, reviews (broadcast)	1.4	1.4
m. Ceremonious speeches, sermons	2.4	0.0
n. Lectures, seminars	2.8	1.3
o. Written text	2.3	2.3
Total instances <i>ergens</i>	3.5	2.8

Table 4.1: The components of the CGN and the frequency of *ergens* (number of instances of *ergens*/number of words in that part of the corpus * 10,000) in Netherlandic Dutch and Belgian Dutch.

In table 4.1 we see that the incidence of *ergens* is higher in the Belgian part of the corpus than in the Netherlandic part. This is especially the case in the first part of the corpus (a-h), which consists of spontaneous speech in all kinds of different settings. As we will see later, this might be explained by the broader use of *ergens* in Belgian Dutch. In addition, *ergens* is found more frequently in spontaneous and informal settings, like spontaneous conversations than in more formal settings like the news, ceremonial speeches and broadcasted interviews and discussions.

The Eindhoven corpus, a Netherlandic Dutch corpus, consists of the following genres:

²For a more elaborate description of how this corpus was designed see <http://lands.let.kun.nl/cgn/ehome.htm>.

³There are no Belgian recordings for section e, therefore these instances have been left out of the comparison.

Component	Tot. words	<i>ergens</i>	<i>ergens</i> / Tot.words x10,000
camb (corpus of officials)	56,679	0	0
dbl (daily newspapers)	152,384	13	0.9
gbl (family magazines)	155,771	26	1.7
gtl (spoken language)	183,622	28	1.5
obl (opinion magazines)	157,830	12	0.8
pwe (non-specialist scientific texts)	148,788	8	0.5
rno (novels and short stories)	167,649	30	1.8
Total	1022723	117	1.1

Table 4.2: The components of the Eindhoven corpus and the frequency of *ergens* in each of those components

In table 4.2, we see again the tendency for *ergens* to be used more frequently in informal genres. It is found most frequently in family magazines and in spoken language as well as in novels and short stories.

The corpus of novels consist of Dutch novels that were selected on two criteria: 1. They had to be stylistically as diverse as possible and 2. their publishing dates should be as close as possible while still fitting requirement 1. The corpus of novels has the following characteristics:

Author	Words	<i>Ergens</i>	<i>ergens</i> / words x10,000
H. Bouazza <i>Paravion</i> (2003)	52,862	11	2.1
R. Dorrestein <i>Het duister dat ons scheidt</i> (2003)	87,928	30	3.4
A. Grunberg <i>De Asielzoeker</i> (2003)	120,259	23	1.9
H. Haasse <i>Sleuteloog</i> (2002)	39,403	6	1.5
G. Reve <i>Op weg naar het einde</i> (1963)	55,340	22	4.0
Total	355,792	92	2.6

Table 4.3: The novel corpus and the frequency of *ergens* in its components

In table 4.3 the number of instances of *ergens* is relatively high in the work of Dorrestein and Reve. In the former this can be linked to the frequent use of modal *ergens*, in the latter the theme of the work (travelling letters) is an explanation for the frequent use of locative *ergens*.

As was argued more extensively above, we need more categories of *ergens* than were described by the dictionaries in order to categorize all cases in the corpora. The frequencies of the uses of *ergens* can be found in table 4.4.

	Novels	%	Eindhoven Corpus	%	CGN	%
Place	62	67.4	79	67.5	1863	67.6
Prep. Obj	21	22.8	19	16.2	296	10.7
Time	2	2.2	2	1.7	123	4.5
Modal	7	7.6	12	10.2	139	5.0
<i>Feelings</i>	5	5.4	4	3.4	51	1.8
<i>Point of view</i>	2	2.2	2	1.7	37	1.3
<i>Somehow</i>	0	0	6	7.3	51	1.8
?	0	0	6	5.1	307	11.1
Scale	0	0	0	0.0	14	0.5
Total	92		117		2757	

Table 4.4: The frequencies of the types of interpretations of *ergens* in three corpora. The numbers in the subcategories of modal add up to the total number of modal *ergens*.

As can be seen from table 4.4, the frequency of modal interpretations differs between the three corpora, whereas the frequency of locative *ergens* is almost constant. This suggests that the presence of modal *ergens* may be more dependent on genre and/or register.

What we may conclude from this survey of the three corpora is that *ergens* is relatively frequent in informal language. On the one hand, this may have to do with the indefiniteness of *ergens*. The use of an indefinite may not be appropriate in more formal settings. On the other hand, it may have to do with modal *ergens*, which is characterized by the Van Dale dictionary as colloquial language, suggesting that it is mainly found in informal, spoken language. This last explanation is strongly supported by the fact that 113 of the 139 modal instances of *ergens* in the CGN (81%) occur in the spontaneous speech sections of the CGN, whereas these sections contain only 59 % of the total number of words in the corpus.

4.1.2 Context, collocations and interpretation

A reference to the context of a form can refer to all kinds of information, from the common ground of interlocutors to a specific linguistic phenomenon in the direct environment of that form. All factors that fall under the term *context* may play a role in interpretation. In order to find out which of the various features of the context plays a role in the interpretation of *ergens*, we will start with the linguistic context. More specifically, we will start with the collocational behavior of *ergens*.

The definition of a collocation

Textbooks on collocational analysis (e.g. Manning and Schütze, 1999) generally mention three types of requirements for words to be considered a collocation: 1. the distance between a word and the target word. 2. the statistical probability that forms occur together and 3. the requirement that the collocation is part of a phrase. The

first criterion is mainly a way to define the term collocation, the second has as its main goal to rule out that the frequency of a co-occurrence is due to chance and the third is a way to avoid 'pollution' of the results in terms of clusters of function words that occur very frequently together, but which are due to grammatical rules that are not relevant to the study at hand (e.g. articles that are collocated with nouns, coordinators that co-occur with articles, etcetera).

However, in the case of collocation research on particles, not all of these criteria are applicable. First, there is the fact that particles are not part of a phrase. That would not be a problem, if we knew what other part of speech we are looking for. However, the definition of a construction or a trigger is so broad that we cannot always restrict ourselves to specific parts of speech here. This problem is discussed by Van der Wouden (2001) and he also shows that despite these problematic issues the phenomenon of collocation is highly relevant for particles. Second, the statistically necessary assumption that words are divided randomly over a text, is obviously not correct. Languages do have word ordering principles. For this reason, and because of the large sample size, it is relatively easy to get statistically significant results (Manning and Schütze, 1999, 166). This means that we cannot draw definitive conclusions from the fact that two words occur together significantly more often than chance (Gries, 2010). The best we can do is rank the collocations with respect to each other. This leaves us with the question of how we can adapt the collocational technique to the study of particles.

There are two ways to avoid the pitfalls that are normally avoided by using the criterion of 'being part of a phrase'. The first is to make a contrastive analysis between two groups. For instance, if the co-occurrence of a form with *ergens* may be due to word order principles, we might contrast a group of instances in which *ergens* is in the first position of the clause with a group in which *ergens* is not in this position, to see whether its collocate is restricted to the cases in which *ergens* is in first position. Another option is to compare two groups that are interpreted in a different way and see whether their collocations correlate with each interpretation. This will allow us to see whether the occurrence of a specific form in the context of *ergens* correlates with one specific interpretation.

The second way to avoid overgeneralizing is to be conservative. If two forms occur frequently within 10 words of each other, there may be many explanations for this situation which have little to do with the interpretation of *ergens*. However, if a form occurs very frequently directly next to *ergens* and does not have a very general grammatical function (like an article or an auxiliary), the chance is already higher that we have found a meaningful collocation that plays a role in the interpretation of *ergens*. This is why I will start with a conservative number of words and, when expanding the scope, I will continue to be critical as to why the forms are frequently found together.

A more general way of restricting ourselves is of a more semantic nature. Although it is interesting to see which forms have co-occurrences with *ergens* often enough to be a potential collocation, a semantic relationship between *ergens* and the form under consideration makes it far more likely for a form to be a trigger for a spe-

cific interpretation of *ergens*. Practically, this means that when looking at a list of the most frequent collocations of *ergens*, it is important to remain critical and to investigate whether the frequent co-occurrence may be due to other processes which have nothing to do with the interpretation of *ergens*. The fact that a frequent collocation is part of the same semantic field as one of the interpretations of *ergens* or the frequent presence of the same type of words in the list are indications that there may be a link between the frequent co-occurrence of that form with *ergens* and the interpretation of *ergens*. A closer investigation will reveal whether a specific interpretation of *ergens* correlates with the presence of some of these collocates.

A last question that needs to be answered is when a form can be considered to be the same as another form. In the case of verbs, for instance, it is not uncommon to consider all inflections of a verb as one collocational entity. This raises the question when we consider a group to belong together. Can we, for instance, call the group of mental state predicates one collocational entity? Since in this study the goal is to find the general patterns in language that may be connected to an interpretation, we will look at this more generalized type of collocation as well as at more specific collocations.

4.2 A collocation analysis of *ergens* by category

The first thing I did when I started to work on *ergens*, was to categorize all instances of *ergens* in the CGN. This can be seen as a native speaker judgement on the interpretation of *ergens*. This categorization was not influenced by the results from the surveys and will be used below to test the results of the surveys on a larger set of data.

4.2.1 Locative *ergens*

As we saw in the previous chapter, the presence of locative prepositions and verbs seemed to have a strong influence on the interpretation of *ergens* as locative. It would be interesting to find out whether this generalization can account for a large number of the locative instances of *ergens*. However, if we take all instances of locative prepositions into account that are in a 5 word distance of *ergens*, we run the risk of counting the instances of *ergens* double that have more than one preposition in their direct surroundings or a preposition and a locative verb. Therefore, we will start with just one position at a time. In 28%⁴ of the locative instances of *ergens* in the CGN, *ergens* is directly followed by a locative adposition. Another 4% is added to that if we take into account verbs that are frequently found next to *ergens* and imply a location

⁴There are 1863 instances of locative *ergens* according to my classification. These numbers may be influenced a bit by the fact that there are some instances of locative *ergens* that have non-locative instances of *ergens* in their direct context, which could not be excluded from this count. However, since these cases constitute less than 2% of the locative instances of *ergens* this should not blur the overall picture.

	Number	% with respect to locative <i>ergens</i>
locative adpositions	1125	60%
locative verbs	796	43%
locative adverbs	183	10%
Total locative instances of <i>ergens</i>	1863	100%

Table 4.5: The number of locative word classes five words before and five words after *ergens* in the CGN. The percentages are percentages of the number of that category with respect to the locative instances of *ergens*. The percentages do not add up to 100% because they are not mutually exclusive.

like, for instance, *go, put down, lie, eat* and *drink*⁵. This means that by taking into account only one position in the sentence and two categories of words, we can already account for almost one third of the instances of locative *ergens*. If we now broaden our horizon a bit and add the adverb *anders* ‘else’, we can add another 12 %. In total we have accounted for 44% of the locative instances of *ergens*, looking only at the position directly after it.

If we do the same for the position directly before *ergens*, we find that this position gives less clear results. All locative adverbs, verbs and adpositions in this position can account for 12% of the instances of locative *ergens*.

What we can conclude from this survey of the two positions directly next to *ergens* is that collocational information is likely to play an important role in the interpretation of locative *ergens*. The words directly before and after it already provide cues for the interpretation of about 50%⁶ of the cases. This suggests that if we broaden our scope to five words before and after *ergens* we are likely to cover an even larger proportion of the data, especially because the total number of locative adpositions, locative verbs and locative adverbs in the ten words around *ergens* is very large, as can be seen from table 4.5.

However, this does not mean that there is always a direct and simple connection between collocations and the interpretation of *ergens*. As was already shown in the previous chapter, some features of the context can dominate over others. For instance, the collocation *daar ergens* ‘somewhere there’ is in most cases a locative expression, but it can also be used as a temporal expression, as in example (1).

- (1) Speaker 1: Vorig jaar was 't een zaterdag. 't is al wat 'k weet.
last year was it a Saturday. It is all what I know.

Speaker 2: uh in alle geval 't moet daar ergens zijn omdat de weekends
in all case it must DAAR ERGENS be because the week-ends

⁵ Although eating and drinking do not automatically imply a location, in Dutch they are often used to express the phrase *go to/be at a restaurant*. In this use they are often accompanied by a location and, if the precise location is not particularly relevant, *ergens*.

⁶ A few of the collocations before and after *ergens* overlap, this is why this percentage is a bit lower than the sum of the percentages mentioned above.

dus een beetje opgeschoven zijn hè? ja.
 thus a bit moved are isn't it? Yes.

Last year it was on a Saturday. That's all I know. Anyway, it has to be *daar ergens*, because the week-ends have moved a bit, haven't they? Yes.⁷

In example (1) the only reason to think that *daar ergens* 'somewhere there' is a temporal reference is the overall context. The speakers are discussing at what date a celebration will be held. Both the text before and after this fragment continues to be about that date, which makes the only reference that makes sense a temporal one. If one isolated the clause *in alle geval... daar ergens zijn* from its overall context, *ergens* would get a locative interpretation due to the presence of the locative marker *daar* 'there'. The date is seen as a point on a time scale, which is again a clear example of the metaphorical nature of the relationships between some of the interpretations of *ergens*.

Examples such as (1) make clear why we need a theory of interpretation that consists of more factors than a meaning and general pragmatic principles. On the one hand, it is clear that this temporal interpretation is dependent on the broader context, which would make it a good candidate for a pragmatic addition to a basic meaning. On the other hand, however, an attempt to translate the phrase into German or Italian already shows that a temporal use of a locative phrase like this is not readily accepted in all languages. This means that Dutch speakers need to know that their language allows for the temporal use of *daar ergens*, which implies that the choice for a temporal interpretation is not purely pragmatic, but has a conventional component.

In addition, example (1) is a good illustration that although collocations may play an important role in the interpretation of a particle like *ergens*, there is by no means an automatic relation between the presence of a collocation and a specific interpretation. The final interpretation is also influenced by the expectations of the addressee, who will choose the interpretation that makes the most sense, even if - given a specific collocation - statistically the chance would be higher that another interpretation was intended.

4.2.2 Temporal, scalar and about/around-functions of *ergens*

In the CGN, the temporal interpretation of *ergens* is always connected to a direct temporal reference in the context, generally directly before or after *ergens*, but as was shown by the previous example, in rare cases it can also be somewhat further away. This means that there is a relatively strong one to one relationship between this type of interpretation and temporal markers in the context.

However, as was also discussed in the previous chapter and illustrated by example (1) the boundaries between the temporal use of *ergens* and its scalar use are

⁷From CGN comp-d/vl/fv700207.

blurry. The connection between the temporal interpretation and the locative interpretation may well be a metaphorical place on a scale and there are many examples in which these interpretations overlap.

However, there are also cases in which *ergens* is used to denote a place on a non-temporal scale. Very frequently (in 8 of 14 instances) the preposition *tussen* 'between' is present. In other examples we find *rond* 'around' or *in de [number]* 'in the [number]'. These expressions generally denote imprecision, which explains why this category also has an overlap with the category *about/around* not only semantically, but also with respect to its collocational characteristics.

The connection between the temporal, scalar and *about/around* interpretations is interesting because it shows that the *about/around* category is more connected to the non-modal interpretations than to the modal interpretations. This suggests that the development to *about/around* was an extension that arose apart from the modal categories.

The category *about/around* is the only category in which there is a clear collocational difference between Belgian Dutch and Netherlandic Dutch. The collocation *ergens iets* 'about/around' is used by speakers from Belgium only. Although there are other Belgian examples that also sound a bit odd to a speaker from the west of the Netherlands (see also section 3.3), this is the only collocation in which there is a clear cut division between speakers from the two countries. An example is given in (2). This interpretation of *ergens iets* is found 19 times in the CGN.

- (2) Speaker 1: Ik bedoel geen superwinsten daar maar dus uh wel wel wel
I mean no super profit there but so ehm PTCL PTCL PTCL

veiligheid. Speaker 2: nee nee maar wel veilige belegging hè.
safety. No, no but PTCL safe investment PTCL.

Speaker 1: en dan ergens iets negentien procent of zo uh beleggen
and then ERGENS IETS nineteen percent or so ehm invest

die maar op de beurs.
they PTCL on the stock market

Speaker 1: I do not mean super profit but some safety. Speaker 2: No, no but you do mean a safe investment, don't you? Speaker 1: and they invest *ergens iets* nineteen percent or something on the stock market.⁸

4.2.3 Modal *ergens*

Now we will turn to the modal interpretations of *ergens*. These interpretations are the most difficult to describe and distinguish, therefore it will be interesting to see whether they show collocational characteristics that are tied to their category. One

⁸From CGN comp-a/vl/fv400809.

CGN	1st	%	2nd	%	3rd	%	eq. 1st	%	Gen 2nd	%	?	%	Total	%
CGN														
Feelings	26	18	1	1	17	11	2	1	5	3	0	0	51	34
Point of view	6	4	0	0	30	20	0	0	1	1	0	0	37	25
Somehow	7	5	3	2	32	22	2	1	6	4	1	1	51	34
?	1	1	1	1	6	4	0	0	1	1	0	0	9	6
													148	100
<u>Eindhoven</u>														
Feelings	1	8	0	0	1	8	2	17	0	0	0	0	4	33
Point of view	0	0	0	0	2	17	0	0	0	0	0	0	2	17
Somehow	1	8	0	0	5	42	0	0	0	0	0	0	6	50
?	0	0	0	0	0	0	0	0	0	0	0	0	0	0
													12	100
<u>Novels</u>														
Feelings	2	29	0	0	3	43	0	0	0	0	0	0	5	71
Point of view	1	14	0	0	1	14	0	0	0	0	0	0	2	29
Somehow	0	0	0	0	0	0	0	0	0	0	0	0	0	0
?	0	0	0	0	0	0	0	0	0	0	0	0	0	0
													7	100

Table 4.6: Person of the predicate in clauses with modal *ergens* in the CGN, Eindhoven corpus and novels. Eq 1st is *equivalent to 1st person*, i.e. constructions that contain a first person pronoun in an oblique case. Gen 2nd is a generically used 2nd person.

of the arguments that we used in the previous chapter was whether the situation described pertained to the propositional content of the clause or to the subjective opinions of the speaker. In other words, whether *ergens* referred to the subjective view of the speaker or to the physical world. It was suggested that the difference between the *feelings*-option and the *point of view* option was dependent on the strength of the subjective connection between the expressed view and the speaker. We also described some features of the context that seemed to be characteristics of the groups of modal interpretations. The *feelings* interpretation seemed to coincide frequently with first person pronouns and mental state predicates/ subjective verbs. From table 4.6 it is clear that 1st person verbs or constructions that contain an accusative first person pronoun (eq. 1st) are much more frequent in this category, as was expected on the basis of the survey results.

In table 4.7 below, an overview is given of the type of verbs or predicates that are used within a clause in which modal *ergens* is found. From this table it is clear that mental state predicates are more frequently found in the *feelings* category, as was expected on the basis of the surveys. Statements with a copula seem to be more frequently used in the *point of view* category. The interpretation *somehow* has the largest number of predicates that do not have a particularly subjective value. These findings confirm our tentative conclusions from the surveys.

A feature that seemed to distinguish the *point of view* interpretation from the other modal options, was the presence of an adversative expression after *ergens* such as *aan de andere kant* 'on the other hand'. The expression *aan de andere kant* 'on the other hand' or a variation on this expression is found 5 times in the CGN in the *point*

CGN	MSP	%	Subj. Pred.	%	Copula	%	Other	%	Total	%
Feelings	40	27	1	1	5	3	5	3	51	34
Point of view	7	5	2	1	23	16	5	3	37	25
Somehow	5	3	5	3	10	7	31	21	51	34
?	1	1	0	0	0	0	8	5	9	6
									148	100
<u>Eindhoven</u>										
Feelings	2	17	2	17	0	0	0	0	4	33
Point of view	0	0	0	0	2	17	0	0	2	17
Somehow	0	0	0	0	4	33	2	17	6	50
?	0	0	0	0	0	0	0	0	0	0
									12	100
<u>Novels</u>										
Feelings	4	57	1	14	0	0	0	0	5	71
Point of view	0	0	2	29	0	0	0	0	2	29
Somehow	0	0	0	0	0	0	0	0	0	0
?	0	0	0	0	0	0	0	0	0	0
									7	100

Table 4.7: The types of verbs in the three corpora split by modal category. MSP is an abbreviation of mental state predicate.

of view category (14% of the instances in this category), 2 times in the *feelings* category (4%) and never in the *somehow* category. Since the total number of instances per modal group is not very high, it is hard to draw any firm conclusions, but we may say that the corpus data do support the results from the surveys.

If we zoom in on the individual collocations somewhat further, it becomes clear that the verb *vinden* ‘be of the opinion, to find, to think’, is frequently found together with the *feelings* interpretation of *ergens* (13 times out of 51 instances in the CGN, i.e. 25%) whereas in the *point of view* and *somehow* categories it is only found 1 time (respectively 3 % and 2% out of 37 and 51 instances).

A collocation that is found in the *somehow* category only is the verb *kloppen* ‘be correct, right’. Within this category *ergens* and *kloppen* are found in each other’s vicinity (5 words before and 5 words after *ergens*) four times (8% of the 51 cases of *somehow* in the CGN). The collocations *wel* and *toch* are more problematic. Although both particles are frequently found around modal *ergens*, this seems to be a general characteristic of (modal) *ergens*, since both particles occur frequently in all three modal categories and in cases of locative *ergens* as well. Therefore, they cannot be used to distinguish the categories from each other.

Summarizing, we have found the following collocational characteristics for each of the modal markers:

For the category *in someone’s feelings or thoughts* the following characteristics are more frequently found than in the other categories:

1. first person pronouns
2. mental state predicates and subjective predicates
3. the verb *vinden* 'to be of the opinion'

The following characteristics are more frequently found in the *from a certain point of view* category:

1. the frequent use of the copula
2. subjective adjectives in combination with the copula
3. variants of the phrase *aan de andere kant* 'on the other hand'
4. the *ergens is dat* 'ergens is that' construction

The following characteristics are more frequently found in the *somehow* category:

1. third person verbs
2. non-subjective predicates
3. the verb *kloppen* 'to be right, correct'

The next step is to see what proportion of the examples is covered by these features. If we take for every category the items of the list of characterizations above and exclude the cases that show these characteristics, we are left with a relatively small number of cases. For the *feelings* category there are 7 cases left out of 51, which is 13 % in this category. For the *point of view* category there are more cases left, 12, which is 32% of this category. In the case of the category *somehow* this is 6 cases or 12%. This means that although the correlation is not perfect, these features allow people to predict which interpretation is likely to be the correct one in most cases. We will now discuss the exceptions which do not show the general characteristics of their category to find out why the exceptions are part of a category.

Four of the seven exceptions within the *feelings* category in the CGN do not contain first person pronouns or mental state predicates. These examples seem to show some sort of free indirect speech. The speaker is talking about the thoughts and feelings of a group of people or another person, suggesting that he has access to their thoughts and feelings. This type of text accounts also for almost all instances of modal *ergens* in the novels.

- (3) en uh dat vermoeit de mensen op een gegeven ogenblik waarschijnlijk
en ehm that tires out the people at a certain moment probably

dus ook ergens wel. Want om dat jaren vol te houden dat zal dus niet
then also ERGENS PTCL. For to that years continue that will PTCL not

meevallen.
 turn out for the better
 And ehm that will *ergens* tire people out at some point probably. For it will be hard to continue that for years.⁹

In example (4), *ergens* is found in a question which is putting words in the addressee's mouth. This makes the necessary connection between the addressee's feelings and the sentence.

- (4) Speaker 1: Da 's leuk hoor. Maar is dat een stiekem een droom *ergens* of?
 That is nice PTCL. But is that a secret a dream ERGENS or?
 Speaker 2: Nee joh 'k doe dat nou maar niet meer.
 No dude I do that PTCL PTCL not anymore
 Speaker 1: That's nice. But is that some sort of hidden dream *ergens* or what?
 Speaker 2: No dude I won't do that anymore.¹⁰

In example (5) it seems that with *ergens* the minds of a group of people is meant, but the following sentence shows that the speaker is also talking about him/herself. This makes this use of *ergens* more in line with the other cases of the *feelings* category.

- (5) Maar *ergens* leeft dat idee nog wel zonne. Nee maar ik ja bij mij dus
 Maar ERGENS lives that idea PTCL PTCL PTCL. No but I, Yes with me PTCL

 ook. En we zitten dus op de moment wel in in de juiste omgeving
 also. And we sit PTCL at the moment PTCL in in the right environment

 om publiek daarvoor te vinden.
 to public for that to find.

 But *ergens* that idea is still alive. Also for me. And at this moment we are in the right position to find public for it.¹¹

In the last example the speaker describes the feelings of another person, which in the end results in direct thought. The use of *ergens*, however, is still in the description. The speaker seems to create a mental space in which the person rebels against a certain image he has of himself.

- (6) Ik denk dat hij ook wel uh *ergens* rebelleert tegen de idee 'k ben nu
 I think that he also PTCL ehm ERGENS rebels against the idea I am now

 iemand die elke dag antidepressiva neemt. 'k kan mij dat voorstellen
 someone who every day antidepressants takes. I can me that imagine

⁹CGN comp-b/nl/fn000096.

¹⁰From CGN comp-b/nl/fn000105.

¹¹From CGN comp-c/vl/fv701091.

dan dat je daar niet graag uh al dat dat hij zichzelf zo niet
 then that you there not willing ehm all that that he himself as such not

ziet, 'k zal het zo een beetje uitdrukken hè.
 sees, I will it this way a little express PTCL.

I think he rebels against the idea I am now someone who takes antidepressants every day. I can imagine that you, ehm that he does not see himself that way, let me put it somewhat more that way.¹²

The exceptions from the *point of view* category show either first person pronouns, which is generally a feature of the *feelings* category, or they do not show any other characteristic of the *point of view* category as described above. However, most of these instances do contain a feature that was already described above: contrasts. In many cases the contrastive elements that give rise to a contrastive interpretation are particles: *ook*, 'also'; *ook wel*, 'also focal particle'; *toch wel*, focal particle focal particle; *wel*, focal particle; *maar*, 'but, however'; *maar ja*, 'but particle'. However, these elements have, just like *ergens*, several functions, which means that we cannot separate them so easily from their counterparts with other functions. Much of the contrast also comes from the contrastive content of the sentences.

There are several cases which contain first person pronouns and mental state predicates as in example (7). The only reason they get a *point of view* interpretation instead of a *feelings* interpretation seems to be the contrastive elements. It may be that this is also the connection between the two types of interpretations. Two points of view within one person can be expressed as two places (or points of view) within one mind. If this mind is not explicitly mentioned, we are only left with the points of view, which may or may not belong to the speaker or the character, as in example (8).

- (7) Ik vind 't gewoon provoceren hoor. Maar ja ik
 I am of the opinion it simply provoke PTCL. But PTCL I
 vind ik vind 't ook wel humor hebben ergens.
 am of the opinion I am of the opinion it also PTCL humour have ERGENS.
 I think it is simply provocation. But I think it is also funny *ergens*.¹³

- (8) Speaker 1: Maar ja ergens heeft ze natuurlijk ook wel een beetje gelijk
 But PTCL ERGENS has she of course also PTCL a bit right
 hè. Speaker 2: Ja da 's waar.
 PTCL. Yes that is true.
 Speaker 1: But *ergens* she does of course have a point. Speaker 2: Yes, that is right.¹⁴

¹²From CGN comp-c/vl/fv901155.

¹³From: CGN comp-a/nl/fn000989.

¹⁴From: CGN comp-d/vl/fv900226.

The last category we discuss is *somehow*. Although there are relatively few exceptions to the general collocational trends in this category, they show less uniformity than the exceptions in the other two categories. In example (9) we find a first person pronoun, which normally is one of the characteristics of the *feelings* category.

- (9) Ik heb met mijn bemoeienissen evenveel bedrijven de grond in
I have with my interferences as many companies the ground in

geboord als dat ik er weer op de rails heb gezet. Het één kan
drilled as that I there again on the track have put. The one can

zeker niet los gezien worden van het ander en aangezien
certainly not apart seen be from the other and since

dat een nietig eindresultaat oplevert deug ik dus ergens niet.
that a zero final result results in am any good I thus ERGENS not.

I have brought down as many companies as I have put back on the rails. The one can certainly not be seen apart from the other and since this results in a zero end-game, I *ergens* am no good.¹⁵

In this example, it is not completely clear whether *ergens* is to be seen as a point of view from the perspective of the people judging the speaker or that *ergens* modifies the extent to which the speaker is no good. Since in both cases *ergens* is part of the subjective view of the opponents, this is not completely clear. The second interpretation can also be seen as an implication of the first one, but they don't need each other. It is the second reading that allows us to interpret *ergens* as 'somehow' here instead of a *point of view* interpretation.

A comparable example, which also shows a first person pronoun, although it is interpreted as *somehow*, can be found in (10).

- (10) Hoe dan ook vraagt het voor ons een aanpassing om dat te begrijpen
Anyway, asks it from us an adjustment to that to understand

maar ik denk dat uhm dat wij ergens neutraler spreken. Zij
but I think that ehm that we ERGENS more neutral speak. They

beseffen niet dat zij echt een vorm van dialect spreken.
realize not that they really a form of dialect speak.

Anyway, it requires us to adjust to understand it, but I think that we *ergens* speak more neutral. They do not realize that they really speak some sort of dialect.¹⁶

¹⁵From CGN comp-o/nl/fn001148.

¹⁶From CGN comp-b/vl/fv400115.

In this example, both a *point of view* interpretation and an interpretation of *ergens* as *somehow/ up to a certain extent* is possible. A reason for this ambiguity may be that *ergens* is not part of the same clause as the mental state predicate *denk* ‘think’. This results in a thought that is expressed in indirect speech. This way, *ergens* may be interpreted either as referring to the personal view of one of the critics or as a modification of the statement in the indirect speech.

This concludes our survey of the collocational characteristics of modal *ergens*. In the next section, the results of this chapter will be summarized.

4.3 Conclusion

What we can conclude from the study of the linguistic contexts of modal *ergens* is that although collocations do not automatically lead to a specific interpretation, there are patterns that may lead language users in a certain direction. On the basis of the surveys and the corpus study, the following information seems to be used to interpret *ergens*. Within the three corpora, the Corpus Gesproken Nederlands (CGN), the Eindhoven corpus and a corpus of novels, the characteristics that are mentioned below could account for more than two thirds of the modal examples.

When there are locative markers that allow for an indefinite locative interpretation and there are no indications that a speaker is making a subjective statement, the addressee will interpret *ergens* as locative. On the other hand, if there are indications that the speaker is making a subjective statement and *ergens* is not needed in an argument position, *ergens* will be interpreted as modal, even if there are locative elements in the context. When there are temporal markers in the direct context of *ergens* or a locative interpretation does not fulfill Grice’s maxim of relevance, whereas a temporal interpretation would fulfill this requirement, *ergens* will be interpreted as temporal. When there is a scalar preposition or expression in the direct context of *ergens* and the interpretation is not temporal, the interpretation will be scalar. However, if the scale that is used is very rough and there are other indications of estimation, like hesitation, the interpretation will be *about/around*.

Modal interpretations can be recognized from the fact that a locative interpretation is impossible or irrelevant. In addition, there are often subjective markers around. If these markers are directing attention to the personal view of the speaker or another person, for instance by means of first person pronouns and mental state predicates, the interpretation will be *in someone’s feelings or thoughts*. If there is less connection to the personal view of the speaker, for instance in a copular construction or if there is a strong emphasis on contrastive views, for instance by means of the phrase *aan de andere kant*, or one or more of the particles *ook, wel, toch, maar*, and the content of the proposition is still clearly subjective, the interpretation will be *from a certain point of view*.

If the content of the proposition is not strongly subjective, often recognizable by third person verbs, and there is little contrast between views, little connection to a person’s personal view and no relevance in a locative interpretation, the interpretation will be ‘*somehow*’. If there is for some reason very little context and this

At some place:

- locative adpositions
- verbs implying a location

At some point on a scale:

- scalar adpositions
- scalar constructions
- numbers

In someone's feelings or thoughts:

- mental state predicates
- first person pronouns
- subjective predicates
- the verb *vinden* 'to be of the opinion'

Somehow:

- Third person verbs
- Mainly non-subjective predicates
- the verb *kloppen* 'to be right, correct'

At some moment:

- temporal markers

About/around:

- numbers
- scalar adpositions
- scalar constructions
- expressions of vagueness

From a certain point of view:

- frequent use of the copula
- subjective adjectives in combination with impersonal copula constructions
- variants of the phrase *aan de andere kant* 'on the other hand'
- the *ergens is dat* 'ergens is that' construction
- contrastive markers like *ook*, *wel*, *maar*, *toch* also, *focus particle*, but, *focus particle*

context does not provide any (constructional) cues, the most concrete and most frequent interpretation of *ergens* will be chosen: the locative interpretation.

These criteria only work within the general limits of the relevance criterion and the expectations of the addressee from the common ground. That is why the characteristics that were mentioned would not be enough to teach a computer how to interpret *ergens*. In addition, more data and surveys may refine this description.

