

Preverbal *ge-* in Old and Middle English

Thomas McFadden

ZAS, Berlin

1 Background

This paper aims to work toward a proper understanding of the role of preverbal *ge-* in Old English (henceforth OE) and its disappearance in the course of Middle English. This prefix is reminiscent of its cognates in Modern German and Dutch (also written *ge-*)¹ in its distribution, but even a cursory examination of the details reveals it to be quite distinct, as we will see. The proper characterization of that distribution, and of its diachronic development, has proven to be extremely difficult. I have thus carried out a large-scale corpus study using the *York-Toronto-Helsinki parsed corpus of Old English prose* (Taylor et al. 2003)² and the *Penn-Helsinki parsed corpus of Middle English*, 2nd ed. (Kroch & Taylor 1999),³. This paper will report the results of the first phase of the project, involving patterns in the data that could be identified primarily on the basis of automatic searches in the corpora. These patterns serve as the empirical basis for an improved description of the facts, and ultimately for a more precise theoretical hypothesis about the nature of *ge-* than any found in previous work. I will propose specifically that *ge-* in OE was the default realization of a *res(ultative)* head in the sense of Ramchand (2008). It is important to note at the outset that the results I will present are preliminary. The predictions of the proposed analysis must still be tested in a planned second phase of the project, involving a close reading of a manageably sized sample of relevant clauses from the corpus, examining in particular details of interpretation that are not reflected in the corpus annotation and cannot be searched for electronically. The second phase will also apply more sophisticated statistical methods to the data, in particular a multivariate analysis that can test more systematically for interactions between variables than I have been able to do here.

¹ These are also written *ge-*, but have quite different pronunciations, due to the effects of regular sound changes in the three languages. While (Standard) German has [gə], and Dutch has [χə], the Old English pronunciation was something like [jɛ] or [jə]

² <http://www-users.york.ac.uk/~lang22/YCOE/YcoeHome.htm>

³ <http://www.ling.upenn.edu/hist-corpora/>

1.1 Germanic comparisons

The prefix *ge-* is the descendant of a common Germanic element **ga/gi*, the cognates of which are found in most of the other (old) Germanic languages. The traditional analysis of the original prefix in its preverbal use in Proto-Germanic is that it marked perfectivity or resultativity (Streitberg 1891; Lloyd 1979; van Kemenade & Los 2003 and see further below), but our understanding is complicated by the distinct developments it has undergone in the daughter languages.⁴

In the Scandinavian branch, the reflex of **ga/gi* was lost prehistorically, just like the other unstressed verbal prefixes, so that by the Old Norse period it no longer appears as a prefix.⁵ In Gothic, the cognate *ga-* was a prominent part of the language, in the verbal morphology and elsewhere. Its distribution shows strong similarities to what we find in OE and the old West Germanic languages, though the relatively small corpus available to us makes it difficult to achieve a clear understanding of its exact role. Nonetheless, there is an extensive (mostly older) literature attempting to do just that (see especially Streitberg 1891; Lloyd 1979; Eythórsson 1995 for data and discussion).

In the old West Germanic languages, the prefix was well preserved, again playing an active role in the verbal morphology. In OE in particular, *ge-* was still an obviously productive part of the verbal system, with a behavior similar to that in Gothic, which has eluded straightforward characterization. It is clear that it interacts with aspect and aktionsart, argument structure and (lexical) semantics, but as is often the case when such factors are involved, and in particular when we don't have access to native speaker intuitions to help sort things out, it's far more difficult to say which of these factors defines its primary function. Indeed it may well be that some further, more abstract factor, which remains to be identified, is most important, with the superficial ones previously identified being only indirectly related. Matters are not helped by the fact that *ge-* was

⁴ The prefix could in fact attach to nouns, adjectives and even adverbs in the older languages, and the different uses clearly are ultimately related. However, their common origin lies quite far back in time, and by the time of the historically attested languages, the connection between them is rather tenuous, so that it no longer makes sense to propose a unified syntax/semantics for them. For this reason, I will focus here solely on the preverbal uses of the prefix and set the other ones aside.

⁵ What actually seems to have happened is that pre-tonic unstressed vowels syncopated, and most of the resulting consonant clusters were subsequently simplified by deleting the first consonant. This effectively deleted the entire prefix in most instances, and even in cases where allowable initial clusters would have resulted, the consonant portion of the prefix was eliminated analogically. However, in at least some cases where the origin of this consonant in a productive prefix was obscured, i.e. where the relationship to non-prefixed forms of the same verb was severed, it remained.

lost as a productive element in the course of the Middle English (henceforth ME) period, as I will discuss in some detail in Section 4. Present Day English (henceforth PDE), like Old Norse, has only traces of *ge-* where its origins as a prefix have become obscure, mostly in non-verbal uses (e.g. the *e-* in *enough*). This means that we can get no clues from the modern language, where native speaker intuitions would have been available.

German and Dutch differ on this point, having preserved the prefix as a productive element of verbal morphology to the present day. This brings both advantages and disadvantages, however, as these languages have clearly undergone significant innovations in the use of the prefix. Due to the availability of native speakers (as well as vast corpora), we can achieve a very accurate description of the distribution of *ge-* in these languages. Whether this is really helpful for understanding OE *ge-* remains doubtful, however. While we find at first glance similarities between the three languages on this point, a quick comparison shows crucial differences in the details, hence taking the German and Dutch patterns as a starting point is likely to be misleading.

In Modern German and Dutch, the prefix has two distinct verbal functions. My presentation here will be based on the German facts, but the characterization applies in its essentials to both languages.⁶ First, it appears productively on the participial form of the verb used (along with various auxiliaries) to form the periphrastic perfects and passives (henceforth the PPP, for perfect/passive participle). In German, this is restricted to verbs with stress on the first syllable, but is otherwise completely regular and productive.⁷

(1) a. **Initial stress:**

| Inf. | | PPP | Gloss |
|-------------|---|------------|--------------|
| 'zählen | ~ | gezählt | 'count' |
| 'trinken | ~ | getrunken | 'drink' |
| 'mailen | ~ | gemailt | 'e-mail' |

b. **Non-initial stress:**

| Inf. | | PPP | Gloss |
|-------------|---|------------|--------------|
| er'zählen | ~ | erzählt | 'tell' |
| spa'zieren | ~ | spaziert | 'walk' |

⁶ The distribution of *ge-* is not identical in the two languages (see fn. 7), but the differences (as far as I am aware) are not relevant to the broader points being made here.

⁷ The relevant condition is different in Dutch, where it is not stress-placement, but the presence or absence of a competing inseparable prefix that matters.

Second, it appears non-productively as a derivational prefix, as in the examples in Table 1. Note that there is no consistent semantic relationship between such *ge-* prefixed verbs and their non-prefixed counterparts, and in some cases it is obscure what connection there could be between them at all. This of course

Table 1: German derivational *ge-*

| | | | |
|-----------------|-------------|-------------------|---------------|
| <i>brauchen</i> | ‘need, use’ | <i>gebrauchen</i> | ‘use’ |
| <i>fallen</i> | ‘fall’ | <i>gefallen</i> | ‘please’ |
| <i>hören</i> | ‘hear’ | <i>gehören</i> | ‘belong to’ |
| <i>denken</i> | ‘think’ | <i>gedenken</i> | ‘commemorate’ |
| <i>stehen</i> | ‘stand’ | <i>gestehen</i> | ‘confess’ |

represents a fairly typical pattern with derivational morphology. When used productively, such morphology typically makes a consistent semantic contribution, but over time derived forms can become ‘lexicalized’, i.e. the connection to their original constituent parts can be weakened or lost, with the lexical item that was once productively derived henceforth following a distinct path of diachronic semantic development from its parts. Consider in this connection examples like English *transmission*, which have long played an important role in the theoretical discussion of derivational processes. This development seems to be especially favored when the derivational morphology involved ceases to be used productively in the language, as is the case in the German pattern at hand. While some of the other unstressed prefixes like *ver-*, *ent-* and *zer-* can still be used, at least occasionally, to create new verbs with reasonably predictable meanings, there is no such productive use of *ge-*.⁸

Whatever account we may adopt of these developments in German in particular or of ‘lexicalization’ in general, it seems clear that we must distinguish the two uses of *ge-* just described. We have nothing to gain from attempting to unify, in our synchronic grammar of German, the productive inflectional use on PPPs with the non-productive derivational use, or its counterparts in our synchronic grammar of Dutch. Instead, a historical explanation of the homonymy we observe seems appropriate. That is, we have here two prefixes that are pronounced the same because they represent divergent developments of a single prefix in the prehistories of these languages, but no longer have any connection in the synchronic grammars. A clue to the distribution of the original unified prefix comes from certain notable patterns in earlier historical stages of Ger-

⁸ So e.g. *sich ver-X* can be created to mean something like ‘to X in an erroneous fashion’, *ent-X* to mean ‘to de/un-X’, and *zer-X* to mean ‘to X to pieces’.

man. First, in Old High German we still find semi-productive alternations in the presence of the prefix with certain verbs, showing a consistent semantic effect. E.g. we find *sizzen* ‘be sitting’ alongside *gisizzen* ‘sit down’. Second is the fact that, into the Middle High German period, the prefix is **not** used in the PPP of certain inherently perfective/resultative verbs like *quëman* ‘come’ (see Braune & Reiffenstein 2004 on both points). We’ll see in Section 2.6 that this somewhat surprising fact is paralleled in Old English, and I’ll offer an account for it in Section 3.3.

Turning back to OE now, the main impression of similarity with German and Dutch comes from the fact that, here as well, *ge-* quite typically shows up on PPPs, e.g. in periphrastic ‘perfect’ constructions:⁹

- (2) ac hēo hæfde **ge**coren Crist hyre to brȳdguman
but she had chosen Christ her to bridegroom
‘...but she had chosen Christ as her bridegroom.’
(coelive,+ALS_[Eugenia]:349.401)
- (3) forðan þe his gebedda **ge**faren wæs of līfe
because his bedfellow gone was from life
‘...because his wife had passed away.’
(coelive,+ALS_[Maur]:131.1567)

However, unlike in German and Dutch, *ge-* is not an integral part of the PPP. A significant portion of PPP examples lack the *ge-* prefix, even though the verb is morphologically compatible with it, as in (4).

- (4) sē gelēaffulla Oswald ... wæs \emptyset -cumen to Cynegylse
the faithful Oswald ... was come to Cynegils
‘The faithful Oswald ... had come to Cynegils.’
(coelive,+ALS_[Oswald]:131.5455)

Furthermore, *ge-* is found with rather high frequency on other verb forms beyond the PPP, e.g. on the finite past form in (5):

⁹ Examples taken from the corpora in the Penn-York series will be given with the source information from their ID tag in the corpus, which identifies the specific corpus, the source text and information on the page and sentence number. E.g. (2) comes from the YCOE, from Ælfric’s Lives of Saints, in particular the life of St. Eugenia, and is (part of) token 401 (tokens correspond roughly to sentences and are numbered sequentially through the sample for a particular text) found on page 349 of the print edition used in the creation of the corpus. I direct the reader to the documentation of the corpora for details on how to identify source details on the basis of the ID tags.

- (5) Sē geworhte ealle þing
he created all things
'He created all things.'
(coaelive,+ALS_[Christmas]:66.51)

Crucially, *ge-* in examples like this does **not** look like the derivational *ge-* in German verbs like *gefallen* (though it may be similar to its ancestor). As we will see, it is far too frequent, too widespread across lexical verbs, and too regular in its semantic contribution.

1.2 Prior approaches

The wide range of rather tricky facts about the distribution of OE *ge-* and its cognates has led to an array of proposals about its meaning, function and grammatical status, which are typically quite abstract, often vague, and sometimes even completely empty. The most extreme view was that *ge-* was simply meaningless. According to Thomas Benson, *Vocabularium Anglo-Saxonicum* (1701) “*Ge-* apud Saxones semper fere superfluum” (‘Among the Saxons, *ge-* is almost always superfluous’). While later scholars have generally not adopted this view, they have repeatedly thought it worthy of mention as an indication of how difficult it is to pin the prefix down. Another approach, which was developed at length by Lindemann (1970), but has found little resonance, is that *ge-* expresses abstract direction. According to him, “the action expressed by any verb to which [*ge-*] is prefixed is directed toward some thing or in a direction forward and outward” [p. 37].

The most popular proposal, not just for OE, but also for its cognates in the other old Germanic languages, is that *ge-/ga-/gi-* is a marker of perfective aspect. The idea is associated in particular with Wilhelm Streitberg, who was inspired by comparisons with aspectual prefixes in the Slavic languages Streitberg (1891: etc.). This view soon fell into disrepute, as it became clear that the distribution of the prefixes in the Germanic languages was rather different in detail from the patterns observed in Slavic. An approach in terms of aspect has been rehabilitated more recently, however, in work (e.g. by Lloyd 1979; Eythórsson 1995) that attempts to take into account the differences in the aspectual systems of the two language families. The guiding idea here is that, difference from what is found in Slavic does not imply that what is involved is not aspect. Lloyd in particular discusses the differences between Slavic and early Germanic aspect in detail and argues in the end that a single system of primitives underlies both systems, but that they differ in exactly which distinctions in those primitives they mark. Thus, if the term ‘perfective’ is reserved for

the aspectual category marked by prefixation in Slavic, what we find marked by the Germanic prefix *ge/ga/gi* must be something else, which Lloyd dubs the ‘complexive’.¹⁰

The approach that I will argue is most promising is related to these aspectual proposals, but operates at a slightly different level. This is that *ge-* expresses resultativity. While it is often true that perfectives are resultative, the converse is often not the case. That is, an inherently resultative predicate (like *break the window*) can quite easily be used imperfectively, e.g. with an additional progressive component (like *I was just breaking the window when the police arrived*). A role for resultativity in the prefix *ge-* has been proposed e.g. by van Kemenade & Los (2003) for various stages of Dutch and English. This also connects to analyses of verbal particles in some of the modern Germanic languages (see e.g. Ramchand & Svenonius 2002; McIntyre 2003), where parallels are quite apparent.

In Section 2 I will present in detail the findings of my corpus study on the distribution of our prefix in OE and in particular how it interacts with other identifiable properties of the verb or the clause containing it that might be expected to be relevant. Then in Section 3 I will propose a specific version of the resultative analysis of *ge-*, taking it to be the default spellout of Ramchand (2008)’s *res* head. I will show how this can account not only for the broad facts of its distribution, but also for the small but consistent details that don’t seem to fit with a naïve idea of resultativity. I turn in Section 4 to developments in the ME period, when the prefix began to disappear from the language. While I will not be able to propose a clear answer to why *ge-* was lost, I will show that my analysis of its structural status in OE can provide some insight into how this loss interacted with other contemporary changes, in particular in the periphrastic perfect system.

2 The distribution of preverbal *ge* in OE

In this section I will present the main OE data which will serve as the basis for the analysis developed in subsequent portions of the paper. I am reporting here the results of a study of the complete YCOE (Taylor et al. 2003), a corpus con-

¹⁰ What Lloyd claims specifically is that the aspectual function of Gothic *ga-* was to mark “the complexive report of a completed action” [p. 85]. He uses the term complexive aspect to refer to when the observer/reporter describes an eventuality from a perspective that is outside of time, from which she can view all phases of a completed action. Discussing how this differs from the Slavic perfective (which for Lloyd is built on, but distinct from, the complexive) would take us too far afield here.

taining approximately 1.5 million words of OE prose, tagged for part of speech other grammatical features and parsed for syntactic structure. The searches were carried out with the CorpusSearch program (<http://corpussearch.sourceforge.net>), which was designed to search corpora in the Penn-Treebank format on the basis of lexical forms, POS tags and structural notions like dominance and precedence. This makes it possible to execute quite sophisticated searches efficiently.

In searching the corpus, I adopted the following strategy. First, I ran searches to identify the main verb of each clause, since this is where preverbal *ge-* is primarily expected to appear.¹¹ Second, I classified each such main verb according to whether it was prefixed with *ge-*, with some other prefix which might be expected to be in competition, e.g. *be-* or *for-*, or not prefixed at all.¹² Then I ran a series of searches to code each clause for properties that might be expected to have an effect on the distribution of *ge-*, either because they were reported to do so in the previous literature, or because I could imagine a plausible connection to other factors that had been discussed. Finally, for each such property I checked to what extent it did in fact correlate with the choice between *ge-* and no prefix. As we will see in the following, there is a great deal of variation in whether and how much these factors actually affect the distribution of *ge-*, which can provide clues about what the function of the prefix actually was.

2.1 The broad patterns

The first and simplest result of my searches of the corpus is that *ge-* is extremely common. Out of 166,544 clauses examined,¹³ 42,366 (25.4%) had *ge-* on their main verb. Even setting aside PPPs, 30,862 of 153,622 main verbs (20.1%) had *ge-*. This is our first and perhaps clearest indication that *ge-* really did play an active and central role in the OE verbal system, and that it was quite un-

¹¹ The various (pre-)auxiliary verbs of the language essentially never have *ge-*, and I have elected, for the time being, to set aside verb forms used as attributive adjectives. This is by no means an innocuous move, but was motivated by the need to keep the volume of data manageable.

¹² The logic here is fairly simple – with some very few exceptions, a single verb form cannot bear two prefixes simultaneously, so *ge-* will be simply ruled out on a form prefixed with *be-*. This means that when we're trying to figure out the conditions on the appearance of *ge-*, a form prefixed with *be-* doesn't tell us the same thing as a form with no prefix. For the first stage of the research being reported here, I chose to set aside the examples with other prefixes, and to focus on differences between forms with *ge-* and forms with no prefix, though I did in some cases consult the data on other prefixes with specific lexical verbs, as we will see.

¹³ As noted above, this excludes all clauses whose main verb has a prefix other than *ge-*.

like its modern German and Dutch cognates. Its productive use is quite simply not restricted to the participial forms showing up in periphrastic perfects and passives.

The second broad result is that I have found some basic confirmation of previous claims: *ge-* tends to be favored in environments suggestive of perfectivity, telicity and resultativity, and disfavored elsewhere. At this rough level, it is difficult to distinguish among the different prior proposals, as there is a tendential relationship between perfectivity and resultativity. It is only when we consider certain details that a particular kind of resultative analysis begins to stand out as the most appropriate. So before I introduce the specifics of my proposal, I will here go through the data on the different factors that are relevant to the distribution of *ge-*, or at least might have been expected to be.

2.2 The form of the main verb

Given the fact that *ge-* is a crucial component of the productive formation of PPPs in modern German and Dutch, we have reason to suspect that the specific form of the main verb will have some effect on the frequency of *ge-* in OE. We will see that this is indeed the case, and that here as well, *ge-* is extremely common on PPPs. However, a very important recognition is that, unlike in German and Dutch, *ge-* does show up, in significant numbers on **all** morphosyntactically defined forms of the main verb. Since this is unfamiliar from the other languages, it will be instructive to have examples demonstrating its appearance in each of them here.

First, we have the present participle, which corresponds to the PDE form in *-ing*, but was usually formed in OE with the suffix *-nde*, most commonly occurring with a form of auxiliary *be*.

- (6) & swā wæs **geendiende** þis wilwendlice liif
and so was ending this temporal life
'and thus [he] was ending this temporal life'
(cobede,Bede_4:9.286.1.2881)

Then we have the *to*-infinitive, i.e. an infinitival form of the verb (in OE typically in the dative, ending in *-enne*), preceded by *to*. As in PDE, these can appear as the sole verb of a non-finite clause or in a periphrastic construction with auxiliaries *have* or *be*.

- (7) and næfð nāne mihte [menn **to gehælenne**]
and not-has no power [men to heal]
'and has no power to heal men' (coaelhom,+AHom_4:86.569)

Next come finite verb forms, i.e. those marked for tense and agreement.

- (8) Sē geworhte ealle þing
he created all things
'He created all things.' (coelive,+ALS_[Christmas]:66.51)

Then we have bare infinitives, which typically show up in combination with the pre-modals and other auxiliaries, but are also found occasionally in certain non-finite clause types (see Los 2007 for detailed discussion of the different types of infinitive and their distribution in OE and ME).

- (9) þæt menn hit **gehyran** mihton;
that men it hear may
'so that men may hear it' (coelhom,+AHom_1:451.233)

We also have examples of imperative forms of the verb, which show up in much the same contexts as in PDE.

- (10) and þonne þū eft cymst, **geoffra** þīne lac.
and when you again come, offer your sacrifice
'and when you come back, make your offering'
(coelhom,+AHom_16:19.2269)

And finally we have the PPP, which occurs primarily in periphrasis with auxiliary *have* or *be* in the perfect or passive.

- (11) ac hēo hæfde **gecoren** Crist hyre to brȳdguman
but she had chosen Christ her to bridegroom
'...but she had chosen Christ as her bridegroom.'
(coelive,+ALS_[Eugenia]:349.401)

All the same, even though *ge-* can show up on any form of the verb, its distribution across them is not even. Rather, there are marked differences in its frequency on the various forms, as shown in Table 2, ordered from the least to the most frequently occurring with *ge-*. What we see is that the distribution of *ge-* is skewed in the same direction as it is in German and Dutch, but not nearly so far. That is, *ge-* is extremely frequent with PPPs, but it is nowhere near categorical. Similarly, it is extremely infrequent with present participles, but far from categorically absent. This is in line with what we might expect if *ge-* has to do with perfectivity or resultativity — perfects and passives, where the PPP mostly shows up, tend to be perfective and resultative, while present participles tend to be used for the description of ongoing states or activities, which are generally imperfective and need not be resultative.

Table 2: Form of the main verb

| Form | <i>ge-</i> | no | % <i>ge-</i> |
|--------------------|------------|--------|--------------|
| Pres. Ptc. | 107 | 1493 | 6.7 |
| to Infin. | 430 | 2177 | 16.5 |
| Finite | 23723 | 102434 | 18.8 |
| Bare Infin. | 4329 | 11188 | 27.9 |
| Imperative | 2273 | 5468 | 29.4 |
| PPP | 11504 | 1418 | 89.0 |

2.3 Tense, mood and negation

Given that the morphosyntactic form of the main verb is relevant for the distribution of *ge-*, we might expect this to carry over to more fine-grained distinctions like tense and mood. Indeed, if *ge-* has something to do with perfectivity or resultativity, it is plausible to think that tense in particular will make a difference, given well-known interactions between tense and aspect. E.g. in PDE, the progressive appears happily in both present and past, but the unmarked aspectual form is heavily restricted in the present, so that with eventive predicates a non-episodic (typically habitual) reading is forced:

- (12) a. I was eating the dosa.
 b. I am eating the dosa.
- (13) a. I ate the dosa.
 b. # I eat the dosa.

Example (13b) is odd out of the blue, because a habitual reading with a definite object requires a special context, e.g. as the answer to a question like ‘What do you usually eat at this restaurant?’ On the other hand, the unmarked aspect is perfectly felicitous with an episodic reading in the past, as in (13a). Thus if OE *ge-* is involved somehow with aspect, we might also expect it to interact with tense marking.

Looking at the numbers on the frequency of *ge-* according to the tense of the clause, we find that past tense has a small but clear favoring effect ($\chi^2 = 573.782$, $p < .0001$), as seen in Table 3.¹⁴ This small effect can perhaps

¹⁴ Note that what is at stake here is the tense of the clause, not necessarily of the verb form which we are considering with respect to whether it bears *ge-*. In the various periphrastic constructions, the tense of the clause will be marked on the (highest) auxiliary, whereas

Table 3: Tense of the finite verb

| Tense | <i>ge-</i> | no | % <i>ge-</i> |
|--------------|-------------------|-----------|---------------------|
| Pres | 15496 | 54329 | 22.2 |
| Past | 23105 | 60878 | 27.5 |

be understood if *ge-* has something to do with completion, or the reaching of some result state. Whether or not a particular eventuality goes to completion is perhaps easier to judge and also more relevant when it lies in the past than in the present or future. Note relatedly that in many languages a perfective-imperfective distinction is restricted to or at least predominantly expressed in past forms, not present ones.

Mood seems less relevant. OE finite verbs distinguish subjunctives from indicatives, though a large number of forms are actually ambiguous between the two. Clear subjunctive forms have a somewhat higher frequency of *ge-* than clear indicatives, but ambiguous forms show the highest frequency, as can be seen in Table 4. The differences here are statistically significant¹⁵ but there

Table 4: Mood of the finite verb

| | <i>ge-</i> | no | % <i>ge-</i> |
|--------------------|-------------------|-----------|---------------------|
| Indicative | 23051 | 75443 | 23.4 |
| Subjunctive | 5234 | 14713 | 26.2 |
| Ambiguous | 9857 | 22840 | 30.1 |

is reason not to take this too seriously. The difference between indicative and subjunctive is small in absolute terms, and the fact that ambiguous forms don't end up in between the two clear categories suggests that something else is going on here.¹⁶

it is the main verb that we are examining for the presence of *ge-*. Of course, only finite clauses will have tense marked at all, the various infinitival and participial clauses lacking such marking. The numbers here thus do not add up to the same total as in some of the other tables, where all clauses are reflected.

¹⁵ For example, for the difference between indicatives and subjunctives $\chi^2 = 73.396$, $p < .0001$.

¹⁶ I.e. if there is a real difference between indicatives and subjunctives in their behavior with *ge-* then, assuming that the ambiguous category contains a mixture of forms intended as indicatives and forms intended as subjunctives, it should show a behavior somewhere in between the two categories. What may be going on here actually is that there is an interaction with tense marking. Perhaps the largest group of forms that are systematically ambiguous be-

Another possibility I investigated is whether negation has an effect on the appearance of *ge-*. While the connection may not seem so obvious, Postma (2002) has shown that the cognate prefix *ghe-* in Middle Dutch actually had a preverbal use as a negative polarity item, so we might imagine that OE would exhibit something similar. Table 5 shows, however, that it does not. Here again,

Table 5: Polarity of the clause

| | <i>ge-</i> | no | % <i>ge-</i> |
|---------------------|------------|-----------|--------------|
| Negative | 3123 | 9953 | 23.9 |
| Non-negative | 39756 | 116964 | 25.4 |

the difference we see between negative and non-negative clauses is statistically significant ($\chi^2 = 14.082$, $p = .0002$), but this is only because the data set is so large. The absolute difference we see is tiny, and in any case goes in the opposite direction of what Postma observed for Middle Dutch.

2.4 Prepositional and adverbial elements

A standard diagnostic of aspectual distinctions, in particular those at the Aktionsart level, is the licitness of certain PPs and adverbials. For example, we can classify predicates according to whether they can felicitously combine with PPs like *in an hour* or *for an hour* (roughly, telic vs. atelic predicates). Given the size of the corpus being examined here and the complexities of the possible PPs, it was not feasible at this initial stage to divide things up according to specific PPs and adverbials.¹⁷ It is, however, relatively easy to search for whether a particular clause contains a PP or adverbial of any kind. Consider then the frequencies for *ge-* under these conditions reported in Table 6. The prefix *ge-* is more frequent with both PPs and adverbials than without, and again in both cases the difference is statistically significant (for PPs $\chi^2 = 822.793$, $p < .0001$, for adverbs $\chi^2 = 44.395$, $p < .0001$). Again, this is at least in large part simply due to the extremely large sample sizes. When we look at the actual size of the difference, we find that with adverbs it is just 1.5%, whereas with PPs it is

tween indicative and subjunctive are the past 3sg forms of weak verbs. If ambiguous forms tend to be past, then perhaps they tend to take *ge-* at a higher rate for this reason rather than anything having to do with mood. This is one of many points that will be investigated in the proper multivariate analysis planned for future work.

¹⁷ This will require taking a sample out of the full corpus to examine in more detail, and thus will be considered for the next stage of the project.

Table 6: PPs and adverbs in the clause

| | <i>ge-</i> | no | % <i>ge-</i> |
|------------------|------------|-------|--------------|
| PP | 20850 | 51655 | 28.8 |
| No PP | 22029 | 75262 | 22.6 |
| Adverb | 15687 | 44175 | 26.2 |
| No adverb | 27192 | 82742 | 24.7 |

6.2% increase, over four times the effect. The effect with adverbs is certainly small enough that, though statistically significant, it may not be particularly meaningful. That with PPs may reflect something more real, but this cannot be determined until a more detailed examination of a sample of the corpus is carried out.¹⁸

2.5 Date

Given the fact that *ge-* disappears in the course of the ME period, it is worth looking into whether it was already in retreat in OE. In other words, we want to see whether the frequency of *ge-* correlates with the date of a particular text. For the OE period it is exceedingly rare that we know exactly when a particular text was composed, or even when the surviving manuscripts (which are usually later, often considerably so) were written. The best we can do with a reasonable degree of certainty is typically a range of a few decades. Furthermore, as the collection of available texts is quite limited, if we tried to assign too narrow a date range to each, we would end up with unacceptably small amounts of data for any particular range. To deal with these issues, historical corpora usually set up a limited number of longer periods, and assign each text to one of these, so that they can be grouped together for analysis in roughly contemporaneous samples that are large enough to do basic statistics. The *YCOE* corpus, following the Helsinki Corpus on which it is largely based, divides OE into four periods, the first from the earliest attestations to 850, the second from 850 to 950, the third from 950 to 1050, and the third from 1050 to 1150, after which the ME

¹⁸ One possibility is that the frequency of *ge-* is increased by the presence of complement PPs which affect the aktionsart of the main predicate, e.g. by introducing a *telos*, as in *Sandra swam* vs. *Sandra swam to the shore*. The idea would be that such a complement role is more frequently played by PPs than by adverbials. Since complement vs. adjunct status of these elements is not consistently annotated in the corpus, this cannot be searched for automatically, but must be determined by examination of individual examples.

portion of the Helsinki corpus picks up. In what follows, I will collapse together periods one and two because the first contains too little text to be comparable.

With this background, consider now the frequency of preverbal *ge-* for the three periods of OE given in Table 7. The first thing to notice is that the frequen-

Table 7: Date of the text

| Period | <i>ge-</i> | no | % <i>ge-</i> |
|------------------|-------------------|-----------|---------------------|
| pre-950 | 15330 | 44721 | 25.5 |
| 950-1050 | 19865 | 58792 | 25.3 |
| 1050-1150 | 74 | 430 | 14.7 |

cies in the first two periods are remarkably close, indeed essentially identical. In fact, even with such large numbers, what little difference there is comes out as not being statistically significant ($\chi^2 = 1.341$, $p = .2468$). This shows us quite clearly that *ge-* was completely stable, neither increasing nor decreasing for most of the OE period. The second thing is that there is a clear and sudden drop in frequency between the second and third periods. At first glance, this suggests that perhaps at this point the decline of *ge-* had begun that would continue through the ME period. However, and this is the third crucial point, we must be very careful about how seriously we take this data point. There are two reasons to be skeptical. For one thing, note that the absolute number of clauses we're looking at here is very small in comparison to the first two periods — two orders of magnitude smaller. This is because of the extremely limited amount of English text that survives from the relevant period, due to the collapse in the use of written English following the Norman conquest. The number of examples here is not so small that valid statistical reasoning is impossible (and the difference does come out as statistically significant, $\chi^2 = 29.707$, $p < .0001$), but it is small enough that we do have to be concerned about the representivity of the sample.

The second reason to be skeptical is also related to the Norman conquest and the collapse of the Old English scribal tradition. By the late tenth century, a quite consistent, standardized form of West Saxon OE had established itself as the written form used in all centers of writing around the country. Like most standardized languages, it was quite conservative, and by the time of the conquest clearly no longer reflected the contemporary spoken language in many respects. When the scribal tradition was broken by the Norman conquest, the propagation of this standard ceased or was at least severely weakened. Thus, to the extent that people wrote anything in English at this time, they were far more

heavily influenced by their own speech than by the inherited standard. This means that there is a quite sharp break in nearly every property of the language we find in the texts of late OE and early ME compared to what came before. But this clearly does **not** imply that there was a series of catastrophic changes in the living language at the time. Rather, there was a catastrophic change in writing practices, such that the written language suddenly caught up with perhaps two centuries' worth of more gradual changes in the spoken language. In other words, to the extent that the difference between the second and third periods in the table above reflects a real change in the language, it was probably more gradual than it appears, spread out over the previous century. In any case, what we can conclude is that *ge-* was nowhere near disappearing, and indeed was stable for most of OE, but that its decline was beginning towards the end of the period.

2.6 The identity of the main verb

The area where the most interesting results are to be found is in the lexical identity of the main verb, i.e. the item on which the prefix *ge-* either does or doesn't appear. Before we get to the data, a quick word on the corpus work it took to get at it is in order. Unfortunately, identifying lexical verbs with searches of the *YCOE* corpus is not nearly as easy as searching for most of the other factors being discussed here. This is essentially because there are vastly more lexical verbs in OE (or of course any language) than there are verb forms, polarity categories, chronological periods etc. More to the point, the *YCOE* is not lemmatized, i.e. beyond functional items and a few other extremely common items, the lexical identity of a word form has not been determined and is not tagged. The POS-tag on a verb will indicate that it is a verb and provide information about its morphological form and grammatical properties, but not whether it is a form of e.g. *speak* or *eat* or *desire*. This means that identifying specific lexical items requires writing queries that can recognize them based on their form, which is time-consuming and prone to errors. There is an effect of diminishing returns as well due to Zipf's Law, which tells us that the vast majority of lexical verbs will only appear a handful of times in the corpus, many of them only once. Thus it is only really worth the effort of doing the work to recognize a small number of extremely common items.

That is precisely what I did, writing my queries to recognize 31 lexical verbs based on their forms, in addition to 'have', 'be', 'do' and the pre-modals, which are specifically tagged in the corpus. This successfully identified 54,380 verb forms with specific verbal lexemes as indicated in the tables below. There

were an additional 74,395 verb forms that were not recognized, and these are listed below as ‘unclassified’. Table 8 is a complete list of all of the verbs, plus the unclassified category, sorted in ascending order of the percentage of *ge-* vs. no prefix, to give an overview of the situation.

Table 8: Identity of the main verb

| Verb | Gloss | <i>ge-</i> | no | % <i>ge-</i> |
|-------------------|--------------|------------|-------|--------------|
| (pre-)modals | | 0 | 2575 | 0.0 |
| <i>bēon/wesan</i> | ‘be’ | 1 | 30127 | 0.0 |
| <i>habban</i> | ‘have’ | 13 | 5053 | 0.3 |
| <i>cuman</i> | ‘come’ | 29 | 4687 | 0.6 |
| <i>sendan</i> | ‘send’ | 15 | 947 | 1.6 |
| <i>drincan</i> | ‘drink’ | 17 | 779 | 2.1 |
| <i>etan</i> | ‘eat’ | 26 | 538 | 4.6 |
| <i>fēran</i> | ‘go’ | 64 | 1282 | 4.8 |
| <i>beodan</i> | ‘command’ | 58 | 1001 | 5.5 |
| <i>cwedan</i> | ‘say’ | 553 | 9145 | 5.7 |
| <i>gān</i> | ‘go’ | 128 | 1927 | 6.2 |
| <i>secgan</i> | ‘say’ | 288 | 3783 | 7.1 |
| <i>sprecan</i> | ‘speak’ | 90 | 1134 | 7.4 |
| <i>andwyrdan</i> | ‘answer’ | 37 | 457 | 7.5 |
| <i>sellan</i> | ‘give’ | 362 | 2182 | 14.2 |
| <i>wunian</i> | ‘dwell’ | 202 | 1093 | 15.6 |
| <i>writan</i> | ‘write’ | 30 | 158 | 16.0 |
| <i>sittan</i> | ‘sit’ | 131 | 649 | 16.8 |
| <i>seoþan</i> | ‘boil’ | 3 | 14 | 17.6 |
| <i>fōn</i> | ‘grasp’ | 159 | 728 | 17.9 |
| <i>hatan</i> | ‘call/order’ | 560 | 2309 | 19.5 |
| <i>dōn</i> | ‘do’ | 933 | 3681 | 20.2 |
| <i>slēan</i> | ‘smite’ | 87 | 325 | 21.1 |
| <i>faran</i> | ‘go’ | 241 | 772 | 23.8 |
| <i>acsian</i> | ‘ask’ | 156 | 486 | 24.3 |
| <i>nemnan</i> | ‘name’ | 217 | 601 | 26.5 |
| <i>þencan</i> | ‘think’ | 328 | 777 | 29.7 |

Table 8: Identity of the main verb (continued)

| Verb | Gloss | <i>ge-</i> | no | % <i>ge-</i> |
|---------------------|--------------|------------|-------|--------------|
| <i>wyrcaŋ</i> | ‘work, make’ | 523 | 1227 | 29.9 |
| unclassified | | 30885 | 43510 | 41.5 |
| <i>tēon</i> | ‘pull’ | 89 | 119 | 42.8 |
| <i>weorþan</i> | ‘become’ | 1001 | 979 | 50.6 |
| <i>niman</i> | ‘take’ | 1434 | 1265 | 53.1 |
| <i>halgian</i> | ‘hallow’ | 392 | 108 | 78.4 |
| <i>hāelan</i> | ‘heal’ | 626 | 110 | 85.1 |
| <i>sēon</i> | ‘see’ | 2714 | 188 | 93.5 |

The first thing to note here, which is a very important message to take away, is that the variation is massive. It goes from verbs that are literally never prefixed with *ge-* to one that bears it a full 93.5% of the time, and fills out the space in between fairly evenly.

Now let’s zoom in a bit to get a better idea of what’s going on in detail, by splitting up that full range of variation into a few smaller chunks. In the each of the tables to follow I will include the ‘general total’ at the bottom for comparison, i.e. the overall frequency of *ge-* across all verbs. First, at the very bottom of the range, we’ll take the ‘auxiliary’ verbs. Note that what we’re looking at here are not the actual auxiliary uses of these verbs (where *ge-* also never appears), but rather their main verb uses, since in general here we are interested in whether the main verb of a clause bears *ge-*. The frequency of *ge-*

Table 9: Main verb uses of ‘auxiliary’ verbs

| Verb | Gloss | <i>ge-</i> | no | % <i>ge-</i> |
|----------------------|--------|------------|--------|--------------|
| (pre-)modals | | 0 | 2575 | 0.0 |
| <i>bēon/wesan</i> | ‘be’ | 1 | 30127 | 0.0 |
| <i>habban</i> | ‘have’ | 13 | 5053 | 0.3 |
| general total | | 42366 | 124178 | 25.4 |

with these verbs given in Table 9 is essentially zero, and given the very high frequency of their appearance in the corpus, we can be quite confident in the accuracy of this result. Now, given the background assumption from previous work that the *ge-* prefix has something to do with perfectivity or resultativity,

such extremely low frequencies are not really surprising. All of these verbs are statives, and so are not expected to appear in perfective or resultative uses. This is a case where the precise details of what is behind the distribution of *ge-* are perhaps not so crucial, since we expect it to be incompatible with statives under most reasonable proposals. As we move further into the other verbs on the list, different specific theories will make clearly different predictions, and the ways that specific verb classes behave will help us to choose among them. When considering each group I will continue to initially speak in terms of what is surprising or expected based on a vague notion of perfectivity or resultativity, and then make use of the surprises to help lead us to a specific proposal.

Turning now to the lexical verbs, we consider first in Table 10 the group with markedly low frequency of *ge-*.¹⁹ The rarity of *ge-* with some of these is

Table 10: Lexical verbs with low frequency of *ge-*

| Verb | Gloss | <i>ge-</i> | no | % <i>ge-</i> |
|----------------------|-----------|------------|--------|--------------|
| <i>cuman</i> | ‘come’ | 29 | 4687 | 0.6 |
| <i>sendan</i> | ‘send’ | 15 | 947 | 1.6 |
| <i>drincan</i> | ‘drink’ | 17 | 779 | 2.1 |
| <i>etan</i> | ‘eat’ | 26 | 538 | 4.6 |
| <i>fēran</i> | ‘go’ | 64 | 1282 | 4.8 |
| <i>beodan</i> | ‘command’ | 58 | 1001 | 5.5 |
| <i>cwedan</i> | ‘say’ | 553 | 9145 | 5.7 |
| <i>gān</i> | ‘go’ | 128 | 1927 | 6.2 |
| <i>secgan</i> | ‘say’ | 288 | 3783 | 7.1 |
| <i>sprecan</i> | ‘speak’ | 90 | 1134 | 7.4 |
| <i>andwyrðan</i> | ‘answer’ | 37 | 457 | 7.5 |
| general total | | 42366 | 124178 | 25.4 |

again relatively easy to understand. The group of speech verbs e.g. (*cwedan*, *secgan*, *sprecan* and *andwyrðan*) are all plausibly essentially activities in Ak-

¹⁹ The cut-off between this group and the next is of course arbitrary. Unlike with the previous group of verbs, which could be distinguished as auxiliaries, independent of their behavior with *ge-*, there is no clear grammatically defined division here. I have chosen to draw the line between *andwyrðan* and *sellan* as there is a marked jump in frequency of *ge-* between them, from 7.5% to 14.2%. The next cut-off point, between *wyrðan* and *tēon*, also corresponds to a jump in frequency, from 29.9% to 42.8%, and also marks off the lexical verbs that combine with *ge-* at a higher frequency than the mass of unclassified ones.

tionsart terms, meaning that they aren't telic and thus don't normally have resultative uses. The two 'go' verbs are plausibly also activities, though it will depend here quite a bit on the details of individual contexts. Motion verbs are frequently activities in their basic uses, but relatively flexible in Aktionsart terms, being easily converted to accomplishments e.g. by the addition of appropriate PPs indicating a goal.

Initially unexpected is the behavior of *sendan*, *drincan* and *etan*. We would expect these, especially the latter two, to be telic in most cases, and thus if *ge-* marks perfectives or resultatives, it seems that it should be common here. I will come back to *drincan* and *etan* in Section 3.3, where we will see that their behavior can actually provide some support for a particular analysis of the function of *ge-*. With *sendan*, the story seems to be a bit simpler. While *ge-* is the most common verbal prefix of its type in OE, and the one whose distribution and meaning present the most challenges, it is really just one member of a larger system, as alluded to briefly above. It turns out that with *sendan*, other prefixes – specifically *a-* and *on-* – are overwhelmingly used in telic contexts, not *ge-*. Table 11 provides the numbers for *sendan* with various prefixes vs. with no prefix. We see then that while *sendan* may have a markedly low frequency with

Table 11: Frequency of various prefixes with *sendan*

| Prefix | Frequency |
|-------------------------------|------------------|
| none | 979 |
| <i>a-</i> | 405 |
| <i>on-</i> | 96 |
| <i>ge-</i> | 15 |
| <i>to-</i> | 8 |
| <i>for-</i> | 6 |
| <i>of-/be-/in-/ut-</i> | 9 |
| total pref. | 539 |
| % pref. | 35.5 |

ge- in particular, it has a rather normal frequency of prefixation overall.²⁰

²⁰ It should be noted here that while I have not systematically considered the other prefixes in my examination of the corpus results so far, my searches did identify them, precisely so that I could exclude them from the count of forms with no prefix at all. I was thus able to check the other lexical verbs identified here to make sure that none of them show similar effects to *sendan*, with prefixes other than *ge-* showing up at a high enough frequency to interfere.

The really big surprise among the verbs with a low frequency of *ge-* prefixation is with *cuman*. Verbs meaning ‘come’ are typically highly telic — unlike verbs meaning ‘go’, they include an inherent telos.²¹ Indeed, ‘come’ is typically a telic verb par excellence, and so we expect OE *cuman* to be used primarily in perfective and resultative contexts. Under essentially all accounts that have been proposed for the distribution of *ge-*, we would thus predict a very high frequency with *cuman*. And yet, this particular verb bears the prefix less than one percent of the time, i.e. with a frequency otherwise found only with statives. In (14) we have a straightforward example, which is clearly telic, perfective, resultative and anything else you might expect to be associated with *ge-*, including being a periphrastic perfect built on the PPP. Nonetheless, the form of *cuman* we find is unprefixated:

- (14) *Martha þa gehyrde þæt se Hælend wæs cumen*
 Martha then heard that the savior was come
 ‘Then Martha heard that the savior had arrived.’
 (coaelhom,+AHom_6:49.889)

We will return to the status of *cuman*, and how it might be accounted for, in Section 3.3.

Now let’s consider Table 12, which contains the verbs with markedly high frequency of *ge-*, i.e. those which take *ge-* more often than the average of the unclassified lexical verbs. Here again we have both the expected and the unex-

Table 12: Lexical verbs with high frequency of *ge-*

| Verb | Gloss | <i>ge-</i> | no | % <i>ge-</i> |
|----------------------|--------------|------------|--------|--------------|
| <i>tēon</i> | ‘draw, pull’ | 89 | 119 | 42.8 |
| <i>weorþan</i> | ‘become’ | 1001 | 979 | 50.6 |
| <i>niman</i> | ‘take’ | 1431 | 1265 | 53.1 |
| <i>halgian</i> | ‘hallow’ | 392 | 108 | 78.4 |
| <i>hǣlan</i> | ‘heal’ | 626 | 110 | 85.1 |
| <i>sēon</i> | ‘see’ | 2714 | 188 | 93.5 |
| general total | | 42366 | 124178 | 25.4 |

pected from the perspective of previous attempts to understand the prefix. These

²¹ By default this telos is the location of the speaker at the reference time, but it can be shifted to other salient locations depending on the context.

arre primarily highly telic achievement verbs, like *weorþan*, *niman*, *halgian* and *hāelan*, which would be expected under all approaches to appear frequently with *ge-*. The surprising case here is *sēon*, which might have been expected to be a stative or an activity at least a significant portion of the time, but in fact overwhelmingly takes *ge-*.

2.7 Interactions with auxiliaries

A final type of factor to consider is the presence of different auxiliary verbs in the clause in addition to the main verb. We can naïvely expect effects on the distribution of *ge-* here because at least some of the relevant periphrastic constructions are used to express aspectual distinctions. Additionally, the presence of particular auxiliaries is also extremely easy to search for in the corpora, unlike most other reflections of aspect. Let's begin then with a comparison of all of the auxiliaries, as well as the possibility of no auxiliary, shown in Table 13.²² Clauses with a (pre-)modal auxiliary have a somewhat higher than average fre-

Table 13: Presence of auxiliaries in the clause

| Aux. | <i>ge-</i> | no | % <i>ge-</i> |
|----------------------|-------------------|-----------|---------------------|
| none | 27853 | 113588 | 19.7 |
| (pre-)modal | 3375 | 7441 | 31.2 |
| BE | 9764 | 2494 | 79.7 |
| HAVE | 969 | 43 | 95.8 |
| general total | 42366 | 124178 | 25.4 |

quency of *ge-*, but the difference is not particularly large.²³ Auxiliaries BE and HAVE, on the other hand, show a very strong favoring effect on *ge-*. Clauses with no auxiliary have a somewhat lower than average frequency with *ge-*, but this is just because the examples with BE and HAVE push up the average so high.

²² The various categories here do not add up to the general total because a number of examples have been set aside where there is more than one auxiliary or the situation is otherwise complicated in a way that is not easy to compare to the main categories here.

²³ To be absolutely clear, what is being discussed here is whether, in a clause containing a particular auxiliary, the main verb is prefixed with *ge-*. We are not talking about instances where an auxiliary itself is prefixed with *ge-*. As noted in Section 2.6, this simply does not seem to occur.

Now, at first glance it looks like the preference for *ge-* is stronger with HAVE than with BE. However, there's a good bit more going on here that needs to be unpacked. First of all, while auxiliary *have* is essentially only found in the (ancestor of the) periphrastic perfect, *be* is also used in the passive and the OE ancestor of the progressive. Aspectually speaking, the passive and especially the progressive are entirely different from the perfect, and we do not expect them to behave at all the same with respect to *ge-*. If we restrict our attention to just perfect clauses, we get the numbers in Table 14.²⁴ Now we see that the

Table 14: Perfects, according to auxiliary

| | <i>ge-</i> | no | % <i>ge-</i> |
|-------------|------------|----|--------------|
| BE | 868 | 96 | 90 |
| HAVE | 125 | 4 | 97 |

frequency of *ge-* is much closer to being the same with BE and HAVE, and furthermore that it is approaching being categorical. Still, there appears to be a somewhat stronger preference for the prefix with HAVE than with BE.

It turns out, however, that this difference is spurious, and comes entirely from interaction with lexical effects. 86 of the 96 examples of perfects with BE, where the PPP lacks *ge-*, are with *cuman*. Recall that — for reasons that we haven't figured out yet — *cuman* staunchly resists prefixation with *ge-*. Crucially, OE *cuman* **only** appears with BE in the perfect (McFadden & Alexiadou 2010), so the examples with that one verb are artificially suppressing the overall frequency of *ge-* with auxiliary BE. If we remove the examples with *cuman* from consideration, we get the numbers in Table 15. The difference between HAVE and BE is now essentially gone, and we have the effect that once we correct for lexical oddities, *ge-* is essentially categorical with the periphrastic perfect in OE.

²⁴ The numbers of examples here are much smaller because I have restricted attention to intransitive examples for methodological reasons. The issue briefly is this. In the *YCOE* corpus, PPPs all have the same tag VBN, regardless of whether they appear in a passive, a perfect or some other construction. This means that perfects with auxiliary BE and passives are formally identical, hence not distinguishable by corpus searches. They must rather be identified on a case by case basis by considering the transitivity of the lexical verb and the semantics of the particular example. Fortunately, this time-consuming hand-coding has already been done, in the research leading up to McFadden & Alexiadou (2010), and I have used that as the basis for the numbers reported here. However, since that work was concerned with the alternation between HAVE and BE in the perfect, and that alternation is restricted to intransitives (transitives always using HAVE), the coding was only done for intransitive examples, yielding the restricted sample reported on here.

Table 15: Perfects, according to auxiliary, excluding *cuman*

| | <i>ge-</i> | no | % <i>ge-</i> |
|-------------|------------|----|--------------|
| BE | 861 | 10 | 99 |
| HAVE | 125 | 4 | 97 |

Consider what this means for our search for an explanation for the function of *ge-* in the language. Since its distribution is so categorical once we carefully distinguish contexts, it is plausible to think that it realizes a single, specific grammatical category, rather than marking a vaguer conceptual category that is variably sensitive to multiple factors. Furthermore, this specific category cannot be something that distinguishes the HAVE and BE perfects in OE, since they behave identically. It must rather be a component that all OE perfects have in common, which however is not limited to or diagnostic of the perfect, as it is present at a relatively high frequency in non-perfect clauses as well.

3 An analysis of the OE patterns and some explanation

3.1 The semantics of early English ‘perfects’

I submit that the facts just discussed from the perfect are the key to understanding the role of *ge-* in OE. In particular, they lead quite directly to the proposal that the prefix does indeed have something to do with resultativity, not perfectivity. In order to motivate this we need to first consider some background on the semantics of the perfect. Part of what makes the perfect difficult to get a handle on is that there seem to be multiple readings for it, which are distinguishable in terms of their entailments, yet can be expressed by the same morphosyntactic form, at least in many languages (see Iatridou et al. 2003 and the other contributions in Alexiadou et al. 2003 for useful discussion). For present purposes we need to be able to distinguish between the ‘perfect of result’ and the ‘experiential perfect’, both of which can be expressed by the periphrasis with auxiliary *have* in PDE.

The ‘perfect of result’ entails that the target state of the eventuality described by the main predicate holds at the reference time.²⁵ In (15), e.g. the continuation makes it clear that we’re not just talking about what Beorhtric has done, but what state he is currently in as a result of what he has done. I.e. he is

²⁵ See Parsons (1990); Kratzer (2000) for the difference between ‘target’ states and ‘resultant’ states.

in York at the reference time (which happens to also be the speech time because this is a present perfect) as a result of going there.

(15) Beorhtric has gone to York, and he won't be back until tomorrow.

(16) Beorhtric has gone to York five times already this semester.

The experiential perfect on the other hand entails that the eventuality described by the main predicate is anterior to the reference time. There are not necessarily any implications about whether or not particular consequences of that eventuality continue to hold, aside from the somewhat trivial fact that what has happened cannot un-happen, i.e. the subject will always have the experience of having participated in the eventuality, even if the particular target state of that eventuality no longer holds. Example (16) must be interpreted as an experiential perfect because of the repetition involved. It is not possible for the target state of at least the first four instances of Beorhtric going to York to still hold, since he must have left in the meantime in order to be able to go back. That is, there is clearly no implication that Beorhtric is in York five times at the reference time, which would be incoherent, but rather that he is in the state of having experienced going to York five times in the last semester.

Now, it is well established that the 'perfect' constructions in OE, built with BE and HAVE in addition to the PPP, crucially differ from their PDE descendants in that they were essentially restricted to the perfect of result (see McFadden & Alexiadou 2010 and citations there). The experiential reading was not yet available. In other words, while you could use a periphrastic 'perfect' to express something like (15) in OE, you could not use it to express (16). A simple past form would have been used instead. This means that every time we see a perfect in an OE text, we can conclude that it is a resultative, i.e. there is a target state that is asserted to hold at the reference time. This target state is then something that the perfects with HAVE and BE in OE all have in common. Thus it is a candidate for what *ge-* contributes. Indeed, we can go further, since it is certainly not the case that target states are only found in perfects. Rather, target states can be found in clauses with all kinds of tense and aspects, which is at least broadly in line with the distribution we observe for *ge-*.

3.2 The proposal and its implementation

I would thus like to explore the hypothesis that OE *ge-* marks a particular type of resultativity, being associated with the presence of a target state. This of course has clear connections to some of the earlier proposals mentioned above. van Kemenade & Los (2003) argue that *ge-* is related to resultativity in various

stages of Dutch and English, and resultativity plays an important role in analyses of verbal particles in some of the modern languages (see e.g. Ramchand & Svenonius 2002; McIntyre 2003). It is clearly distinct, on the other hand, from proposals that connect *ge-* to (outer) aspect, in particular Streitberg (1891)'s claim that it marks perfectivity.²⁶

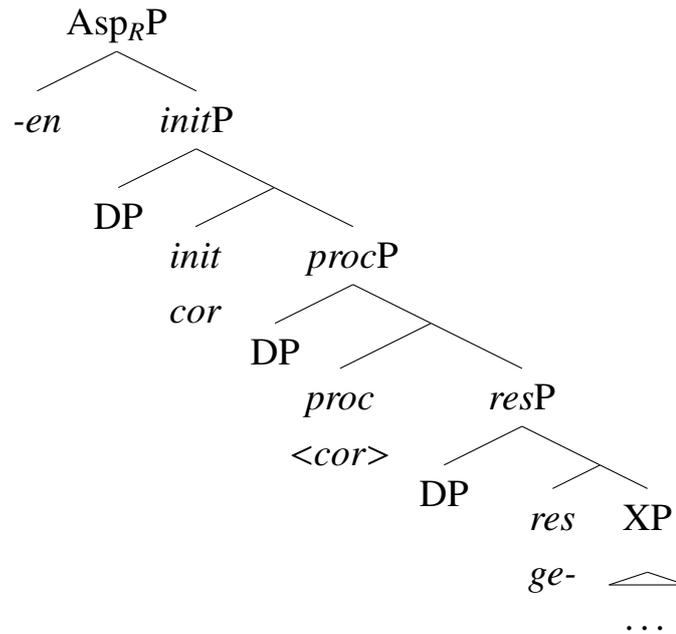
The idea being pursued here thus associates *ge-* with inner aspect or Aktionsart, rather than with outer or viewpoint aspect. Note, however, that I am **not** tying *ge-* to a specific (Vendlerian) Aktionsart, but rather to one of the building blocks that goes into at least two different ones. Modern work on Aktionsart, even if it adopts Vendler's four main categories, typically decomposes them in terms of more basic components (see Dowty 1979 among many others), and the target state is one of these basic components. It is a defining ingredient of the telic Aktionsarten, i.e. of Achievements and Accomplishments, but not of Activities or States (or Semelfactives).

What I would like to argue for now is that the OE data discussed so far actually allows us to be even more specific about what is meant by saying that *ge-* is associated with resultativity and target states. I will implement my proposal in terms of Ramchand (2008)'s verbal decomposition, which identifies the basic building blocks of the Aktionsart structure of predicates with dedicated heads in the syntax. She posits three main heads in what she dubs the 'first phase'. The middle head, called *proc(ess)* encodes a (durative) process. The higher head, called *init(iation)* encodes a state which causes the process, i.e. the state of affairs that sets the process in motion, like the intention of an agent. The lower head, called *res(ult)* encodes a state that is caused by the process, i.e. the target state of the complex eventuality. The Aktionsart of particular eventualities can differ based on which of these heads are present. For example, a typical activity will include *proc* but not *res*, and may or may not include *init*, depending on whether it is externally caused. An agentive accomplishment, on the other hand, will include all three heads, with *init* encoding the agentive causation, *proc* the process portion, leading up to the target state represented by *res*.

Given this background, I would like to propose that *ge-* is the unmarked realization in OE of Ramchand's *res* head. The *-en* or *-d* suffix in PPPs is higher up, in an Asp head outside of *initP* (Kratzer 2000; Embick 2004; McFadden & Alexiadou 2010). Concretely, for a form like the PPP *gecoren* 'chosen' in a perfect clause like example (11) above, we have the structure in (17):

²⁶ There are, however, connections to the more nuanced aspectual proposal of Lloyd (1979) for Gothic *ga-*, which takes into account issues of both inner and outer aspect. A more complete comparison with Lloyd's proposals is planned as part of work in progress.

(17)



In addition to the basic semantics (about which more below), this structural proposal accounts for certain simple morphological facts, e.g. that *ge-* is a prefix while *-en* is a suffix. Given the Mirror Principle (Baker 1985), if the three heads in the structure above combine in a single word, the expected unmarked order will be *res-proc-init-Asp_R*, which is exactly what we get. Note that it is cross-linguistically common for resultative elements to show up as verbal prefixes, and so this seems like a quite reasonable result.

3.3 Covering the data

Let us consider now how this proposal can accommodate some of the surprising data that we turned up for specific lexical verbs. First, how do we deal with the extreme dispreference for *ge-* with *cuman*? Note that the pattern here is out of line with expectations no matter what approach we take to the basic semantics of *ge-*. As noted above, ‘come’ is clearly a telic predicate, with a strong resultative component, and is certainly expected to appear frequently in perfective uses. The idea being pursued here, that *ge-* is associated with target states realized as Ramchand’s *res* head, fares no better on its own, since the semantics of ‘come’ clearly does include a target state. What this means is that we need a theory that can treat *cuman* as some kind of lexical exception, and ideally also make sense of why this of all verbs should be exceptional in this way. While the proposal being made here cannot claim to be uniquely suited in this sense — any reasonable theory will have a mechanism for dealing with lexical exceptions of this kind — the solution it offers is at least more than adequate.

The relevant bit of background is that Ramchand's system allows for single verbal elements to simultaneously realize multiple head positions, subject to lexical restrictions. That is, some lexical verbs can be specified to realize *init* + *proc*, others *proc* + *res* or *init* + *proc* + *res*, and still others just *proc*. Indeed, in a language like PDE, the vast majority of lexical verbs can realize the entire spine of the first phase, including two or three of the relevant functional heads, without any help from prefixes (or suffixes). What I am proposing for OE in contrast is that, as a language-specific property, its lexical verbs are generally not specified to realize the *res* head, which thus must be spelled out separately from the verb root in structures in which it appears, and furthermore that this precisely this is what the prefix *ge-* does. Now, to deal with a verb like *cuman*, we can simply posit that it is exceptional in that it can realize the *res* head in addition to *init* and *proc*. I.e. its lexical entry contains the specification [*init*, *proc*, *res*]. Under the principles Ramchand adopts for how lexical items compete to realize particular bits of structure, an element like *cuman* will span across all three heads, winning out over and thus blocking *ge-*. This of course works technically and covers the empirical ground we need it to cover, but it can be reasonably argued that it goes beyond this minimum to be a fairly natural account of the situation and to actually help make sense of why this particular verb should behave this way.

First of all, in Ramchand's system, what determines what spells out the different heads in the first phase is the lexical information specified for particular verbal items, interacting with general principles for resolving competition. If *ge-* is competing with lexical material to spell out a particular head in that system, it is entirely expected that there will be some lexical exceptions where it gets beaten out, as we find with *cuman*. Second of all, the specification of what heads in the first phase a lexical item can realize is essentially a grammaticalization of its typical semantic behavior. Thus if there are going to be lexical items that are exceptionally specified to be able to realize the *res* head, we would expect them to be precisely those lexical items that most frequently appear with such semantics. In other words, we expect a verb like *cuman*, which is basically always used in resultative contexts, to be able to supercede *ge-*, i.e. to be inherently resultative, not a verb like, say, *faran* 'go', which is sometimes used in resultative contexts and sometimes in non-resultative ones.

What then about the markedly low frequency of *ge-* that we noted with *etan* and *drincan*? These are verbs that may not always involve a clear target state, but certainly will much of the time, when a clearly defined substance ends up being consumed. Consider an example where both of them happen to appear, and where there are clear target states defined for the definite objects they take:

- (18) þā sacerdas þā ... **ǣton** þone mete ... and þæt wīn eal **druncon**
the priests then ... ate the food ... and the wine all drank
‘The priests then ate the food and drank all the wine.’
(coaelhom,+AHom_22:406.3519-21)

Given everything we have said so far, we certainly expect the result states of the food and all the wine being consumed to be reflected by the presence of *ge-* on the two verbs. Furthermore, given the fact that these verbs are not expected to **always** be telic, with a clear target state, the approach we adopted for *cuman* does not seem appropriate.

As it turns out, however, Ramchand (2008)’s theory actually predicts this pattern. She makes a distinction between resultative meanings that come from the structural specification of an actual target state, and those that arise from the presence of a bounded path or theme argument. While the former involve an explicit *res* head in the structure, the latter do not, with the resultative meaning instead being an entailment of how the rhematic material restricts the interpretation of *proc*. That is, they do not actually involve a *res* head. Note then that clauses built around *etan* and *drincan*, as consumption verbs, will primarily be found in structures of this latter type. They realize *init* and *proc*, and combine with ‘incremental themes’ like *þone mete* and *þæt wīn* above, which bound the process, providing an implication of telicity. That is, it is possible at any time to gauge the priests’ progress in completing the eventuality described by examining how much of the food and wine are left. However, no *res* head will be involved in these structures, and thus there is no place for *ge-* to be inserted. We can thus account for why *ge-* is generally not found with these verbs, even though they have a ‘resultative’ interpretation in a pre-theoretical sense. This is a clear advantage over competing proposals in terms of telicity or a less precisely defined resultativity.

4 ME developments

An additional point that makes the analysis of *ge-* in OE just presented particularly attractive is that it offers insights into its development in ME. In this section I report on results from searches on the complete PPCME2 (Kroch & Taylor 1999), which is analogous to the YCOE discussed above, also containing 1.5 million words, tagged for part of speech and grammatical features and parsed for syntactic structure, but covering ME rather than OE. Not surprisingly, the use of the prefix shows a steady decline over the course of late OE and ME, as we can see if we extend Table 7 from Section 2.5 above into ME, as shown in Table 16. Again, the frequency of *ge-* is impressively stable in the

Table 16: Frequency of *ge-* in ME

| Period | Dates | <i>ge-</i> | no-pref | total | % <i>ge-</i> |
|------------|------------------|------------|---------|-------|--------------|
| OE1 | pre-950 | 15079 | 43464 | 58543 | 25.8 |
| OE2 | 950-1050 | 19695 | 57793 | 77488 | 25.4 |
| OE3 | 1050-1150 | 74 | 422 | 496 | 14.9 |
| ME1 | 1150-1250 | 2297 | 30190 | 32487 | 7.1 |
| ME2 | 1250-1350 | 989 | 16850 | 17839 | 5.5 |
| ME3 | 1350-1420 | 1106 | 58519 | 59625 | 1.9 |
| ME4 | 1420-1500 | 162 | 31614 | 31776 | 0.5 |

first two periods of OE, but then begins to drop in period O3. While we noted above that we have to be careful about taking the numbers from this period too seriously given the relative paucity of examples, we can see now that they do fit in well with subsequent developments in ME. The overall frequency of *ge-* is significantly lower than in OE from the beginning of the ME period, and by the end of the ME period, the prefix has essentially been lost, aside from a small number of relics.

It is far from clear what might underlie this decline, regardless of what theory we might adopt for the function of *ge-* in OE. We certainly cannot expect speakers to have had less need of resultativity or of perfectivity or any of the other proposed categories, i.e. to have stopped using *ge-* because they stopped talking about target states. One possibility, given the proposal made for OE *cuman* above, is that lexical verbs increasingly came to be able to realize the *res* head themselves, obviating the need for a separate realization by *ge-*. At some level this must ultimately be the case, since in Ramchand's system, verb roots that can appear in Accomplishments and Achievements in PDE must be analyzed as covering the *res* head. But this just pushes the question one step further down the line: what led (certain) verb roots to expand their realization in this way, at the expense of the inherited prefix *ge-*?

Perhaps the most likely explanation is a relatively mundane morphophonological one, with developments akin to what happened in the prehistory of Old Norse. The prefix was unstressed and had relatively little phonological substance to begin with, starting as /jɛ/ in early OE. In late OE it would have been reduced to /jə/ by regular sound changes affecting unstressed vowels, and by ME, where the usual spelling has become *i-* or *y-*, it was simply /i/ or perhaps even /ə/. While it probably should not have completely disappeared due solely to sound change, it would have been reduced enough to plausibly be suscepti-

ble to morphologically conditioned loss. Note that two of the other prefixes that have survived into the modern language — *be-* and *for-* — had more phonological substance, beginning with an obstruent, although even they have seen their distribution heavily reduced.

Even if we can't be sure about why *ge-* disappeared, the way in which it did so does yield some insights into what it was doing. Its decline in frequency is not uniform across environments, but proceeds rather differently in the two places where *ge-* is most common — perfects and passives, as shown in Table 17.

Table 17: *ge-* in perfects and passives in ME

| Period | Perfect | | | | Passive | | | |
|------------|------------|---------|-------|--------------|------------|---------|-------|--------------|
| | <i>ge-</i> | no-pref | total | % <i>ge-</i> | <i>ge-</i> | no-pref | total | % <i>ge-</i> |
| ME1 | 437 | 424 | 861 | 50.75 | 967 | 1222 | 2189 | 44.18 |
| ME2 | 217 | 265 | 482 | 45.02 | 352 | 1096 | 1448 | 24.31 |
| ME3 | 213 | 1891 | 2104 | 10.12 | 691 | 4730 | 5421 | 12.75 |
| ME4 | 10 | 1247 | 1257 | 0.80 | 85 | 3136 | 3221 | 2.64 |

Note that in the M1 period at the beginning of ME, the frequency of *ge-* is comparable in the two environments. But while the drop in the passive is fairly smooth over the next three periods, in the perfect the frequency remains stable into ME2, before dropping suddenly in ME3.

We can actually make sense of this development if we consider the resultative analysis of *ge-* being proposed here in light of McFadden & Alexiadou (2010)'s findings on the development of the perfect in ME. We showed that in OE and early ME, the periphrastic perfect was only used with a perfect-of-result reading (as discussed above), and thus could only be built on resultative predicates. Starting in the ME3 period, however, the new experiential — crucially non-resultative — use of the perfect with *have* (but not with *be*) arose. Again, this accounts for why *ge-* was so common in the perfect in the early periods, if as proposed here it was the default realization of the underlying resultative structure. More importantly for current purposes, it also predicts the sudden drop in the frequency of *ge-* in perfects, precisely in ME3, due to the influx of the new experiential perfect. This placed no resultativity requirement on the predicates it was built on, thus did not favor *ge-* the way the old resultative perfect had. Indeed, as the table above shows, the marked decrease in the percentage of perfects with *ge-* in that period results not from a decrease in instances of *ge-*, but from a sudden increase in the total number of perfects, as expected. The

old perfect-of-result, which favored *ge-*, continued to be used at similar rates as before, but was swamped by the new experiential one, which did not favor *ge-*. This is a striking parallel to McFadden & Alexiadou (2010)'s finding that the purely resultative *be*-perfect was swamped by the resultative-or-experiential *have*-perfect in the same period.

5 Summary and outlook

In this paper I have presented data and conclusions from the first stage of a large-scale corpus study on the use of preverbal *ge-* in Old and Middle English. On the basis of the evidence obtained so far, I was able to propose an analysis of the prefix which is more explicit about the specific aspectual components involved than previous approaches, and which can cover certain otherwise puzzling factors of its distribution. The preliminary nature of this report is clear from a number of its limitations. Most importantly, the corpus study has so far only involved automated searches, which means that only those patterns have been investigated that can be unambiguously identified on the basis of the parsed structure or annotation in the corpora or on the basis of specific string forms. As a result, a wide array of syntactic and especially interpretive factors have not yet been taken into account. The temporary justification for this is that the volume of data involved is simply too large to examine all of the examples by hand. Additionally, only the most basic level of statistical analysis has been carried out, in particular a series of chi-square tests on the effects of individual factors. No attempt has been made thus far to carry out a proper multivariate analysis to disentangle the effects of the various factors that have been identified. For some factors with very clear effects this is probably not a serious problem, but for others, where non-trivial interactions are clearly involved, we are certainly missing an important part of what is going on.

The continuation of this project will address both of these issues. First, a representative sample of manageable size will be extracted from the collection of data made so far, and the examples there examined and coded by hand for factors that could not be searched for. Second, a more complete statistical analysis will be carried out. One portion of this will be to carry out the multivariate analysis that is so sorely missing at this stage. A second will involve applying more sophisticated tools to overcome the recurring issue described in this paper, where a simple chi-square test returned (a sometimes quite high level of) statistical significance even with a very small effect size, simply because of the huge numbers of examples involved.

References

- Alexiadou, Artemis, Monika Rathert & Arnim von Stechow (eds.) (2003). *Perfect explorations*. Berlin: Mouton de Gruyter.
- Baker, Mark (1985). The Mirror Principle and Morphosyntactic Explanation. *Linguistic Inquiry* 16, 373–415.
- Braune, Wilhelm & Ingo Reiffenstein (2004). *Althochdeutsche Grammatik*. Tübingen: Max Niemeyer Verlag, 15th edition.
- Dowty, David (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Embick, David (2004). On the structure of resultative participles in English. *Linguistic Inquiry* 35, 355–392.
- Eythórsson, Thórhallur (1995). *Verbal Syntax in the Early Germanic Languages*. Ph.D. thesis, Cornell University.
- Iatridou, Sabine, Elena Anagnostopoulou & Roumyana Pancheva (2003). Observations about the form and meaning of the perfect. In: Artemis Alexiadou, Monika Rathert & Arnim von Stechow (eds.). *Perfect explorations*. Berlin: Mouton de Gruyter, 153–204.
- van Kemenade, Ans & Bettelou Los (2003). Particles and prefixes in Dutch and English. In: Geert Booij & Jaap van Marle (eds.). *Yearbook of Morphology*. Dordrecht: Kluwer, 79–118.
- Kratzer, Angelika (2000). Building Statives. In: *Proceedings of the 26th annual meeting of the Berkeley Linguistics Society*, 385–99.
- Kroch, Anthony & Ann Taylor (1999). Penn-Helsinki Parsed Corpus of Middle English, 2nd ed. Univ. of Pennsylvania.
- Lindemann, J.W. Richard (1970). *Old English preverbal Ge-: Its meaning*. Charlottesville: The University Press of Virginia.
- Lloyd, Albert (1979). *Anatomy of the verb: The Gothic verb as a model for a unified theory of aspect, actional types, and verbal velocity*. Amsterdam: John Benjamins.
- Los, Bettelou (2007). *The rise of the to-infinitive*. Oxford: Oxford University Press.
- McFadden, Thomas & Artemis Alexiadou (2010). Perfects, Resultatives, and Auxiliaries in Earlier English. *Linguistic Inquiry* 41 (3), 389–425.
- McIntyre, Andrew (2003). Preverbs, argument linking and verb semantics: Germanic prefixes and particles. In: Geert Booij & Jaap van Marle (eds.). *Yearbook of Morphology*. Dordrecht: Kluwer, 119–144.
- Parsons, Terence (1990). *Events in the Semantics of English: a Study in Subatomic Semantics*. Cambridge, Mass.: MIT Press.

- Postma, Gertjan (2002). Negative polarity and modality in Middle Dutch *ghe*-particle constructions. In: Sjef Barbiers, Frits Beukema & Wim van der Wurff (eds.). *Modality and its Interaction with the Verbal System*. Amsterdam: John Benjamins, 205–244.
- Ramchand, Gillian (2008). *Verb meaning and the lexicon: A first phase syntax*. Cambridge: Cambridge University Press.
- Ramchand, Gillian & Peter Svenonius (2002). The Lexical Syntax and Lexical Semantics of the Verb-Particle Construction. In: L. Mikkelsen & C. Potts (eds.). *WCCFL 21 Proceedings*, 101–114.
- Streitberg, Wilhelm (1891). Perfective und imperfective Actionsart im Germanischen. *Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB)* 15, 70–177.
- Taylor, Ann, Anthony Warner, Susan Pintzuk & Frank Beths (2003). York-Toronto-Helsinki Parsed Corpus of Old English Prose. University of York.