MEDICINE, ASTRONOMY, AFFIXES AND OTHERS: AN ACCOUNT OF VERB FORMATION IN SOME EARLY SCIENTIFIC WORKS

Abstract

The patterns of derivational morphology found in Middle English in general should, theoretically, coincide with those found in an emerging functional variety of English from the same period (scientific writing). This scientific register has only been studied as such in the last decade or so, and often receiving a brief mention in more general works. At the same time, derivational morphology has not been considered as a tool to measure the degree of vernacularisation of late Middle English scientific texts.

In this paper we intend to examine the behaviour of verbs in late Middle English scientific texts from the point of view of derivation and the degree of vernacularisation observed in them (previous work has focused on the analysis of nouns, arguing that they are the most frequent word-category with semantic content in scientific writing). The particular aspect of vernacularisation that concerns us here is the ability of writers to reanalyse and re-use elements of the language. To ascertain whether all fields of knowledge resort to the same linguistic devices, whe have used Chaucer's *The Equatorie of the Planetis* and *Astrolabe* and compared it with a *Remedy Book*.

Keywords: derivational morphology, Middle English, scientific register, vernacularisation.

Resumen

Los modelos derivativos encontrados en Inglés Medio deberían, en teoría, coincidir con los hallados en una incipiente variedad funcional del mismo periodo (el inglés científico). Este registro científico ha sido objeto de estudio en la última década, a menudo, recibiendo una breve mención en trabajos más amplios. Al mismo tiempo, la morfología derivativa no ha sido considerada una herramienta para medir el grado de vernacularización de los textos científicos en inglés medio tardío.

En este artículo pretendemos examinar el comportamiento de los verbos en textos científicos de inglés medio tardío desde el punto de vista de la derivación así como el grado de vernacularización observado en ellos (trabajos anteriores se han centrado en el análisis de los nombres aduciendo que son la categoría léxica más frecuente en inglés científico). El aspecto concreto de la vernacularización que nos interesa es la capacidad de los autores para reanalizar y reutilizar elementos de la lengua. Para comprobar si todos los campos del saber recurren a los mismos procesos lingüísticos, hemos usado *The Equatorie of the Planetis y Astrolabe* de Chaucer y lo hemos comparado con un *Remedy Book*.

Palabras clave: morfología, derivación, Inglés Medio, registro científico, vernacularización.

The patterns of derivational morphology found in Middle English should, theoretically, coincide with the patterns found in an emerging functional variety of English from the same period. Marchand (1969), Matthews (1974), Fernández (1982), Burnley (1992), Kastovsky (1992) and Dalton-Puffer (1996),

as well as a number of classical authors, have done extensive work on the behaviour of the English morphological subsystem at different stages of its evolution. A complementary analysis of the scientific register has only been broached in the last decade or so, however, sometimes receiving only a brief mention in more general works (Barber 1993; Beal 2004). Meanwhile, derivational morphology, mainly affixation, has not yet been considered as a tool to measure the degree of vernacularisation of so-called "late Middle English scientific texts".

Previous work on morphology and semantics has focused on the analysis of nouns, arguing that they are the most frequent word-category with semantic content in late-medieval scientific writing, which in turn relates to the primary function of scientific texts to transmit ideas -concepts, objects, etc. (Sager & *al.* 1980).

In this paper, we intend to examine the behaviour of verbs in late Middle English scientific texts from the point of view of derivation and the degree of vernacularisation observed in them. The particular aspect of vernacularisation that concerns us here is the ability of speakers/writers to reanalyse and re-use elements of the language. The paper is divided into four sections. The first contains some preliminary observations about morphological analysis, Middle English and the emerging scientific register. In section 2 we present the corpus of data, and in section 3 our analysis. The paper concludes with some final remarks in section 4.

1.- MIDDLE ENGLISH MORPHOLOGY
    AND THE EMERGING SCIENTIFIC REGISTER

The productive combination of native bases and affixes in OE comes close to disappearing in the following period. Although a detailed account of Middle English derivational morphology shows the survival of a few residual OE prefixes and suffixes, it also demonstrates the influence of French, Latin and Greek in bringing about one of the biggest changes in the morphological and

lexical stock of the vernacular. Legal texts, documents, wills, charts, and the like, written in French and Latin, introduced a new, more non-Germanic formal aspect to the English lexicon. Language contact, both through oral and written media, favoured the introduction of new vocabulary items (either simple or complex) that, once assimilated by the discourse community, were used in combination with more familiar native bases (*unknowable*). The reverse also occurred: foreign bases were combined with native affixes (*moisten*). As Adams (2001: 11) has recently claimed: "no account of English word formation can ignore the mixed nature of the English vocabulary and the circumstances in which this situation came about". Adams's words remind us of the importance of taking into account the external history of the language to achieve a better understanding of how word-formation processes developed.

It is towards the end of the 14th century when the vernacular began to displace French and Latin from their usual contexts of use, with new functional varieties of English or registers appearing. We, like Voigts (1989) and Taavitsainen (2000), therefore believe that the vernacularisation of the English scientific register dates to the last quarter of the 14th century. This is why we have selected samples from the turn of the century: to study the behaviour of word-formation (derivation) in a particular functional variety of English at its earliest manifestation, and measure the level of vernacularisation of these first texts using derivation.

For a complete examination of any linguistic aspect of a register or special language (to use Sager *& alii*'s terminology), we must consider what other elements should be taken into account apart from the intra-systemic analysis itself. Here are some pragmatic considerations proposed by two different authors that may be useful:

a) According to Halliday's (1988: 140-141) Systemic Functional Grammar, diatypic variation can be examined in terms of Field (discipline or subject-field), Tenor (intended audience) and Mode (oral or written). He defines

register as "a cluster of associated features having greater-than-random [...] tendency to co-occur".

b) Biber (1995: 1) proposes a definition of register as "any variety associated with particular situational contexts or purposes".

The difference between these two authors lies in the fact that for Halliday the scientific register is the result of a realignment of lexico-grammatical elements, whereas Biber focuses on the use of technical vocabulary, among other (morpho-)syntactic constructions, as a register marker.

We will analyse both the intra-systemic and extra-systemic factors in question in our survey of derivational morphology in early scientific writing.

2.- METHODOLOGICAL CONCERNS:
PROCEDURE AND CORPUS MATERIAL

Three samples from two different disciplines have been used in our survey: a *Remedy Book* containing a compilation of medical recipes for medicine, and Chaucer's *A Treatise on the Astrolabe* and *The Equatorie of the Planetis* for astronomy. To make the total number of words for each discipline more or less equal, both the *Remedy Book* (20,788 words) and the *Astrolabe* have been taken *in toto*, but, for the *Equatorie,* only the sample included in the *Helsinki Corpus* has been considered. This gives a total of 21,544 words for the astronomy texts. Graph 1 below shows the total number of words (42,332) per sample and discipline.

Sager, Dungworth and MacDonald (1980) state that sublanguages or special languages (what we term registers) contain primarily nouns and, secondarily, verbs. Although their findings refer to present-day English, it is interesting to observe that this is also the case for pre-modern times, at the very birth of the scientific functional variety.

3.- ANALYSIS OF DATA

Three aspects will be considered in our analysis of data:

> 1.- Type/token ratio of the verbal forms recorded to assess the presence of verbs as opposed to nouns.
>
> 2.- Etymology of verbs, in general and in relation to each discipline.
>
> 3.- Derivational morphology as an indicator of vernacularisation.

3.1.- TYPE/TOKEN RATIO

The samples in our survey comprise 42,332 words: 5,126 of them represent verbal forms (12.10 %), in contrast with the number of nouns (21.85%) found in previous works where the same text samples have been used (Moskowich & Crespo 2006). These figures, referring only to the number of tokens found, are probably a reflection of the fact that scientific writing is less concerned with the description of actions and processes than of artefacts or concepts. Our figures confirm the claim by Sager *& al.* mentioned earlier, that the lexical category verbs comes second in the ranking of word–classes used in this type of writing. This claim is also in line with the figures we have found for each discipline: medicine contains 2,979 tokens, whereas astronomy contains 2,147. The type of texts selected for the present study helps us to explain this circumstance. In Medicine, the compilation of recipes belongs to the *Remedy Book* tradition, representing the lower layer of text-types alluded to by Taavitsainen (2004). It is a practical guide in which the processes of selection of ingredients are described and directions for the preparation of different formulae (actions) are given to an audience with no academic training. By contrast, the two extracts of astronomy texts are more descriptive of ideas and objects and are addressed to a more educated audience[1].

The type/token analysis has revealed that there are 421 types and 5,126 tokens in our survey. To illustrate the most numerous types, we have listed

---

[1] We believe that Chaucer's address to his son Lewis is just a literary device to catch the readers' attention.

those with more than fifty tokens. Table 1 shows that all these types are of Germanic (G, henceforth) provenance.

At the other end of the scale we find instances of non-Germanic (NG, henceforth) types with one single token: amenusen, amounten, aperceyven, causen, moisten, rollen, condescenden, costen, counsellen, desturben, dicen, enhauncen, erren, expansen, destroyen, intercepten, obseruen, sustenen, among others. Within NG forms we can distinguish both derivative and non-derivative types, although derivation is more abundant. This is not the case with Germanic types, where all of them are simple. G forms, however, are more amenable to combination with other parts of speech to create multiword verbs than NG forms. This is the case with taken (taken up, out, doun, awei, of, ouertaken), setten (down, ouer, upon) and lien (on, up, upon, doun, ouer). As Burnley (1992: 445) has put it:

> This emphasis upon the particled verb as the focus of derivation is symptomatic of the change which took place during the fifteenth century by which the formation of verbs became concentrated on the production of particled verbs, and compound verbs ceased to be productive as a type of word formation.

The reason for this may be that these verbs belong to the core vocabulary of the language. They are so internalised by the speaker that she/he is capable of reanalysing and combining them with different linguistic units to create new meanings and thus enrich the lexicon of the vernacular using native resources. We may therefore regard the addition of particles to verbal bases as a productive device of word-formation.

3.2 ETYMOLOGICAL ANALYSIS

The etymological analysis of verbs will help us illustrate a number of pragmatic observations mentioned earlier. Although we have consulted the *Middle English Dictionary*, the etymological classification of verbs has been

simplified by adopting a broad-based division of all occurrences into those of Germanic (G) and non-Germanic (NG) origin. There are, obviously, cases of unknown provenance (UNK), and these have been recorded as well. A combination of origins has been proposed for derivative cases where a base and an affix are mixed. The origins of both elements may coincide (G+G; NG+NG) or not (NG+G; G+NG). Switches into Latin have been also included in our database as CS (Code-switching).

Table 2 below exhibits the general figures calculated according to these parameters.

Our 5,126 verbal forms are distributed in such a way that Germanic origin corresponds to 82.73% of all forms (4,241 tokens, either G only or G+G) and Non-Germanic origin to 16.32% (837 tokens, both NG or NG+NG). This testifies to the advance of vernacularisation in early English scientific writing, at least in relation to the levels of technicality observed in our texts. However, we should also point out that, with respect to the number of types of G and NG provenance, the proportion of NG types is higher. This shows that recourse to foreign terms ultimately proved necessary: lexical gaps in the transmission of scientific concepts had to be filled, but the terms used for the purpose were not yet sufficiently assimilated by the discourse community of the recipient language to be able to reanalyse them and use their components. The creation of new words out of non-native word stock is a common procedure in present-day English scientific language, but this was not the case in medieval times when borrowing was the preferred method (Halliday 1978). Derivative occurrences are more common in forms of a NG extraction, where they often take the form of cultural borrowings[2].

The unbalanced distribution of etymological origins can be appreciated more clearly in the two disciplines of our study when viewed separately.

---

[2] According to Scotton's (1993) classification, cultural borrowings represent items introduced into the recipient language to designate new concepts, objects or processes.

Tables 3 and 4 below display the number of tokens recorded in each case. We observe a predominance of G forms in both medicine and astronomy samples, though the difference between G and NG instances is smaller in astronomy texts. Once again, this may be explained by the fact that the samples in our corpus are targeted at different kinds of addressee. An additional factor in this etymological imbalance is the underlying tradition in the *Remedy Book*, the compilation of medical recipes, which is missing in astronomy. The latter can be regarded as a newcomer in the vernacular.

3.3 Word-formation

Affixation is the only word-formation process we have considered in this analysis. This means that all examples have been classified into simple forms (Ø process), S (for suffixation), P (for prefixation), and S+P, SS or PP for forms combining both types of affixes, or "multiple affixation" in Carstairs-McCarthy's terms (2002). Finally, "other", encompasses both compounding and multiword verbs.

As Halliday (1978: 195) states "Creating new words out of native word stock [...] has not played a very great part in the creation of technical registers in English". In fact, a general examination of word-formation processes in our samples produced the results shown on table 5.

The difference between simple and non-simple forms (91.74% vs 8.12%) is immense. At the level of non-simple forms the difference between derivatives (4.4%) and "other" (3.82%) is much less significant. Notwithstanding, a slightly different picture (see Graph 2) emerges from our discipline-based counts.

To begin with, a lower percentage of simple forms has been found in astronomy along with a greater tendency to use derivative forms, although these "derivative" forms were taken in by language users as whole units, not yet identified as separable entities. The higher percentage of "other" in the medical sample can be explained principally, though not exclusively, on the

grounds of the Scandinavian influence on Middle English. All combinations of lexical verb + grammatical word (either preposition or adverb) of the type described in section 3.2 can be included here.

We now proceed to combine both variables, origin + word-formation process, in each discipline, in the hope of obtaining a new perspective on the vernacularisation of early scientific writings (Tables 6 and 7 below).

Data from the *Equatorie* and the *Astrolabe* manifest a clear tendency to employ simple occurrences of G origin (1,574), in sharp contrast to the 271 instances from NG strata. All switches occur in the *Astrolabe* and correspond to the Latin type *nota,* a carry-over from the underlying classical tradition of the instructional book[3]. The 205 derivative tokens are divided into those which combine Germanic elements (G+G=80) and those which contain NG bases and affixes (NG+NG=125). The interaction of native and non-native elements is observed in "other" in the 7 instances of NG+G corresponding to 7 different types: *ben descriued vp, ben deuided owt, tornen abowte, passen thorow, turnen vp, kowchen adown* and *passen owt*. A special case of multiple affixation or derivation was found in just one form: *condescenden*. The technicality of the samples is not high enough to endow the texts with a higher degree of lexical complexity.

As for the *Remedy Book,* because it exhibits fewer cases of affixation (21), it contains more simple verbal types of both G (2,405) and NG (406) origin, as well as a greater number of tokens grouped under the heading "other" (139), than the Astronomy samples (see Table 7).

Affixation is illustrated by the combination of bases of the same origin: G+G (*ouerhelen, ouerkarven*), NG+NG (*refresschen, restoren*); or different origins, as in *ouerclose* (G+NG). The application of vernacular word-formation patterns to foreign words is observed in *ouerclose*, but also in the 13 instances accounted for by the juxtaposition of NG+G forms: *pouren into, pouren out,*

---

[3] The Latin form coexists in our corpus with the same anglicised imperative *note*.

etc. Verb compounds (*ouerhelen, ouerclose, withdrawen, ouertaken*) coexist with phrasal verbs (*helen ouer, drawen out, drawen doun, turnen up, turnen about, taken awei, taken doun*, etc.), a method inherited from Old English times. The fact that phrasal verb formation was still operative in late Middle English and in a particular register could be regarded as a symptom of the vernacularisation of science from a morphological point of view. On the whole, the scarcity of examples from G+NG or NG+G validates the perception of the language as being in a stage of change described by Burnley (1992: 445-46) as follows:

> […]after analysis of the word structure, there follows a period during which the word is stylistically differentiated from the rest of the lexis. It is synchronically recognisable by speakers of the language as foreign, and its affixes may be used to produce new formations with a restricted set of bases also perceived to be foreign. Such affixes are productive only within a subset of the lexis.

We propose to analyse the classification of derivational processes based on the classes of form involved[4]. These categories include:

* Class maintaining derivation produces lexemes that belong to the same class as the base from which they are derived, as in *holden/biholden*. Prefixation is a type of form-maintaining derivation in English.

* Class-changing derivation produces forms that belong to a different class from that of the base from which they are derived. In English suffixation is a class-changing process, as in *clarifie*.

Of the 228 derivative tokens, only 5 contain a suffix. They correspond to the type *clarifie* that, as we can see, illustrates affixation of NG provenance by means of the suffix *–fie* (usually added to nouns or adjectives to form verbs). As in any other field or discipline written in the vernacular, each independent

---

[4] Some scholars disagree with this classification. Lyons (1970), for example, believes that forms that have undergone class-maintaining derivation cannot be considered as belonging to exactly the same class because the derivative cannot, in turn, undergo the process of derivation by which it was formed.

suffix embodies the general patterns of the process in NG (mainly Latin and French) and G (basically English) languages. *–fie* is attached to a bound base to form a verb. As Harley (2006: 167) has noted: "None of the Germanic suffixes alter the phonological shape of their stems like that", meaning that Latinate suffixes do. *–fie* suffixation is an example of class-changing derivation.

As regards prefixes, we have already observed that the only combination of G+NG is found in Medicine (*ouerclose*).

All instances of a Germanic base combined with a Germanic prefix have been encountered in Astronomy (85 in all), which would seem to contradict the general tendency in scientific discourse to use Latinate forms. However, this may be due to the fact that they are all class-maintaining forms and, therefore, easily recognisable for the readership. There are 136 cases that illustrate the combination NG+NG; out of all of them, only 10 belong to the recipe compilation (including 4 instances of *remeue* and 2 of *amenden*). The NG+G combination does not occur anywhere in the samples of our corpus, further validating Moskowich's (1995) idea that Scandinavian items were not perceived as foreign by English speakers in the Middle Ages.

Some of the prefixes in our data are used exclusively with verbal bases. This is the case of *de-*, (Bauer 1983: 218) in *distillen*. Others (*vnder-* in *vnderstonden*, *with-* in *withdrawen* or *for-* in *forgeten*) have undergone a process of lexicalisation, in some cases reinforcement (Bauer, 1983: 56), already evident at the time our texts were produced. *Re-* is far more common with verbs (as in our examples *refresschen, restoren, remeuen)* than with nouns. Finally, though *a-, be-*, and *en-* are normally considered class-changing prefixes, our examples (*assendith, beginne, beholde, behete, encressith*) do not appear to confirm this tendency.

In general, lexicon-increasing morphology demanded by socio-external conditions seems to correspond to the level of productivity of the vernacular.

Further research may confirm that grammatical constraints affect the outcome of certain morphological processes (Bauer 2004).

4. CONCLUDING REMARKS

Word-formation processes have been considered here as one of the possible indicators of the degree of vernacularisation in English scientific writing at its very birth as a new register within the language. The derivational processes undergone by other word-classes such as nouns, the most abundant category in the so-called scientific register, have demonstrated that such vernacularisation was well underway at the time. Though less striking, results obtained with verbs also show that they too were adapting well in response to the need to transmit in the vernacular ideas and aspects of knowledge that had been always conveyed in Latin.

General word-formation patterns do not seem to vary for this specific use of English. However, in reference to Tenor (type of readership) and Field (discipline), we have observed some peculiarities. Depending on the discipline examined, different etymological origins, combined in various ways, have been identified. Likewise, the use of native or non-native morphological elements in each of the three extracts corresponds with the different levels of literacy and education among their respective readerships.

The variables we have considered (etymology, discipline and word-formation process) have proved to be valuable tools in assessing the degree of vernacularisation of early scientific English.

Begoña Crespo & Isabel Moskowich

University of Corunna

REFERENCES

Adams, V. 2001: *Complex Words in English*. Longman, Pearson Education Limited, Edinburgh.

Alonso Almeida, F. forthcoming: *A Middle English Medical Remedy Book from Glasgow University Library MS Hunter 185*. Universitätsverlag C. Winter, Heidelberg.

Barber, Ch. 1993: *The English Language. A Historical Introduction*. Cambridge University Press, Cambridge.

Bauer, L. 1983: *English Word-Formation*. Cambridge University Press, Cambridge.

Bauer, L. 2004: The function of Word-Formation and the Inflection-Derivation Distinction. In Henk Aertsen, Mike Hannay & Rod Lyall (eds.) *Words in their Places. A Festschrift for J. Lachlan Mackenzie*. Vrije Universiteit, Amsterdam. 283-292.

Beal, J. C. 2004: *English in Modern Times*. Arnold, London.

Biber, D. 1995: *Dimensions of Register Variation. A Cross-Linguistic Comparison*. Cambridge University Press, Cambridge.

Burnley, D. 1992: Lexis and Semantics. In Norman Blake ed. *The Cambridge History of the English Language. Vol. II*. Cambridge University Press, Cambridge. 409-499.

Carstairs-McCarthy, A. 2002: *An Introduction to English Morphology*. Edinburgh University Press, Edinburgh.

Dalton-Puffer, Ch. 1996: *The French Influence on Middle English Morphology: a corpus-based study*. Mouton de Gruyter, Berlin/New York.

Fernández, F. 1982: *Historia de la lengua inglesa*. Gredos, Madrid.

Halliday, M.A.K. 1978: *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. Edward Arnold, Ltd., London.

Halliday, M.A.K. 1988: On the Language of Physical Science. 2004. In Webster, J. ed. 2004: *The language of Science. The Collected Works of M.A.K. Halliday.* Continuum, London/New York 140-158.

Harley, H. 2006. *English Words. A Linguistic Introduction.* Blackwell, Oxford.

Kastovsky, D. 1992: Semantics and Vocabulary. In Hogg, R. (ed.), *The Cambridge History of the English Language, vol. I,* Cambridge University Press, Cambridge. 290-408

Kurath, H. *& al.* 1959-. *Middle English Dictionary.* http://ets.umdl.umich.edu/m/med/

Lyons, J. 1970: *New Horizons in Linguistics.* Penguin, Harmondsworth.

Marchand, H. 1969: *The Categories and Types of Present-Day English Word-Formation.* Beck, München.

Matthews, P. H. 1974: *Morphology. An Introduction to the Theory of Word-Structure.* Cambridge University Press, Cambridge.

Middle English Dictionary On-Line: http://ets.umdl.umich.edu/m/med/

Moskowich-Spiegel Fandiño, I. 1995: Language Contact and Language Change: The Danes in England. *Revista Alicantina de Estudios Ingleses.* 8: 139-153.

Moskowich-Spiegel Fandiño, I. & B. Crespo García 2006, forthcoming: *Lop-webbe* and *henne cresse*: Morphological Aspects of the Scientific Register in Late Middle English. *Studia Anglica Posnaniensia* 42.

Rissanen, Matti *& al.* 1991. *The Helsinki Corpus of English Texts. Diachronic Part.* Helsinki University Press, Helsinki. (Chaucer's The Equatorie of the planetis).

Sager, J. C., D. Dungworth, D. & P. F. Mcdonald 1980: *English Special Languages. Principles and Practise in Science and Technology.* Oscar Brandstetter Verlag, Wiesbaden.

Scotton, C. M. 1993: *Duelling Languages*. Oxford University Press, Oxford.

Skeat, W. W. ed. 1968: *Geoffrey Chaucer's A Treatise on the Astrolabe /* addressed to his son Lowys. Edited from the Earliest Mss. published for The Early English Text Society by the Oxford University Press, Oxford.

Taavitsainen, I. 2000: Science. *The Chaucer Companion*, ed. Peter Brown. Blackwells, Oxford. 378-396.

Taavitsainen, I. 2004: Transferring Classical Discourse Conventions into the Vernacular. *Medical and Scientific Writing in Late Medieval English,* eds. Irma Taavitsainen & Päivi Pahta eds. Cambridge University Press, Cambridge. 37-72.

Voigts, Linda E. 1989. Scientific and medical books. *Book Production and Publishing in Britain, 1375-1475*, eds. Jeremy Griffiths & Derek Pearsall. Cambridge University Press, Cambridge. 345-402.

Graph 1. Word counts and texts.
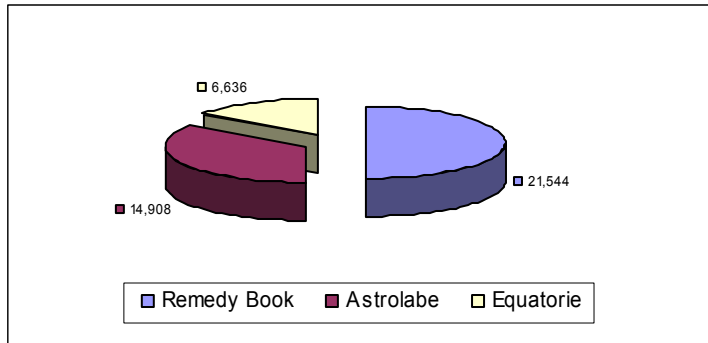
Graph 2. Discipline-based affixation
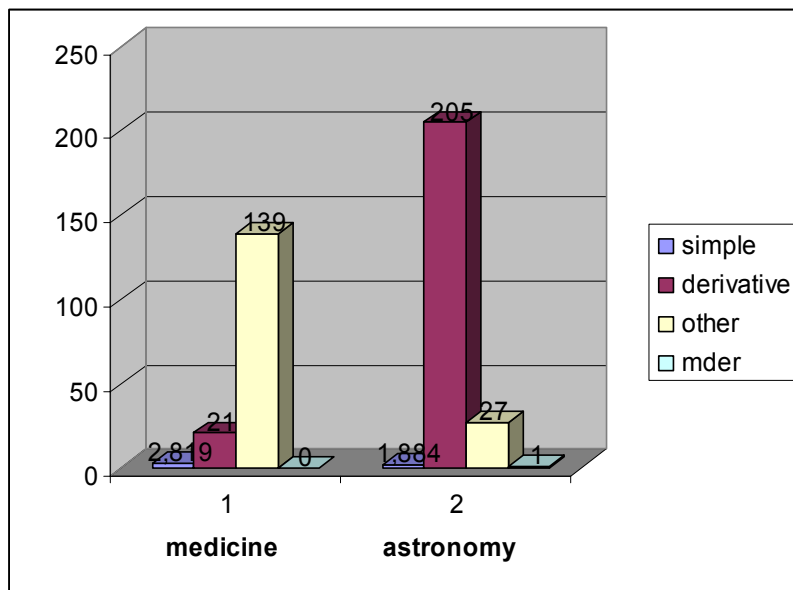
Table 1. Types with more than fifty tokens

Type	Tokens
Ben	790
Taken	553
Maken	191
Setten	190
Don	189
Lien	164
Hauen	113
Sayen	94
Drinken	87
Stampen	77
Knouen	70
Clepen	70
Finden	59
Geuen	53
Schauen	51

Table 2. The Etymology of Verbs

Origin	Tokens
G	3,982
G+G	259
G+NG	1
NG	696
NG+NG	141
NG+G	20
CS	19
UNK	8
TOTAL	5,126

Table 3. Medicine

Origin	Tokens
G	2,407
G+G	129
G+NG	1
NG	406
NG+NG	15
NG+G	13
CS	0
UNK	8
TOTAL	2,979

Table 4. Astronomy

Origin	Tokens
G	575
G+G	130
G+NG	0
NG	290
NG+NG	126
NG+G	7
CS	19
UNK	0
TOTAL	2,147

Table 5. Word-formation processes

Forms	%	Process	Tokens	%
Simple	91.74	S	4,703	91.74
Non-simple	8.12	Der	226	4.4
		Other	196	3.82
		Mder	1	0.01
		TOTAL	5,126	

Table 6. Derivation in astronomy

Process	Origin	Tokens
S= 1,884 tokens	G	1574
	G+G	0
	G+NG	0
	NG	271
	NG+NG	0
	NG+G	0
	UNK	0
	CS	19
DER= 205 tokens	G	0
	G+G	80
	G+NG	0
	NG	0
	NG+NG	125
	NG+G	0
	UNK	0
	CS	0
OTHER= 57 tokens	G	0
	G+G	50
	G+NG	0
	NG	0
	NG+NG	0
	NG+G	7
	UNK	0
	CS	0
MDER	NG+NG	1

Table 7. Derivation in medicine

Process	Origin	Tokens
S=2,819	G	2,405
	G+G	0
	G+NG	0
	NG	406
	NG+NG	0
	NG+G	0
	UNK	8
	CS	0
DER=21	G	0
	G+G	5
	G+NG	1
	NG	0
	NG+NG	15
	NG+G	0
	UNK	0
	CS	0
OTHER=139	G	2
	G+G	124
	G+NG	0
	NG	0
	NG+NG	0
	NG+G	13
	UNK	0
	CS	0
MDER	NG+NG	0

* † *