Hippocrates transformed: crafting a Hippocratic discourse of medical semiotics in English, 1850–1930

Kamran I. Karimullah

This study presents a methodology for adapting corpus linguistics to the genealogical analysis of translation’s role in the evolution of medical concepts. This methodology is exhibited by means of a case study that draws on a number of corpora to explore how two English translators—Francis Adams, a Scottish physician, and Williams H.S. Jones, a Cambridge philologist, classicist and ancient historian—translated a set of terms in Hippocratic medical texts that refer to how the body reveals illness. Drawing on the Genealogies of Knowledge subcorpora of ancient Greek and modern English, it examines some of the ways in which translation contributes to the creation of a Hippocratic semiotic discourse in English whose lexical features differ from those attested to in the subcorpus of Greek Hippocratic texts. A comparative analysis of keyword frequency and collocations of Greek semiotic terms such as sēmeion, and English terms such as sign and symptom reveals the different translation strategies Jones and Adams used to translate the text. The result of this process is a Hippocratic semiotic discourse in English whose lexical features do not reflect those in the Hippocratic texts in a straightforward way.
Introduction

Hippocrates (c. 460 BCE–c. 375 BCE) was an ancient Greek physician who is traditionally regarded as the father of Western medicine. Despite this reputation, it has long been recognised that the 60-odd medical texts that together make up what is typically referred to as the ‘Hippocratic corpus’ could not have been written by a single individual in the 5th or 4th century BCE. Rather, significant variations in writing style, subject matter and date of composition suggest that these texts were the output of a group of writers (Jouanna, 1999). This set of texts presents a number of concepts, attitudes, practices, and terms that, while not always coherent in themselves, were gradually synthesised by notable physicians such as Galen of Pergamum (d. c. 216 CE), into what is today called ‘Hippocratic’ medicine. Such was the endurance of Hippocrates’ medical legacy that even at the end of the 19th century, Hippocratic medical thought and Hippocratic texts remained “meaningful in the day-to-day medical practice” of physicians (Smith, 1979, p. 32; Cantor, 2002). Even as medical research and clinical methods rapidly developed in Europe and the United States in the 19th century, the pioneering French medical researcher Claude Bernard (1813–1879) sought to trace the genealogy of his physiological theory of what was later called ‘homeostasis’ to “the healing power of nature” which Bernard purports was described by Hippocrates more than two millennia earlier (cited in Porter, 1997, p. 340).

Hippocratic medical doctrines relating to semiotics and prognosis in particular continued to be important in modern European medicine until the 19th century, and continue to attract concerted scholarly interest even today (Roselli, 1990; Holmes, 2010, 2015; Manetti, 1993; Burney, 182, 1994; French, 1998). Yet, in the history of medicine and classical reception, translation has been undervalued as a vehicle for mediating the modern reception of Hippocrates. One question that has so far remained unexplored is how modern English translations of Hippocratic texts have mediated the reception of Hippocratic semiotic thought in the English-speaking world. While Hippocrates has long been recognised as an important source for Greek thinking about semiotics (Manetti, 1993), by the 19th century medical semiotic terminology in English was going through a process of evolution in which terms such as sign and symptom—the focus of analysis in this study—were shifting in meaning away from Hippocratic and Galenic usage. To some extent these shifts were part of gradual lexicographical shifts in English medical writing (Tyrkkö, 2006, 2010). But they were also partly attributable to the increasing importance physicians were assigning to laboratory test results in clinical medicine in the 19th century, and to a new vision of the human body (Holmes, 2015; Foucault, 1973; Porter, 1997, pp. 320–333). When 19th-century English translators were confronted with the task of translating Hippocrates, they had to adapt Hippocratic terminology and style to the evolving terminology, style, and conceptual frameworks of modern English medical writing. These translations did not simply transmit Hippocrates to modern English-speaking audiences, but were part of a language-mediated process of transformation in which Hippocratic medical concepts were made to adapt to new cultural, scientific, linguistic, and historical environments.

The consequence of this translation-mediated adaptation was the formation of what we may call a Hippocratic medical discourse in English. This was more than just a collection of English texts with Hippocrates identified as the author. It was a set of English texts attributed to Hippocrates, but whose lexical patterning was dictated in part by the lexical patterning in the source and target languages, and in part by social, political, economic, literary, and institutional factors in 19th-century Britain. A full study of the rise of a Hippocratic medical discourse in English would analyse noteworthy lexical patterns in English and Greek as well as the non-linguistic factors that motivated the patterning in question. In this paper, I shall analyse data relating to the former in order to document how Hippocrates translators activate particular 19th-century medical categories and vocabularies to construct a Hippocratic semiotic theory by translating key texts in the Hippocratic corpus.1

The study is, therefore, genealogical, in line with the loose sense the Genealogies of Knowledge project assigns the term in their publications (Jones, 2019a, 2019b; Karimullah, 2020a; Baker, 2020). In the social sciences and humanities genealogy is typically associated with Michel Foucault.2 Our studies share with Foucault a broad concern with discourse as a pattern in which the way power shapes the historical human subject may be observed; they also share with Foucault a view of history that is neither linear, discontinuous, nor end-directed. Like Foucault our histories focus on relations between forces that impress themselves on discursive patterns rather than the “eruption of an event” (Foucault, 1984a, p. 88); nor do we assume that concepts such as freedom and reason are historical constants. In the case of classical reception and the history of medicine, my study draws inspiration from genealogical studies such as Brooke Holmes’ (2010) study on the emergence of the concept of body in ancient Greek literature, and Helen King’s (1993) study on the how the disease called ‘hysteria’ slowly emerged as a gendered pathology in modern European medicine. Rather than denoting a specific methodology, saying that these studies are genealogical is simply to say that they share a set of assumptions about history, truth, the human subject, the event, knowledge, and discourse. Unlike Foucault, however, our research is aimed at revealing how translation plays a significant but understudied role in the evolution of political concepts in time and across cultures (Baker, 2020; Karimullah, 2020a, 2020b; Jones, 2019a, 2019b). This study is genealogical in the sense that it traces the emergence of a Hippocratic medical discourse in English in the 19th century, while recognising that the emergence of the relations between body, knowledge, and illness this ancient discourse propounds are given a modern inflection by translation.

In addition to its focus on translation, two further features distinguish how this study uses genealogy from how Foucault, as well as Holmes (2010) and King (1993) used it. The first is that this study seeks to supplement genealogical research with methods drawn from corpus linguistics. Second, this study is unapologetically empirical in its attitude towards data collection and quantitative analysis. For Foucault genealogy had certain broad implications relating to what kinds of questions should be asked; for example, research in the social sciences and humanities should reduce the explanatory significance of authorship, emphasising instead questions pertaining to the historical evolution of discursive patterns (Foucault, 1984b); research should take the body, the “inscribed surface of events” (Foucault, 1984a, p. 83), as the starting point of historical analysis; research should analyse less timeless objects such as the self, truth, or morality and turn instead to document the historical descent of such concepts in history (Foucault, 1984a). Throughout his life, even in his seminar lectures at the Collège de France, Foucault’s work was unified by an overarching set of questions rather than in the articulation of a certain methodology. By contrast, like the other contributions in this special issue, this paper shows how research that begins from a genealogical perspective can use methods from corpus linguistics to conduct genealogical research on a variety of themes. In particular, this study shows in detail how genealogical questions in the medical humanities can be combined with corpus linguistics to produce conclusions that are consistent with genealogically oriented research and methodologically cogent. Synthesising genealogy with corpus linguistics takes on the
following features in this study, all of which will be explained in
detail in the section “Data and methodology”. The first relates to
the axis of data collection:

- **Corpora as empirical data**
- Corpus construction determined by sampling frames
- **Comparisons between lexical patterns in corpora in the same language**
- **Comparisons between lexical patterns in corpora in different languages**

Moreover, in line with the increasing methodological and
statistical sophistication in corpus linguistics (Hardie and
McEnery, 2011; Brezina, 2018b), this study takes a more
empirical, or quantitative, approach to the analysis of corpora and
patterns in lexical data. Foucault probably would have scoffed at
the insistence that research should be ‘empirical’. Indeed,
empiricism, especially medical empiricism, is one of those
“suprahistorical” forms of perception the genealogy of which
Foucault sought to uncover in, for example, *The Birth of the
Clinic* (Foucault, 1973). In this study, synthesising genealogy with
the methods of corpus linguistics has the following additional
features. This axis pertains to quantitative analysis:

- Word and word-phrase frequency as primary observed variables
- Word and word-phrase frequency assessed according statistical significance
- Statistical parameters sensitive to syntactical differences between English and Greek
- **Qualitative analysis** guided by statistical measures

Hence, in what follows, I use subsets of ancient Greek and modern English subcorpora compiled by the Genealogies of
Knowledge project to provide a corpus-based analysis of how the
process of translation contributed to the formation of a Hippo-
cratic medical discourse in English. More specifically, I identify
and examine noteworthy lexical patterns around the English
semiotic terms sign and symptom. I argue that these patterns do
not necessarily reflect the patterning of the Greek Hippocratic
source text, but result from the transformative effects of trans-
lation from Greek into English in the 19th-century and early
20th-century British Isles. To address these issues, I restrict my
analysis to two British translators, Francis Adams (1796–1861)
and William Henry Samuel Jones (1876–1963), both of whom
translated Hippocratic texts dealing with medical semiotics.

Adams and Jones made impressive contributions to the
reception of Hippocratic texts in English. Adams, a physician by
training, was an amateur classicist and historian of medicine
from Scotland. His *The Genuine Works of Hippocrates* (1849)
was the first systematic translation of Hippocratic texts from
Greek that set out the entire system of Hippocratic medical
thought for an English-speaking audience. A number of English
translations of Hippocratic treatises existed as early as the late
16th century (Adams, 1849, pp. 47–130), but they were individ-
ual treatises that were more often than not based on Latin
translations of the Hippocratic corpus produced in the 16th
century rather than directly on Greek texts. By contrast, Adams’
was the first English translation to be based on the critically
edited Greek text by the great French philologist and physician,
Émile Littré. Jones’ contributions to the history of ancient
medicine are no less impressive. A Cambridge academic for his
entire professional life, Jones was a renowned classicist and
historian of ancient medicine. He was also an accomplished
classical philologist and the primary editor and translator of the
Loeb classical library’s complete edition and translation of the
Hippocratic corpus.

Both translators held Hippocrates in high esteem and were
committed to preserving the meaning and integrity of the Hip-
pocratic text (Jones, 1923, p. vii; Adams, 1844, p. x; Adams, 1856,
p. xx; Adams, 1849, p. iii). For each, the best way to convey what
Hippocrates said was to produce literal translations that, in
Adams’ words, provided a “true version of the language” of
Hippocrates (Adams, 1849, p. vii). Their translations exemplify
the expectation in the 19th century that translation should prefer
“accuracy to invention” (Talbot, 2015, p. 61; Stray, 1998, p. 297).
In spite of their quest for literalness, scrupulous accuracy, and
faithfulness to the source text, corpus-assisted analysis reveals
how Adams and Jones transformed Hippocrates’ medical dis-
course in subtle but profound ways.

This paper therefore aims to examine how these transforma-
tions are brought about by translation, not to pass judgement on
their quality or to examine translation strategies. Hence, in
section “Data and methodology”, I describe the corpus data,
statistical measures, and the corpus-based methodology that the
study is based on. This section details the methodological prin-
ciples on which the case-study is carried out. These principles fall
along the aforementioned axes, the data collection axis pertain-
ing to the construction of corpora and a quantitative axis per-
taining to statistical analysis. I want to emphasise that the aim of
this section is to make explicit the methodological assumptions
the analysis is based on, and, simultaneously, to offer the
methodology thus outlined as one way genealogical research
might be combined with corpus linguistic methods in the future.
The study then turns to examine the features of the Greek and
English semiotic discourses in Hippocratic texts included in the
corpus. In this final section I show how Jones’ and Adams’
translation techniques contribute to the emergence of a Hippo-
cratic semiotic discourse involving signs and symptoms in English
where such a discourse does not feature in analogous ways in the
original Greek texts. Naturally, whether the methodology out-
lined in “Data and Methodology” is successful depends in part
on how compelling the results of the case-study are as a gene-
alogy of semiotic discourse in 19th-century medical English.
The reader will doubtless form their own view. Nonetheless, in
the conclusion I shall return to consider how successful the
methodology is as a genealogical exercise.

### Data and methodology

**Corpora, frequency, collocation, and metrics.** This study documents key lexical patterns that typify a translated medical
discourse in English and compares them with lexical patterns in
the ancient Greek source texts. The aim is to record differences
between lexical patterning relating to semiotic terms that cannot
be entirely a result of differences between the two languages. To
this end, I use comparable English and Greek corpora (Hardie
and McEnery, 2011, p. 20), concordances, as well as frequency
and collocation analysis to combine observations of macrolevel
lexical patterns in English and Greek medical writing with close
analysis of individual passages in Greek and English. Utilising
corpora in this manner allows us to explore how a modern
Hippocratic medical discourse in English emerged and to assess
the role translation played in this process.

The data on which this study is based come from modern
English and ancient Greek subcorpora built by the Genealogies
of Knowledge research team. I use the Genealogies of Knowledge
concordance browser to generate the core and reference corpora,
collect and visualise word frequency and collocation figures,
and to generate KWIC concordance lines (Luz and Sheehan, 2020).3
The Genealogies of Knowledge subcorpora consist of two core
corpora and two reference corpora.
Core corpora.

- The English core corpus consists of translations of a selection of key Hippocratic texts by Francis Adams (1849) and William Jones (1923–1931), totalling 132,371 tokens (Jones 57,359 tokens; Adams 75,012 tokens). The Greek source texts were central to the development of the Western medical tradition, which explains the attention given to them by Jones and Adams. The texts cover a variety of topics and assume a number of different textual forms. Some deal with the medical theory of the elements and humours, others with prognosis, diagnosis, medical regimen, and medical ethics. Some, such as the *Epidemics*, take the form of case histories, recording in detail the progress of diseases that struck islands in ancient Greece; others, such as the Hippocratic *Aphorisms*, are a collection of wise medical sayings (Table 1).

- The Greek core corpus consists of the original Greek texts attributed to Hippocrates, assumed to have been written between the fifth to the fourth centuries BCE; totalling 102,358 tokens (Table 2).

The sampling frame for these core corpora included a number of considerations. The texts had to be Hippocratic as far as the translators were concerned. This meant that in order to be included in the core corpora, it was not required that the texts themselves were or are classed among Hippocrates’ genuine works, but only that Jones and Adams viewed them as Hippocratic texts (Smith, 1979; Lloyd, 1993; Craik, 2018). Thematicaly, the texts included had to demonstrate some relevance to the kinds of questions about scientific method that are in line with the broad aims of the Genealogies of Knowledge project. Hippocratic texts that do not engage with methodological or epistemological issues related to science—Hippocratic texts on surgery, for example—were not included, even if they were thought to be genuine. Finally, it is important to recognise that I am principally concerned with analysing the lexical patterns of two Hippocratic *discourses*, not patterns in a particular text or set of texts. It was not critical, therefore, in this study that the Greek and English core corpora should be *parallel* corpora, in the sense that the Greek source and English target texts must form a source–target pair (Hardie and McEnery, 2011, pp. 19–20), since the patterns I examine here emerge in the discourse cumulatively. To this extent the pair of core corpora should be thought of as comparable corpora since they have been constructed using the same sampling frame (Hardie and McEnery, 2011, p. 20).

Reference corpora.

- The English reference corpus totals 6,782,599 tokens. It is a corpus of English translations of ancient Greek philosophical, historical, and scientific writings published in the 19th and early 20th centuries contained in the Genealogies of Knowledge modern English corpus. This includes, for example, English translations and retranslations of classical Greek philosophers, such as Plato and Aristotle, and Greek historians such as Thucydides and Herodotus, alongside medical authors such as Galen. The English reference corpus is intended to be representative of the English that was used to translate classical Greek authors into English in the 19th and early 20th centuries.

- The ancient Greek reference corpus totals 3,168,966 tokens. This corpus includes the original Greek writings of classical philosophers, historians, and physicians mentioned above, as well as later Greek authors, such as Diogenes Laërtius (fl. c. 250 CE) and Procopius (d. after 555 CE). The Greek reference corpus is intended to be roughly representative of scholarly Greek usage from classical to late antiquity.

The sampling frame for the Greek and English reference corpora involved a number of considerations. Some, such as the stipulation that the texts included should be thematically relevant to the project’s overarching research agenda relating to the development of political and scientific discourse, are similar to the criteria for selecting the core corpora. Likewise, there is no stipulation that the English and Greek texts included in the reference corpora should form a source–target pair. Two further stipulations are that the texts included in the English reference corpus should be *translations* of ancient Greek works and that they were published before 1945. Finally, the size of the reference corpora was also dictated by the role it was intended to play in the analysis. It is common, but by no means necessarily, for reference corpora to be larger than core corpora, though, as might be expected, there is no strict rule that stipulates how much larger (Baker, 2006, pp. 30–31; Brezina, 2018b, p. 81). In this study, the reference corpus is intended to represent the English that classicists who were contemporaries or near contemporaries of Adams and Jones used to translate Greek texts recognised as pivotal in Western intellectual history.

| Translator | Title | Tokens | % Total |
|------------|-------|--------|---------|
| Jones      | Oath  | 437    | 0.330   |
| Jones      | Airs, Waters, Places | 9030 | 6.822 |
| Jones      | Nature of Man | 4654 | 3.516 |
| Jones      | Prognostic | 7350 | 5.533 |
| Jones      | Law | 1574 | 1.189 |
| Jones      | Epidemics (Books One and Three) | 18,245 | 13.783 |
| Jones      | On Ancient Medicine | 7958 | 6.012 |
| Jones      | Regimen in Acute Diseases | 8111 | 6.127 |
| Adams      | The Oath | 369 | 0.279 |
| Adams      | On Airs, Waters and Places | 9744 | 7.361 |
| Adams      | The Law | 477 | 0.360 |
| Adams      | Aphorisms | 10,467 | 7.907 |
| Adams      | Epidemics (Books One and Three) | 16,975 | 12.824 |
| Adams      | On Regimen in Acute Diseases | 16,608 | 12.547 |
| Adams      | On Ancient Medicine | 7299 | 5.514 |
| Adams      | On the Sacred Disease | 6041 | 4.564 |
| Adams      | Book of the Prognostics | 7032 | 5.312 |
| Total      |       | 132,371 |        |

| Title | Greek title | Tokens | % Total |
|-------|-------------|--------|---------|
| Airs, Waters, Places | Περὶ Ἄρων, Ἡδατῶν, | 7459 | 7.287 |
| Aphorisms | Aphorismoi | 7097 | 6.934 |
| The Nature of Man | Περὶ Φυσεῶς Ἀνθρώπου | 3900 | 3.810 |
| Prognostic | Προγνώστικον | 5221 | 5.101 |
| Oath | Ἑρωκός | 326 | 0.318 |
| Law | Νόμος | 326 | 0.318 |
| On Ancient Medicine | Περὶ Ἀρχαῖας Ἰετρικές | 5531 | 5.404 |
| On the Sacred Disease | Περὶ Ηἰερῶν Νοσου | 4719 | 4.610 |
| Epidemics | Ἐπίδημιος | 41,872 | 40.907 |
| On Generation | Περὶ Γενεσίς | 1951 | 1.906 |
| Regimen in Acute | Περὶ Διατῆς Ὀξεῖν | 11,829 | 11.556 |
| Diseases |       |       |         |
| Total |       | 102,358 |        |
To compare and contrast the various strategies that Jones and Adams used in their Hippocrates translations, to discover what terms they used on which occasions to translate Greek semiotic terminology, or to assess the quality of the translations is not the principal aim of this study. I am interested, rather, in the frequency and collocation features of the Greek and English core corpora, which I treat as lexical features of Hippocratic medical discourses in the source language (Greek) and a modern target language (English). Word frequency stands at the heart of the corpus-linguistic method used in this paper (Baker, 2006, pp. 47–48). Its importance in this study arises from the viewpoint that no choice of word is neutral or random, and that the repetition of salient words or phrases is motivated by an a Hippocratic author’s or translator’s assumptions about or attitudes towards medical knowledge and therapy. Collocation, or the regular co-occurrence of lexical items in a statistically significant way (Baker, 2006, pp. 95–96), is another central element in the corpus linguistics. In this study, collocations derive their significance from the assumption that the particular semantic features of the Greek σημεῖον in the Greek core corpus and of sign and symptom in the English core corpus are not contained in the word in isolation, “but rather subsist in the characteristic associations” that these words regularly participate in (Hardie and McEnery, 2011, pp. 122–123). Relying mainly on frequency and collocational analysis, therefore, I isolate key patterns surrounding semiotic terms in each core corpus, in order to compare some of the features of the Hippocratic medical discourse in the two languages. These comparisons allow me to identify the cumulative effects that translation into English has on the Greek Hippocratic medical discourse. I take the differences observed in the lexical patterning in the Greek and English core corpora as evidence of the transformative effect that translation exerted on Hippocratic concepts. Comparison, therefore, plays an important role in this study. I also compare frequency data and collocational patterns in the English core and reference corpora and in the Greek core and reference corpora. In order to facilitate such comparisons, I shall use a number of statistical measures that are familiar to corpus linguists (Baker, 2006; Manning and Schütze, 1999; Oakes, 1998; Brezina, 2018a, 2018b). These measures are critical in studies that involve trying to assess the magnitude and significance of differences in lexical patterning in comparable corpora (Brezina, 2018a, 2018b, p. 260). They permit us to carry out comparisons between corpora in the same language and between corpora in different languages. Some of them allow me to establish, in relation to the Greek and English core corpora, whether the frequency of certain semiotic terms is high enough to contribute to elaborating a specific semiotic discourse in relation to the reference corpus (the “effect-size”, Gabrielatos, 2018). Table 3 presents the parameters that are used to calculate and compare the magnitude of difference between the keyword frequencies in the core and reference corpora.

Other metrics allow me to judge how likely it is that the differences in the frequencies of semiotic terms in the Greek and English core corpora are motivated or merely owing to chance (the “statistical significance”, Gabrielatos, 2018). Table 4 presents the parameters that are used to calculate the magnitude and significance of collocation patterns in the core corpora. These give a sense of the strength of association between collocations in Greek and collocations in English, and allow us to compare the association strengths between semiotic terms and their collocates in the core corpora. The selection of these parameters was also informed by the differences in the sizes of the core and reference corpora. The use of normalised rather than absolute frequencies, for example, is essential for comparing word frequencies between the Greek and English core corpora and between the respective core and reference corpora. On the other hand, the selection of MI3 as the collocation significance parameter, for example, was dictated by the fact that, unlike MI, it does not privilege rare collocations and, unlike Loglikelihood, is less sensitive to large differences between core and reference corpora (Brezina, 2018b, pp. 84–85).

Finally, despite the importance of quantitative analysis, this study is based fundamentally on qualitative analysis. Like much of the research carried out in corpus-based discourse analysis (Baker, 2006), it makes use of quantitative methods in order to gain greater insight into the analysis of texts, to identify the particular features of semiotic discourses in Greek and English texts, and to strengthen the evidentiary value of particular passages selected for qualitative analysis as representative of the discourse under study.

Motivations for focusing on σημεῖον, sign, and symptom. In Greek Hippocratic texts, corporeal and behavioural indicators of disease served as a key element of prognosis, a medical subdiscipline that receives considerable attention in pivotal Hippocratic texts (Manetti, 1993, pp. 37–39; Jouanna, 2013, pp. xxxii–xxxiii). In this theory, as presented in the Greek texts, the term σημεῖον, meaning indication or sign, features prominently (Manetti, 1993), though Hippocratic authors also used other terms to refer to what we would today call signs or symptoms of disease. Analysis of σημεῖον’s frequency in the Greek core corpus in comparison with the ancient Greek reference corpus indicates that σημεῖον is especially salient in this collection of Hippocratic texts in comparison with Greek scientific and historical literature more generally.

| Statistic ID | Statistic name | Statistic cut-off value | Normalised frequency scale (in thousands) |
|--------------|---------------|-------------------------|-----------------------------------------|
| effect-size  | Log ratio     | 1                       | 10                                      |
| stat. sig    | BIC           | 6                       | 10                                      |

**Table 3 Keyness parameters.**

| Statistic ID | Statistic name | Statistic cut-off value | L and R span | Minimum collocation freq. | Filter |
|--------------|----------------|-------------------------|--------------|---------------------------|--------|
| ENG ADJ      | M13            | 6                       | 1            | 2                         | Function words removed |
| GR ADJ       | M13            | 6                       | 5            | 2                         | Function words removed |
| VERB         | M13            | L10–R10                 | 2            |                           | Function words removed |

*On the calculation and interpretation of the mutual information metrics (MI, M12, and M13), see Oakes (1998), which assumes a value of span = 1. See Brezina (2018a, 2018b, p. 69) for calculating MI using corrected estimated collocation frequency within a window of span *n*. 

It’s important to note that in this study, the authors are comparing the frequency and collocation features of the Greek and English core corpora to understand the cumulative effects of translation on the Hippocratic medical discourse. The keyness parameters (Table 3) help in identifying significant patterns in the corpora, while the collocation parameters (Table 4) provide insights into the strength of association between semiotic terms and their collocates. The use of normalised rather than absolute frequencies and the selection of MI3 as the collocation significance parameter are crucial for making fair comparisons between core and reference corpora. The qualitative analysis, however, is essential for gaining deeper insights into the discourse, indicating that the translation from Greek to English has had a significant impact on the frequency and collocation patterns of semiotic terms in the Hippocratic medical discourse.
Table 5 Raw and normalised frequencies of ἱνακόσια (singular, includes nominative, accusative, genitive, dative) and ἱνακόσια (plural, includes nominative, accusative, genitive, dative) in the ancient Greek core and reference corpora.

| Lemma | Raw frequency core corpus | Normalised frequency core corpus | Raw frequency reference corpus | Normalised frequency reference corpus | Effect-size | Significance |
|-------|---------------------------|----------------------------------|--------------------------------|----------------------------------------|------------|-------------|
| semeion | 53                        | 5.2                              | 490                            | 1.5                                    | 2.17       | 63.14       |
| semeia | 51                        | 5.0                              | 352                            | 1.1                                    | 1.74       | 44.58       |

See Table 3 for keyness parameters.

Table 5, which allows us to compare normalised frequencies of semeion/semelia in the Greek core corpus with those in the Greek reference corpus, reveals that the singular semeion (in nominative, accusative, genitive, and dative cases) occurs more than three times as frequently in the core corpus than in the reference corpus, and the plural, semelia (nominative, accusative, genitive, and dative cases), occurs more than four times as frequently. Moreover, the differences between expected and observed frequencies for both items are large enough to confidently rule out the possibility that the frequency of semeion and its declined forms in the Greek core corpus is merely the result of the typical distribution of semeion in Greek texts: the BIC scores calculated to measure the statistical significance of these differences are well above the conventional cut-off value of 6. These results suggest that the high frequency of semeion in the core corpus in relation to the reference corpus is motivated and contributes to the elaboration of a specific semiotic discourse in the original Hippocratic texts.

While in the Greek Hippocratic corpus semeion was the chief semiotic term, in 19th-century English medical writing symptom and sign were the two principal semiotic terms (Tyrkkö, 2006, 2010). For Francis Adams, William Jones and other translators of classics, this meant that the lexical options available to them were limited to these two terms. This fact becomes evident when we observe that semiotic terms such as token and accident that were used frequently in English medical writing prior to the 19th century hardly occur in Jones’ and Adams’ Hippocrates translations. Token is not used at all by either author; accident is used four times: once by Jones and three times by Adams. The generic term, indication, is only used four times: three times by Adams and once by Jones. Concordance 1 presents examples of how accident and indication are used to render the Greek source text.

Concordance 1. Jones’ and Adams’ use of ACCIDENT and INDICATION in the English core corpus, followed by Greek source and literal English translation of Greek by the Author.

1. In many cases erysipelas, from some obvious cause, such as an accident, and sometimes from even a very small wound, broke out all over the body …

Πολλοί μὲν τὸ ἔρυσικά μετὰ προφάσος ἐπὶ τοῖοι τυγχάνει καὶ πάνω ἐπὶ σύμφοιν τραυματίας ἐφ’ ἀλλ’ τῷ ὀξύτατῷ

[Many are affected by erysipelas all over the body when accompanied by a cause that is by chance or by very minor wound …]

2. Many were attacked by the erysipelas all over the body when the exciting cause was a trivial accident or a very small wound …

Πολλοί μὲν τὸ ἔρυσικά μετὰ προφάσος ἐπὶ τοῖοι τυγχάνει καὶ πάνω ἐπὶ σύμφοιν τραυματίας ἐφ’ ἀλλ’ τῷ ὀξύτατῷ

[Many are affected by erysipelas all over the body when accompanied by a cause that is by chance or by very minor wound …]

3. The diaphragm has obtained its name (frenes) from accident and usage, and not from reality or nature …

Ἀ δὲ φρενές ἄλλος ἄλομον ἔχουσιν θ’ τύχῃ κεκτημένον καὶ τού νόμῳ τῷ δ’ ἔοντι ὀξύτ., ὀὐδὲ τῇ φύσι.: The but diaphragm has an unsuitable name acquired by chance and by custom, and not by nature …

4. When doing everything according to indications, although things may not turn out agreeably to indication, we should not change to another …

Πάντα κατὰ λόγων πιστεύοντει τρυφεῖν τοὺς κατὰ λόγων, μη μεταβλασίαν ἐν ἐτέρων

[When everything proceeds according to the rule but does not turn out according to the rule, do not pass to another …]

In Jones’ translation of the Hippocratic Epidemics (line 2), ‘trivial accident’ is used to refer to a minor wound (σμικρὸς τραυματισμόν), as opposed to a serious threat to health. In Adams’ translation of the Epidemics (lines 1 and 2) ‘accident’ is used to render the Greek term τοῖοι τυγχάνει, or ‘things that happen by chance’; in On the Sacred Disease Adams refers in line 3 to something that is by chance and convention in opposition to what is by nature (phasis). Hence, in the English core corpus accident is never used as a semiotic term—that is, it is not a word in the Hippocratic medical lexicon that is used in contexts in which corporeal indicators in the patient’s body or behaviour are linked to the physician’s task of making a diagnosis of the malady or making a prognosis of the patient’s likely outcome. Unlike accident, Jones and Adams use indication as a semiotic term in their translations but at nowhere near the frequency of sign or symptom. In line 4, Adams uses indication in a way that seems to refer to the corporeal signals on the body or in the patient’s behaviour that gives insight into the progress and outcome of the disease. It is noteworthy, moreover, that in all these instances indication is used to translate generic phrases such as κατὰ λόγον (‘according to reason or rule’), not to translate semeion.

Given the above noted frequencies and patterns of use, the rest of this article will focus on a close analysis of the use of sign(s) and symptom(s) in the English core and reference corpora and semeion/semelia in the Greek core and reference corpora.

Sign and symptom in the Hippocrates translations by Adams and Jones

In the English core corpus—that is, the collection of Hippocrates translations by Adams and Jones—we observe that symptom is much more frequent than sign. Part of this is owing to the fact that Adams and Jones tend to use symptom more than sign. Figure 1 shows that Jones tends to use sign about half as often as symptom and that Adams uses sign infrequently, about once every 10,000 words.

Symptom is the ninth most frequent non-catchword word in the English core corpus, whereas sign and signs are considerably
less frequent. Table 6 allows us to compare the normalised frequencies of sign and symptom in the English reference corpus with their frequencies in the English core corpus. Symptom is a keyword in the English core corpus, but sign is not.

It is also interesting to observe that sign and symptom—especially the latter—are used far more frequently in relative terms in the English core corpus in comparison with the English reference corpus (Table 6). These observations may be accounted for in part by the fact that Adams and Jones are faithfully rendering the texts in major Greek Hippocratic writings where Hippocrates talks about diagnosis and prognosis in terms of corporeal signs or indicators in the patient. Yet for historical reasons the high relative frequency of symptom in the English core corpus cannot be accounted for by the high frequency of sumptōma in the Greek core corpus, since sumptōma is not used as a semiotic term in the Hippocratic corpus (Holmes, 2015, p. 192). Cognates of sumptōma occur seven times in the Greek core corpus, but, as the following examples from the corpus show, they are not used as semiotic terms.

Concordance 2. Cognates of SUMPTŌMA in the Greek core corpus.

1. Θεωρεῖν δὲ οὕτω διῄ—[μεταβολαί] χρωμάτων, συμπτώσεις φλεβῶν, ὑγκαν υποκονδρίῶν, ἀναρρήψια, καταρρήψια
It is therefore necessary to observe: changes in the complexion, collapsing of the vessels, swelling in the abdomen, motion upwards, motion downwards …

2. ἵνα πάλιν ἁρχήν ἀναθέρειος λαμβάνῃ τὸ σώμα· μηδὲ τὰς ξυμπτώσιας ἐκ τὸ ἐσθατὸν ἄγειν, σφαλερὸν γὰρ,
[So that once again from the start the body is able to rejuvenate, but the reduction [of flesh] should not be taken to the farthest limit, for it is danger …]

3. Ἐν γυμνασίῳ σημείον, ὦ ἀνας ὃ ἔτεος στάχυν ὃς ἔξειζ άσπερ ἐκ ὀσχεῶν, ἢ ξυμπτώσις ἐκ ἐπάρσιος.

[An indicator (smeion) during exercises: sweat flowing drop by drop that discharge as from pipes or a collapse from swellings.]

In most instances, therefore, the cognates of sympţoma refer to the collapsing of the body’s parts or capacities, though in the Aphorisms, the author says that weight-loss may be advantageous to the patient, but “these reductions (tas sumptōsias) should not be taken to the extreme”.

The pervasiveness of these semiotic terms, especially symptom, in the English core corpus cannot be explained by a high relative frequency of smeion in Greek either. Table 7 compares the frequency of this pair of English semiotic terms in Jones’ and Adams’ translations with the frequency of smeion/smeia in the Greek core corpus.

Table 7 tells us that in their English translations Adams and Jones use the singular forms, sign and symptom, taken together slightly more frequently than smeion is used in the Greek core corpus, Adams at nearly the same frequency but more than 1.5 times as frequently in Jones. Moreover, each translator uses their plural forms taken together more times in English than could correspond to the Greek, Jones more than five times as frequently, and Adams more than three times as frequently. While it is true, then, that the high frequency of semiotic terms in English reflects the presence of a discourse of medical semiotics in Greek, these observations suggest that this discourse plays a more prominent role in the English core corpus of Hippocrates translations than in the Greek core corpus of Hippocratic texts.

It is also unlikely that the high frequency of symptom in the English core corpus can be explained by a general tendency in 19th-century classics to render smeion as symptom in a more or less one-to-one fashion. To the contrary, translators of classical Greek in the 19th and early 20th century generally translated smeion with sign, and reserved symptom for explicating-translations of passages that speak about a variety of corporeal affections. The following examples drawn from the reference corpora illustrate how different 19th-century translators of Thucydides and Plato used sign and symptom in their English translations. The pair of passages is from Thucydides’ History of the Peloponnesian War (accompanied by literal English translations by the present author), and is followed by a selection of 19th-century and early 20th-century translations.

1. ἐγὼ δὲ οὖν τε ἐγόντες ἔλεξα, καὶ φήμ νός ἅ τι σκοπέων, εἰ ποτε καὶ ἀδική ἐπισέξου, μάλιστ’ ἂν ἔξοι τι προειδοῖς [I, for my part however, shall speak of what sort of thing it is, and from what things, someone seeing them, should it happen again sometime, should have more knowledge beforehand …]

a. For my own part, I shall merely relate the manner of it; and, having been myself sick of it, and seen others afflicted, I shall point out those symptoms of the

![Fig. 1 Normalised frequency of SIGN and SYMPTOM in Jones and Adams, per 10,000 words.](image-url)
early 20th-century translations: translations by the present author) is drawn from Greek passages

Table 7 Comparison of raw (RF) and normalised frequencies (NF) of SĒMEION/SĒMEIA in Greek core corpus (all cases) with SIGN/SIGNS and SYMPTOM/SYMPTOMS in Hippocrates translations by William Jones and Francis Adams.

|            | RF(sign) + RF (symptom) | RF(sēmeion) in Greek core corpus | NF(sign) + NF (symptom) | NF(sēmeion) in Greek core corpus | Effect-size |
|------------|--------------------------|----------------------------------|--------------------------|----------------------------------|-------------|
| Jones      |                          |                                  |                          |                                  |             |
| Singular   | 46                       | 53                               | 8.0                      | 5.2                              | 0.63        |
| Plural     | 149                      | 51                               | 26.0                     | 5.0                              | 2.38        |
| Adams      |                          |                                  |                          |                                  |             |
| Singular   | 60                       | 53                               | 8.0                      | 5.2                              | 0.18        |
| Plural     | 159                      | 51                               | 21.2                     | 5.0                              | 1.64        |

See Table 3 for effect-size parameters.

malady, from a consideration of which any one may have some previous knowledge of it … (Bloomfield, The History of Thucydides, 1839).
b. But I shall describe its actual course, and the symptoms by which any one who knows them beforehand may recognise the disorder … (Jowett, Thucydides, 1881).

2. τὸ μὲν οὖν νόσημα, πολλὰ καὶ ἄλλα παραλπόντα ἄτοπιας, ὡς ἐκάστως ἐτύγχανε τι διαφεροῦντος ἐτέρω πρὸς ἐτέρον γιγνόμενον, τοιούτον ἦν ἐπὶ πάν τινας ἰδέας. [Indeed, the disease, putting aside many other unusual things, for each thing having happened, something differing from it coming about, is this way in general.]

a. Such, then, was the general nature of the disease; for I pass over many of the unusual symptoms, since it chanced to affect one man differently as compared with another (Forster Smith, History of the Peloponnesian War, 1919).
b. Such was the general nature of the disease: I omit many strange peculiarities which characterised individual cases (Jowett, Thucydides, 1881).

And this pair of examples (accompanied by literal English translations by the present author) is drawn from Greek passages from Plato, and is followed by a selection of 19th-century and early 20th-century translations:

3. οὔτω δὲ καθάπερ τὰ πράγματα τὰ μὲν ἄλλοις ἠμοτπεν, τὰ δ’ οὖ, καὶ περὶ τὰ τῆς φωνῆς ἀν’ ημεία τὰ μὲν οὐχ ἠμότπεσθαι [So, just as things join together with one another, and other things do not, in the case of signs from utterances these do not join together, yet [others] join together …]

a. So, then, just as of things some fit each other and some do not, too some vocal signs do not fit… (Fowler, Sophist, 1921).
b. And as there are some things which fit one another, and other things which do not fit, so there are some vocal signs which do, and other which do not combine and form discourse (Jowett, Sophist, 1871).

4. … ἓ δὲ τοῦ σώματος νοσοτριφία ἀπέργυγος αὐτὸν τῶν πολιτικῶν κατέχει, τὸ δ’ ἠμέτρον οὕτι ἔτερον λέγεν, τὸ δαμαύουν ημείαν: ἢ γὰρ ποῦ τινι ἄλλω … […] but caring for his sick body kept him away from engaging himself with city affairs. Our divine indication hardly merits mentioning, for it has not come about in others anywhere …]

a. … is kept in check by ill-health, which excludes him from a public life, though in all other respects he has every inducement to desert philosophy. I need not mention the supernatural sign, which restrains me; for I fancy it has been granted to few … (Davies and Vaughan, The Republic of Plato, 1852)

b. … but ill-health kept him away from politics. My own case of the internal sign is hardly worth mentioning, for rarely, if ever, has such a monitor been given to any other man (Jowett, The Republic, 1871).

In examples 1 and 2 Thucydides gives a first-hand account of the plague of Athens (Craik, 2001). There are no semiotic terms in the Greek text, but, in the medical context, several translators use symptom to explicate the passage. In Plato we see that sēmeion tends to be translated with sign, which can be explained in part by the philosophical rather than medical context of these passages. Together these examples illustrate the fact that the choice of semiotic term was dictated in large part by context or even genre, with symptom being the term suited for medical registers in histories, and sign for philosophical discussions, as in Plato’s dialogues.

I conclude that the magnitude of the difference between the frequency of symptom and sign in the English core corpus and sēmeion in the Greek core corpus does not reflect a high frequency of sumptôme in the Hippocratic Greek, nor does it result from a straightforward, one-to-one lexical correspondence between sēmeion in Greek and symptom or sign in English translations. Rather, I would argue, it suggests that the prevalence of symptom, particularly in light of the less significant role played by sign, in the English core corpus reflects how Adams and Jones interpret and translate Hippocratic texts that deal with issues involving semiosis.

These findings are corroborated by collocational analysis, which I turn to now. In at least one respect symptom and sign seem to be used as virtual synonyms in the English core corpus. On the other hand, when sign and symptom collocate together, and when they form verbal collocations, differences between them emerge. I begin by considering adjectival collocates of sign and symptom, and then instances where sign and symptom collocate with each other, and, finally, verbal collocates.

Tables 8 and 9 show how some adjectives collocate with sign and symptom, respectively.

Tables 8 and 9 indicate that sign and symptom attract a similar set of modifiers. When either occurs there is a fair likelihood that one of these adjectives will occur in proximity to it. This is not the case in the Greek, however. As we observe in Table 10, of the more than 100 instances of sēmeion in the Greek core corpus a small number of the rough Greek equivalents of the above modifiers collocate on the left or the right as adjectival modifiers or predicates in a statistically significant way.

The reasons for this have to do in part with the syntax of ancient Greek, and in part with the way the Hippocratic authors wrote. Unlike in your typical English sentence, it is quite normal
Table 8 Adjective collocates of keyword SIGN, raw frequency = 66.

| Collocate | Raw frequency of collocation | Raw frequency of collocate | M13 |
|-----------|------------------------------|---------------------------|-----|
| bad       | 9                            | 231                       | 12.6|
| favourable | 6                            | 58                        | 12.9|
| deadly    | 3                            | 28                        | 10.9|
| critical  | 4                            | 41                        | 11.6|
| good      | 7                            | 163                       | 12.0|
| fatal     | 3                            | 98                        | 9.1 |
| mortal    | 2                            | 23                        | 9.4 |
| dangerous | 2                            | 64                        | 8.0 |

See Table 4 for collocation parameters; metric id: ENG ADJ.

Table 9 Adjective collocates of keyword SYMPTOM, raw frequency = 348.

| Collocate | Raw frequency of collocation | Raw frequency of collocate | M13 |
|-----------|------------------------------|---------------------------|-----|
| bad       | 22                           | 231                       | 14.1|
| critical  | 3                            | 41                        | 8.0 |
| dangerous | 6                            | 64                        | 10.3|
| deadly    | 8                            | 28                        | 12.8|
| fatal     | 11                           | 98                        | 12.3|
| favourable | 10                           | 58                        | 12.7|
| good      | 11                           | 163                       | 11.6|
| mortal    | 10                           | 23                        | 14.0|
| salutary  | 4                            | 7                         | 11.8|
| unpleasant| 2                            | 5                         | 9.2 |
| worst     | 5                            | 41                        | 10.2|

See Table 4 for collocation parameters; metric id: ENG ADJ.

Table 10 Collocates of keyword SÊMEION (all cases), raw frequency = 104.

| Collocate | Raw frequency of collocation | Raw frequency of collocate | M13 |
|-----------|------------------------------|---------------------------|-----|
| kokos (bad) | 10                           | 314                       | 8.3 |
| chrêstos (favourable) | 5                            | 36                        | 8.4 |
| thanatôdes (deadly) | 4                            | 51                        | 6.9 |
| krisimos (critical) | 1                            | 34                        | 1.5 |
| periestikos (indicating recovery, opp. thanatôdes) | 1                            | 8                         | 3.6 |
| phlauros (unfavourable) | 35                           | 2                         | 4.5 |

See Table 4 for collocation parameters; metric id: GR ADJ. Collocations indicated in bold fall below minimum threshold of statistical significance.

for adjective modifiers in the Hippocratic corpus to be separated from the noun they modify. Consider the following passage from the Coan Prenotions.

Αἱ προαληθήσασι τρόπων χολερῶδες πρὸ τῶν τόκων, τίτκουσι μὲν ρημίως, παρέξασι δὲ, κακοπρόσωπας, ἄλλως τε κήν τι κατὰ φάρμαγα όχλη, ὡς τι τῶν ἐν πολεμίῳ κακοπροσώπων ἐπιφανεῖς σημαίων.

[Women who feel pain prior to giving birth in a choleric way give birth easily, but then become feverish and are in a malignant [condition], otherwise something in the throat troubles them, especially if there is something troubling the throat or a certain of the malignant indicators of fever appear.]

In this sentence a number of words intervene between the definite article τὸν (’the’), the modified noun, sêmeion (‘indicators’), and the adjective, kakaôethôn (’malignant’). Hence, our association metrics for adjective construction have to be calculated for a larger window than their English counterparts (compare the Span reported in Tables 3 and 4). This larger window allows us to pick up all the adjectives associated with sêmeion, but the magnitude of the association is smaller than in English. In relation to the way Hippocratic authors wrote, on the other hand, a sentence such as the one above or the following from book six of the Hippocratic Epidemics, with both the semiotic term and the adjective, are relatively infrequent:

Σημεία θανατώδεα, ἀνὰ ρινῶν θερμῶς ἀτριῶς· πρὸτερον δὲ ρίς ψυχρὸν πνεύμα ἀφίησιν· τὰ ζωτικὰ ἔναντι.

Deadly indicators, hot vapours [rising] upward from the skin. Before, the nose discharges cold pneuma. Vital [indicators, sc. sêmeia?], the opposite.

What is more usual is how zôïka (vital) is used in the above passage and in the following example from the Hippocratic Aphorisms. In these cases, the adjective is used as a substantive, or the noun it modifies has been omitted for brevity. In English, the adjective is awkward without a noun for it to modify. In the case of ta zôïka, from the fact that it is a neuter nominative plural, it is natural to conclude that it is modifying an omitted neuter nominative plural sêmeia. And Jones and Adams do not fail to draw this conclusion.

Ὡς ἀρα τόιοι περιετίες τεταρτάοισσό τιούν αἵμα ἐκ τῶν ρινῶν ρηθι, ποινρῆν. When blood flows from the nose in those who have quartan fevers, it is [something] bad.

Which Jones translates as

When in patients suffering from quartan fevers there is bleeding at the nose, it is a bad symptom.

And Adams

When in quartan fevers blood flows from the nostrils it is a bad symptom.

The fact that sêmeion is a neuter singular strongly suggests to the reader that ponêron is a neuter singular adjective modifying an omitted sêmeion. Hence, the well-known brevity of Hippocratic authors encourages Adams and Jones to supply unstated nouns for the unaccompanied adjectives. Moreover, their assumption that a Hippocratic semiotic medical discourse is a central element in Hippocrates’ thinking over-determines the interpretation of these ambiguous passages in favour of a semi-otically oriented reading and explication of the text.

We also observe that in the English core corpus sign and symptom frequently collocate together within a fairly restricted window (M13 = 9.6, L5–R5). In these instances, symptom tends to be used as the more generic term for medical indicators, whereas sign is typically used to refer to particular subcategories of symptoms. In a minority of cases, sign appears to refer in a more general way to medical indicators than symptom. For example, in the Hippocratic Epidemics 1.11, the author states:

πεπασμὸν ταχριῶν θρίσσος καὶ ἀφάδελφον ὑγίης σημαίνουν, ἀμα δὲ καὶ ἀπέπτα καὶ ἐς κακᾶς ἀποστάσις
Concoctions signify that the crisis is swift and that recovery is certain, whereas the raw and the unconcocted that change into bad swellings [signify] non-crisis, pain, time, death, or a relapse of them. One must consider from among others which one of them is most of all.

Which Jones, for example, translates as

Coctions signify (sêmainōsin) nearness of crisis and sure recovery of health, but crude and unconcocted evacuations, which change into bad absorptions, denote absence of crisis, pain, prolonged illness, death, or a return of the same symptoms. But it is by a consideration of other signs that one must decide which of these results will be most likely.

In this passage, the use of sign and symptom likely results from Jones’ desire to explicate an ambiguous Greek text, where, in this instance, signs appears to be the more generic term. The collocation of the terms seems to be precipitated by the Hippocratic author’s distinction in the passage between specific medical indicators, here called symptoms, mentioned early in the passage, and other types of indicators, here called signs, that go unmentioned.

When sign and symptom collocate it is not typical, however, for sign to have the more generic sense. In fact, it is more common for symptom to serve as the more generic term. Concordance 3a shows first the Greek text accompanied by a literal English translation; Concordance 3b presents Jones’ or Adams’ translations of the same Greek text.

Concordance 3a. SIGN and SYMPTOM with my literal translation of the Greek. The Greek text in lines 3 and 5 are the same. Adams and Jones translate them differently as seen below.

1. ἢ γάρ ὑπετέκτως φαίνεται φέρον τὸ κακὸν ἢ καὶ ἄλλο τι τῶν περιεστικῶν σημείων πρὸς τόουτον ἐπιδεικνύο, τὸ νόσημα
   [For if he [the patient] shows he is bearing the illness with ease, or if he displays besides some other indicators that signal recovery, the disease …]

2. μελαιόνωμεν δὲ παντελῶς οἱ δάκτυλοι καὶ οἱ πόδε ἴσον ἀληθρὸν τῶν πελενδῶν εἰσὶν ἄλλα καὶ τάλλα σημεία σκέπτεθαι χρή· ἢ γάρ
   […the fingers and toes becoming completely black is less deadly than their being white, and it is necessary to consider the other indicators, for if …]

3. εἰ μέντοι χρῆ εἰδέναι περὶ τῶν τεκμηρίων καὶ τῶν ἄλλων σημείων, ὅτι εν παντὶ ἔτει καὶ πάση χρή
   [It is necessary to know well, however, about the proofs and the other indicators that are in every year and all places …]

4. τα τε κακὰ κακὸν τι σημαίνει καὶ τὰ χρήστα γάγαθον, ἐπί καὶ ἐν Λίμη καὶ ἐν Δῆλῳ καὶ ἐν Σκύθη καὶ σήμεια προγραμμάτων σημεία ἀληθεύεται
   [The bad indicates something bad, and the good, something beneficial, since the aforementioned indicators are true in Libya, Delos, and Scythia.]

5. εἰ μέντοι χρῆ εἰδέναι περὶ τῶν τεκμηρίων καὶ τῶν ἄλλων σημείων, ὅτι εν παντὶ ἔτει καὶ πάση χρή
   [It is necessary to know well, however, about the proofs and the other indicators that are in every year and all places …]

Concordance 3b. Collocation of SIGN and SYMPTOM.

1. For if the patient should show himself bearing up against the illness, or manifest, in addition to the signs mentioned before, some other symptom indicating recovery, the illness

2. but when fingers or feet become quite black it is a less fatal sign than their becoming livid. But the other symptoms also must be attended to. For if

3. However, one must clearly realise about sure signs and about symptoms generally, that in every year and in every land

4. signs indicate something bad, and good signs something favourable, since the symptoms described above prove to have the same significance in Libya, Delow and Scythia.

5. One should likewise be well acquainted with the particular signs and the other symptoms, and not be ignorant how that, in every year, and at every season

By comparing the Greek and English concordance lines we may draw a few conclusions. Lines 3 and 5 suggest that the collocation of sign and symptom was precipitated by the collocation of sêmeion and tekmerión (‘sign’ or ‘proof’) in the Greek. In this case both translators selected sign to render tekmerión. Jones renders tôn tekmerión as ‘sure signs’ and tôn allôn sêmeión as ‘symptoms generally’; Adams renders tôn tekmerión as ‘particular signs’ and tôn allôn sêmeión as ‘other symptoms’. In Line 1 sign is added by Jones to distinguish between the specific disease indicators that signal recovery mentioned earlier in the paragraph from other kinds of indicators that indicate recovery from the disease. Likewise, in Lines 2, Jones uses sign and symptom to distinguish between specific medical indicators mentioned earlier in the passage from ‘the other indicators that are necessary to examine’. Finally, in Line 4, sign is used to explicate the Greek phrase, which literally says ‘and the beneficial [indicator indicates] the good’, whereas symptom is used to express a universal medical principle, namely, that wherever the ‘symptoms (sêmeia)’ he has spoken about in the treatise come to be, even in places as remote as Delos, Scythia and Libya, the prognoses they supply the physician with come true.

Thus, when symptom and sign collocate our translators seem to use them with different ends in mind. This impression is corroborated when we examine how sign and symptom collocate with verbs. Verbal collocations suggest that, somewhat different from the case of adjectives discussed above, Adams and Jones treated them as performing different semantic roles. We note that symptom shows a statistically significant tendency to collocate with what we might call medical cause and effect verbs and medical cognition verbs, whereas sign does not. When we compare the association metrics in Table 11a, b, we see that the lexical patterns that sign and symptom attract diverge considerably in relation to verbs of cause and effect.

Even when we take into account the different frequencies of sign and symptom in the English core corpus, Table 11b shows that for Adams and Jones signs do not come about in the same, concrete, corporeal way that symptoms do. Focusing more closely on symptoms we observe that the relationship between them and cause and effect verbs is sometimes a loose one. In order to get a clear picture of how symptom collocates with these verbs, I decrease the window span to ±4. When this is done, clear trends emerge. The following series of graphs shows the frequency with which symptom occurs with respect to this set of cause-and-effect verbs. Figures 2–4 show that symptoms tend to occur, supervene, and appear, and these collocations happen in a regular way (one position to the left of the keyword) in the English core corpus. On the other hand, Figs. 5–7 show that symptom does not occupy a position to the left or right of attend, experience, and relieve in a regular way.

Thus, sign tends to collocate almost exclusively with adjectives such as good, bad, and fatal as it tends to be associated with the culmination of the prognostic process in Hippocratic medicine, when the physician draws a conclusion about the prognosis of the
disease based on the quantity and quality of symptoms. Drawing on a fund of past experience he knows that these corporeal indicators accompany the disease in a predictable way, and enable him to make a reliable prognosis.

Symptom collocates in two distinct ways. We saw that it shares many of the noun-modifier collocates with sign (Tables 9 and 10), but unlike sign it also shows a tendency to collocate with verbs that indicate that symptoms are phenomena that unfold over time and whose existence accompanies or somehow depends on the existence of other phenomena (Table 11a; b; Figs. 2–4). Symptoms are parts of or elements in the unfolding of the disease rather than cognitive elements in the process of making a prognosis. Of course, this distinction between symptom and sign is not Hippocratic. It is, rather, imposed on Hippocratic medical discourse by our translators. For example, Concordances 4a, b present the concordance lines for parallel English and Greek passages:

Concordance 4a. Concordance lines from English core corpus.

1. So pay attention to the other symptoms also, since, if another bad symptom supervene, the case is hopeless; (Jones)
2. Nor was there haemorrhage from the nose in any case when these symptoms occurred, but only slight epistaxis. (Jones)
3. These **symptoms**, appearing at the commencement, are indicative of strong delirium, and patients so affected generally die … (Adams)
4. These **symptoms** also occur in the site of the diaphragm, but much less frequently; for the diaphragm is a broad, expanded … (Adams)
5. And these **symptoms** occur sometimes on the left side, sometimes on the right, and sometimes on both. (Adams)

**Concordance 4b.** Parallel Greek text from Greek core corpus with literal English translation by the present author.

1. προσέχειν οὖν δέ τού τόν κόσμον καὶ τοίς ἄλλοις σημείοισιν, ὡς ἢν τι καὶ τῶν ἄλλων σημείων πονηρόν ἐπιφανῆται, ἀνέλπιστος ὁ ἀνθρώπος.

   [It is necessary, therefore, to give mind to the other indicators also, for should the bad consequences of the other indicators appear, the person has no hope.]
2. οὐδὲ ἡμορραγήσεις ἐκ τῶν ὤνδει, οὐκ ἔχει ταύτα συμπίπτοι, ἀλλ’ ἢ σμικρὰ ἔστασεν.

   [Nor were there any bloody discharges from the noses when these happened, being but a small trickle.]
3. Ταύτα δ’ ἐν ἄρχῃν ἐπιφανώμενα παραφροσύνας δηλωτικά ἐστι σφοδρῆς, καὶ ἐπί τὸ πόλον θνησκοῦν.

   [These indicators of derangement that appear at the beginning are violent, and for the most part they die.]  
4. Γίγνεται δὲ καὶ ὑπὸ φρένοις ἴχνον, ἴσονν οὖν πολλόν διάτασις μὲν γὰρ φρένων πλατεία καὶ ἀντικείμενη, φόνος.

   [They [pains, sc. ὀξύναι] also appear to all be force, though a great deal less, below the belly, since the diaphragm is a natural formation that is wide and offers resistance …]
5. διὸ καὶ οкажет κατὰ τῷ τότε τῆς κεφαλῆς ἀλγεί, ἀλλ’ ἐν μέρες ἐκέρατος, ἢτα ἐπισύνας.

   […] for which reason pain is not always felt in the same part of the head, but in part of one or the other, and at times all over.]

In Line 1 **symptom** is a straightforward translation of **sēmeion**. In Lines 2–4 Adams and Jones use **symptom** to explicate ambiguous definite articles and supply the previously stated subject of the verb **γίγνεται**. And in Line 5 nothing in the Greek directly corresponds to these symptoms. Concordances 4a, b also show that English verbal forms such as **occur**, **supervene**, and **appear** correspond to a variety of verbs, such as **epiphainó** (’come into view’), **gignomai** (’come into existence’), **sumpinti** (’to co-occur’), and **algé** (’to experience [pain]’, ’suffer’). Moreover, **symptoms** has a strong connexion to physical processes, coming about with ’all force’, appearing at the beginning in a ’violent’ way, being accompanied by ’bloody discharges’, and situated in the region around the viscera, and in the head.

In addition to cause-and-effect verbs, **sign** and **symptom** collocate in noteworthy ways with cognition verbs, the second of the classes of verbs mentioned above. According to Jouanna (2013, p. xlv; Manetti, 1993), medical indicators, **sēmeia**, play an important role in a prognostic ’calculus’ that enables the physician to form a judgement about whether, when, and even how the patient will recover from their illness. Noting that these judgements take on a fairly regular syntactical form (Jouanna, 2013, pp. iii–lvi; Roselli, 1990), Jouanna points to other verbs used in the Hippocratic Prognostic relating to cognitive aspects of determining, judging, knowing, and understanding disease. Other items, however, gather around verbs having to do with medical epistemology. Jouanna (2013, p. xlvii) also mentions the verb **sullogizomai** (to calculate together) as an activity that is involved in assimilating the evidence from various indicators in order to arrive at a prognosis. Yet, in this instance, the Hippocratic author does not speak about calculating on the basis of a complete set of indicators (**sēmeia**), but, more generically, on the basis of ’the good as well as the bad things that follow upon (τά δὲ ἐπιγενόμενα ἀγάθα τε καὶ κακὰ) disease’.

We learn, therefore, that while **sēmeion** is, in fact, involved with verbs referring to acquiring knowledge about a disease, it is not the only word that is used in such contexts. Indeed, Table 12 summarises how **sēmeion** clusters around medical cognition verb stems that occur frequently in the Greek core corpus, and shows that this clustering appears to be relatively sparse and difficult to assess quantitatively.

The association metric values presented in Table 12 are small, and suggest that we cannot be confident that **sēmeion** forms a genuine collocation with any of these Greek verb stems. Looking closer at the two verbs that appear to be most associated with **sēmeion**, namely **krinó** (roughly meaning to judge) and **sképë** (roughly meaning to view, examine, consider), yields ambiguous results. **Krínó** occurs with **sēmeion** in the Hippocratic Prognostic, the Coan Preations, and the Epidemics. In the latter, for example, the Hippocratic author says:

Αἱ ἑτοὶ τῆς νοσόσαις ἀποστάσεις εἰ κρίνουσιν, σημείον ἢ πυρόδεια ἔνθαν, μὴ παρεταίνουσι, καὶ δύσφορα ἔνθαν, εὐθυρώς φέροσιν…

Abscesses in diseases, if they reach the crisis [εἰ κρίνουσιν], are an indication [sēmeia], should they be fierce, that they [patients] do not become feverish, and should they be hard on them, that they bear it well.…. In this passage **krinó** means to reach a medical crisis, which in Hippocratic medical theory has a cognitive element. It is the critical period in the illness during which the disease, having matured, takes its most intense form. Should he read the patient’s
indicators correctly, it is at this stage, too, that the skilled physician is able to determine whether the patient will survive their illness.

On the other hand, all co-occurrences of σήμειον and σχέδον appear in the Hippocratic Prognostic, which suggests that these instances cannot be taken as representative of patterning in the Greek core corpus as a whole, but only of patterning in this text. Hence, the evidence from the Greek core corpus for the claim that σήμειον tends to cluster around medical cognition verb stems is inconclusive. If there is a Hippocratic medical semiotic theory in the Greek texts, the evidence reviewed above suggests that it is not explicit in the Hippocratic texts themselves; or that such a theory is limited to a small number of treatises, such as the Hippocratic Prognostic, but is not necessarily representative of discursive patterning in the rest of the Greek core corpus.

Comparing these results from the Greek core corpus with a similar collocation analysis in the English core corpus yields telling results. Table 13a, b indicate that in the English core corpus both sign and symptom collocate with cognition verbs in ways that suggests that these indicators are crucial to the cognitive process of making a prognosis and diagnosis.

Sign collocates significantly with indicate and determine, whereas symptom attracts a larger variety of cognition verbs. Symptom collocates significantly with indicate, know, determine, and expect in the English core corpus, as illustrated in Concordance 5.

Concordance 5. SIGN and SYMPTOM collocating with cognition verbs.

1. ... according as they appear along with favourable or unfavourable symptoms, indicate diseases of a short or long duration. (Adams)
2. And whenever they interpret symptoms with a view to determining the right method ... (Jones)
3. danger of dying; when, therefore, you perceive these symptoms, announce this prognostic, if you shall judge proper, after making the suitable observations. (Adams)
4. where he had pain formerly, for these symptoms occur in the commencement of suppurations. One then may expect the rupture of the abscesses to take place from ... (Adams)
5. I know of no woman who died if any of these symptoms showed themselves properly, but all to my knowledge had abortions if they chanced to fall ill when with child. (Jones)

This result provides further evidence that the frequency and collocation behaviour of symptom and sign in the English core corpus do not straightforwardly reflect the collocation patterns in the Greek source text. More importantly, the fact that symptom is found occurring repeatedly throughout the English core corpus taken together with the frequency and variety of collocations of symptom with this set of verbs suggest to the reader who is unfamiliar with the Greek source text that there is a Hippocratic semiotic medical discourse in which semiosis is strongly linked to prognostic and diagnostic medical knowledge. In actual fact the discursive patterns that constitute the evidence for the existence of a Hippocratic semiotic discourse have been largely introduced by the English translators. The Hippocratic Prognostic is a good example of how the process of translation into English supplies a semiotic framework for the Hippocratic text where the Greek source-text is theoretically underdetermined. In the following passage, the author speaks about what the patient’s hands (cheires) and breathing can tell us about the patient’s prospects for recovery.

Πνεύμα δὲ πυκνόν μὲν ἐνώ πόνον σημαίνει ή φλεγμονήν ἑν τοίνυν ὑπὲρ τῶν φρενῶν χωρίσων· μέγα δὲ ἀναπνεοῦν καὶ διὰ πολλοῦ χρόνου παραιροσύνην σημαίνει· ψυχρὸν δὲ ἐκπνεούμενον ἑκ τῶν πόνων καὶ τοῦ στόματος ὀλέθριον κάρτα ἢ δὴ γίνεται.

On the other hand, rapid breathing indicates that there is pain or inflammation in the region above the belly. Breathing deeply with long intervals in between indicates delirium. Exhaling what is cold from the nose and mouth is extremely deadly as well.

It is evident from the Greek that no semiotic terms are present, though the verb σημαίνω is used to speak about how an assortment of gestures and conditions tell us about what the disease is, its severity, and the patient’s prospects for recovery. Jones twice introduces sign into this passage, concluding the first paragraph with “all these signs are bad” and exhaling cold air is a “very fatal sign” in the second; in this same passage, Adams says “all such symptoms are bad” and that exhaling cold air is a “very fatal symptom”. By supplying sign and symptom to explicate the Hippocratic text, our translators create a medical semiotic for Hippocrates by, in effect, saturating the text with English semiotic terms. The translators’ decisions and interpretations are, however, well-founded, resting on the lexicographical history of medical semiotic language in English since the 15th century. Moreover, the expectation that Hippocrates must be speaking about signs and symptoms in a way analogous to modern categories also relies on a narrative about the Western medical tradition that sees Hippocrates standing at its head.

Conclusion
In this article I have shown how corpora can be used to give new insights into translation’s transformative role in classical reception. Focusing on Francis Adams’ and William Jones’ Hippocratic translations, I have illustrated how combining a number of historical Greek and English corpora allows us to understand how a Hippocratic semiotic discourse emerged in English in the 19th and early 20th centuries. We have seen that, owing to the shrinking terminological options available to them since at least the middle of the 19th century, both translators interpreted the source-text in ways that took for granted the centrality of a
Hippocratic semiotic theory in the source-text, and that sign and symptom were the most suitable terms to convey Hippocrates’ semiotic discussions to English-speaking audiences. Careful comparisons of word frequency in the Greek and English core corpora with their respective reference corpora showed how the translators over-determined the semiotic import of ambiguous passages. Collocation analysis revealed differences in verbal collocations between the Greek and English core corpora, and showed how the translators linked knowledge with semiotic terms more strongly in English than in the Greek, and introduced distinctions between symptoms and signs into Hippocratic thought that could not exist in the source-text. Translation therefore played an active role in determining how these translations shaped how physicians in the British Isles who consulted these translations understood Hippocrates’ legacy and his continued relevance to their medical practice.

The ramifications of this study stretch beyond the disciplines of classical reception, translation studies, and the history of medicine. By focusing on discourse rather than translation strategies, this study genealogically links Adams’ and Jones’ Hippocrates translations to 19th-century developments in science, literature, medicine, and classics that transformed how 19th-century scholars perceived the relationship between body, knowledge, and disease. This study also details how physicians and scholars read this body–knowledge–disease relationship into the history of western medicine and its foundational texts.

Finally, this study also presents a methodology for harnessing the tools of corpus linguistics for genealogical analysis of discourse. I reiterate that this, and the other contributions in this Genealogies of Knowledge special issue, engage in genealogical research that is broadly in line with Foucauldian ideas about genealogy in the humanities and social sciences. Yet, the methods I described above differ in important ways from Foucault’s approach to genealogy. In outlining a detailed methodology derived from corpus linguistics, this study evinces a stronger commitment to a particular methodology than Foucault did. Moreover this study’s statistical orientation and its reliance on quantitative methods is bolstered by a faith in empirical methods that Foucault did not share. Despite these differences, however, I expect that its conclusions are nevertheless consistent with and contribute to genealogical research in classics and the medical humanities. In particular I discussed the role of translation, as a socially mediated practice, in the emergence of a semiotic discourse with ancient and modern lexical features. Further, I clearly delineated the properties of core and reference corpora, the sampling frames for building the corpora, the criteria of validity for comparisons between core and reference corpora, the ways of assessing keyness and collocation strength, and the influence of trends in English translations of the classics and English medical lexicography before and during the 19th century. Taken together, I suggest that these elements supply genealogical research with a method that may allow us to bring classics and medical translation into the centre of continuing debates about the emergence of the modern subject. The insights drawn from this case-study suggest the combination of genealogical research and corpus linguistics methods is a promising avenue for future research.

Data availability

The Greek Hippocratic texts used in this study are available to the public under a Creative Commons license at A Digital Corpus for Graeco-Arabic Studies: https://www.graeco-arabic-studies.org/texts.html. The English translations by Jones and Adams are available from the Perseus Digital Library’s digital repository hosted on GitHub: https://github.com/PerseusDL/canonical-greekLit/tree/master/data/tlg0627. The data sets for linguistic analysis are available upon request. The searches, word-frequency lists, and concordances can be reproduced using the Genealogies of Knowledge concordancer, available for download from the modnlp repository on Source Forge: https://sourceforge.net/projects/modnlp/; or from the Genealogies of Knowledge website: http://genealogiesofknowledge.net/software/.

Notes

1 Along similar lines, the Genealogies of Knowledge English print subcorpus has been used to understand the reception of Greek political ideas in modern English translations of ancient Greek authors as in part a product of the social, political, and cultural conditions of Britain in the 19th century (Jones, 2019a, 2019b).
2 Foucault (1984a) provides clearest exposition of the principle assumptions, import, and ramifications of genealogy as an approach to the study of truth, moral value, and the human subject.
3 The Genealogies of Knowledge concordance browser provides a modular architecture and tools for natural language processing and visualisation: https://sourceforge.net/projects/modnlp/.
4 The Genealogies of Knowledge browser builds on software designed for the Translational English Corpus Project (https://www.alc.manchester.ac.uk/translation-and-intercultural-studies/research/projects/translational-english-corpus-tee/) (Luz, 2011). What counts as a token in the Genealogies of Knowledge browser depends on the language of the subcorpus under study (English, Latin, Arabic, Greek). In the English subcorpora, blanks (including tabulation spaces, line breaks, etc.), hyphens and punctuation characters are treated as word delimiters. A token is defined as a string of characters delimited by any of these word delimiters (Luz, 2011, p. 136).
5 The aim of the study is to focus on the cumulative effects of these translation on lexical patterning in English. Nevertheless, the analysis presented below frequently examines how certain characteristic lexical patterns distinguish Jones and Adams.
6 For a complete list of the Greek texts included in the Genealogies Greek subcorpus: http://genealogiesofknowledge.net/corpora/greek-corpus/; for the modern English: http://genealogiesofknowledge.net/corpora/modern-english-corpus/.
7 The Genealogies English subcorpus contains many translations of classical texts printed after 1945. These translations represent scholarship and attitudes towards translation and classics one or two generation after Jones and Adams.
8 For studies that adopt a similar approach see, for example, the work on climate change discourses in English, French, and German media in Grundmann and Krishnamurthy (2010).
9 For a complete list of the Greek texts included in the Genealogies Greek subcorpus: http://genealogiesofknowledge.net/corpora/greek-corpus/; for the modern English: http://genealogiesofknowledge.net/corpora/modern-english-corpus/.
10 The Genealogies of Knowledge software does not at this time support parts-of-speech tags in any of the subcorpora. On POS-tags and future development of the Genealogies of Knowledge concordancing software, see Luz and Sheehan (2020).
11 Of course, there are many other medical cognition verb stems whose frequency in the Greek core corpus was too small to be included in the quantitative analysis. For example, the verb stems ζετέω, επιστρέφω, συντάσσω, επιτελέω, έξυπνώ, έξυπνώζω, and νεώτερος were omitted from consideration because they occur in the subcorpus infrequently.
12 Italics added.

References

Adams F (1844) The seven books of Paul Aegineta. The Sydenham Society, London
Adams F (1849) The genuine works of Hippocrates. The Sydenham Society, London
Adams F (1856) The extant works of Aretaeus, the Cappadocian. The Sydenham Society, London
Baker M (2020) Rehumanizing the migrant: the translated past as a resource for refashioning the contemporary discourse of the (radical) left. Palgrave Commun 6(12):1–16
Baker P (2006) Corpora in discourse analysis. Bloomsbury, London
Brezina V (2018a) Statistical choices in corpus-based discourse analysis. In: Taylor C, Marchi A (eds) Corpus approaches to discourse: a critical review. Routledge, London, pp. 258–280
Brezina V (2018b) Statistics in corpus linguistics: a practical guide. Cambridge University Press, Cambridge
