Patrick Hanks

Lexicography and technology in the Renaissance and now

Abstract

The development of lexicography has always been closely integrated with and dependent on technological developments. In this paper, I draw attention to some parallels between lexicography in the Renaissance and the present day.

In 16th-century Europe, the character of lexicography underwent a radical change as a result of two inventions: printing from movable type and metal type-founding. It was also profoundly affected by the so-called 'rebirth of learning'. Part 1 of this paper discusses the technology and philosophy of Renaissance lexicography.

In the 21st century, computer technology and developments in scholarship are beginning to have a comparable radical effect on lexicography. It is too early to say where this will lead, but one thing is already certain: traditional printed dictionaries, after five hundred years of unparalleled success, are a thing of the past. The business model according to which a publisher pours huge sums of money over many years into developing a dictionary in the hope of reaping rich rewards from sales of the dictionary as a printed book each year is dead. Dictionary users nowadays, seeking information about the spelling, use, or meaning of a word, do not turn to a printed book: they turn instead to a web site. Whether they are well served or ill served by such web sites is a matter for debate.

At the same time, there has been a revolution in cognitive and philosophical understanding of the nature of language, and in particular the role of lexis, which lexicographers have been slow to come to terms with. Traditional Leibnizian models of word meaning and phraseology have been superseded by the work of philosophers such as Wittgenstein and Putnam, anthropologists such as Rosch and Tomasello, and linguists such as Sinclair.

A third revolution has taken place in the collection of lexicographical evidence. Traditional lexicographers relied on extremely time-consuming and inefficient reading programmes for the collection of evidence, or (following the lead of generative linguists) simply invented the evidence by introspection. It has been said that linguistics of this kind is the only 'scientific' discipline in which it is regarded as acceptable first to invent evidence, then to explain what has been invented, then to claim that something of general validity has been 'discovered'. A modern alternative approach – empirical linguistics –, particularly relevant to lexicography, is to collect large quantities of data (e.g. an electronic corps of texts or speech) and investigate how people actually use words with the aid of various kinds of computational and statistical methods. Generative linguists, defending their use of introspection to create evidence, sometimes raise a counter-accusation that empirical linguistics has no role for introspection and intuitions. Some unthinking empirical linguists even agree. But this is self-evidently not true. There is a role for intuitions in empirical linguistics, namely the *interpretation* of data. The empiricists' true objection is to the invention of evidence.

In Part 2 of this paper, I present one of many possible new approaches to lexicography, namely Corpus Pattern Analysis (CPA). This picks up a theme promulgated by the late John Sinclair, that "many meanings depend for their realization on the presence of more than one word" (Sinclair 1998). It is a truism that meaning very often depends on context, yet much modern lexicography ignores context. And what is meant by 'context' anyway? I discuss some of the many problems that are lying in wait for lexical analysts who seek to adopt a phraseological approach to lexicography. Nevertheless, I shall conclude that lexicography needs to pay more attention to phraseology than is currently customary.

A third revolution has taken place in the collection of lexicographical evidence. Traditional lexicographers relied on unbelievably time-consuming reading programmes for the collection of evidence, or simply invented the evidence by introspection (navel-gazing). It has been said that linguistics (generative linguistics) is the only 'scientific' discipline in which it is regarded as acceptable first to invent evidence, then to explain what has been invented, then to claim that something of general validity has been 'discovered'. A modern alternative approach, empirical linguistics (including various approaches to lexical analysis), is to collect large quantities of data (e.g. an electronic corps of texts or speech), followed by explanation of what has been collected using various kinds of large-scale computational and statistical analysis. Generative linguists, defending their use of introspection to create evidence, sometimes claim that empirical linguistics has no role for introspection and intuitions. But this is self-evidently not true. There is a role for intuitions in empirical linguistics, namely the *interpretation* of data. The empiricists' objection is to the use of intuitions to invent evidence, not (of course) to the use of intuitions to interpret it.

1. Renaissance lexicography and printing technology

1.1 Dictionaries in England and Europe before Cawdrey

Surveys of English lexicography, starting with Murray (1900), give the impression that the first English dictionary was Robert Cawdrey's *Table Alphabeticall*, published in 1604. This little book is a dictionary of hard words, mostly 'inkhorn terms' – learned words that were introduced in profusion from Latin into English by scholars during the 16th century. Cawdrey's book is addressed mainly to women – who, in the 16th and 17th centuries were rarely fortunate enough to receive a Latin education and as a result tended to be excluded from intellectual discourse and in some cases could not even understand the Latinate English words (known as 'inkhorn terms') addressed to them in sermons. Apart from this, the *Table Alphabeticall* is a historical curiosity of comparatively little intellectual or cultural interest. It had no ambition to be a 'complete' inventory of the lexicon. The notion that a dictionary should serve as an inventory of the lexicon of a language was not an innovation of English lexicographers, but was a goal that had been pursued (for Latin) by several important lexicographical works in Continental Europe in the 16th century.

The prominence assigned to Cawdrey's *Table Alphabeticall* by Murray (1900) and by subsequent Anglocentric writers such as Starnes and Noyes (1946) had the unfortunate effect of deflecting attention from the rich lexicographic tradition of the European Renaissance in the 16th century, in which English was only one of several participant languages – a rather minor one, as we shall see. Starnes (1963) tried to correct the false impression given by his earlier work, but apparently in vain. Landau (1984, 2001) describes *Cawdrey's Table Alphabeticall* as a seminal work, adding that it is "the least inspiring of all seminal works". The *Table Alphabeticall* is indeed uninspiring, but it is not a seminal work.

The word ***dictionary*** itself came into English as an inkhorn term in the mid 16th century. The *Oxford English Dictionary* (OED) shows that the Medieval Latin word *dictionarium* was coined as early as 1225 and was used to denote a collection of Latin words arranged according to subject, rather than in alphabetical order. More exotic synonyms such as *glossarium* 'glossary', *cornucopia* 'horn of plenty', *elucidarius* 'elucidator', and *thesaurus* 'treasure house' also became widespread.

OED comments:

> Dictionaries proper are of two kinds: those in which the meanings of the words of one language or dialect are given in another (or, in a polyglot dictionary, in two or more languages), and those in which the words of a language are treated and illustrated in this language itself. The former were the earlier. (OED second edition, s.v. *dictionary*)

So what were these Renaissance dictionaries before Cawdrey? What did they consist of, how and where did they originate, who compiled them, and what was their purpose?

Scholarly studies by Starnes/Talbert (1955), Starnes (1963), Considine (2008), and an excellent chapter by Bately (2009) in Cowie's monumental *Oxford History of English Lexicography* have gone some way towards correcting the misleading impression perpetuated by Landau and others. Bately shows how lexicography developed as a scholarly and cultural activity during the 16th century. She observes that lexicographers both of Latin-English dictionaries and of other foreign language-English dictionaries turned to the continent for models and sources.

> So, when, in 1538, Thomas Elyot [...] produced his unidirectional Latin-English *Dictionary*, the authorities he cited included French, Dutch, and Italian contemporaries, who, like him, were seeking to provide the linguistic tools demanded by the 'New Learning'. It was the monolingual Latin *Dictionarium* of 'Calepinus' – Augustinian friar Ambrogio Calepino of Bergamo –, first published in 1502, that was his chief source. And when Elyot's dictionary was reissued in 1542 as the *Bibliotheca Eliotae – Eliotis librarie*, it was from the *Dictionarium Latino-Gallicum* (1538) of French printer Robert Estienne [...] that much of its new material was derived.

More will be said about Calepino below. And it should be noted here, at the outset, that Estienne was much more than a printer in the modern sense. He was a classical scholar, an editor, a publisher, and a Humanist thinker, conversing on equal terms with the leading Parisian intellectuals of his day.

1.2 The development of printing and typography

The development of lexicography in 16th-century Europe was dependent on the development of printing technology and the associated crafts of punch-cutting and typefounding. Dictionary-like compilations pre-dated printing, but dictionaries as products for widespread general use only became available because of the rapid reproduction of identical copies that printing made possible. Collections of words with glosses were created in monasteries as manuscripts throughout Europe in the later Middle Ages. Mostly, these works consisted of collections of Latin words glossed into vernacular languages, for the benefit of young novices learning to read Latin texts. Sometimes the Latin words arranged (more or less roughly) in alphabetical order, sometimes thematically. For propagation each manuscript had to be laboriously copied out by hand, and each act of copying could produce only one copy at a time, each with its own idiosyncrasies and copying errors. The invention of printing by Johannes Gutenberg in about 1440 in Strasbourg (subsequently moved to Mainz) changed everything, not only for lexicography but for all other fields of knowledge, as discussed by Eisenstein (1979). Suddenly, rapid replication and massive dissemination of identical copies of a text – including large and complex texts such as dictionaries – became possible.

Three components contributed to and are intertwined with the development of lexicography and with each other: 1) the invention of printing, 2) the rediscovery of classical Latin literature, philosophy, and art (including lettering), and 3) the development of challenging thinking that constituted the Reformation. The history of all these events has been intensively studied, but their interaction bears re-examination, for an understanding of it will crucially affect our appreciation of the early history of European lexicography. Let us first look at the relationship between printing and lexicography.

After Gutenberg, a key figure is Nicolas Jenson, a man of German extraction who was born in 1420 in Sommevoire, France (about half-way between Paris and Strasbourg). By the 1450s, Jenson had risen to become controller of the French royal mint at Tours. In 1458 he moved to Mainz, where he evidently became fascinated by the technology of printing with movable type, recognizing its potential for the rapid dissemination of knowledge. To this technology, he devoted the rest of his life. After a few years as a printer and publisher in Mainz and Frankfurt, Jenson moved to Italy, where, in Venice in 1468, he set up shop as a printer, publisher, and typographer. Between 1468 and his death in 1480 he edited and printed about 150 books, mostly editions of Latin theological tracts, but also some Latin classics, some Greek, an Italian guide to medicinal herbs, and miscellaneous other works. Jenson was not the only printer and typographer in Venice in the 1470s, but he is surely the most important of them.

Let us look a little more closely at his typographic principles, which were to play such an important role in the development of lexicography in subsequent decades, not only in Venice, but also as far afield as Paris, Lyons, Frankfurt, and Geneva. Jenson's type styles were based on the clean lines and subtle distinctive serifs of the lettering on monumental inscriptions that had been created by anonymous Roman stonemasons and other craftsmen a millennium and a half earlier. An important part of Jenson's contribution to the Renaissance was his replacement of the heavy black lettering style of medieval manuscripts, which had served Gutenberg for a model, with the more sharply defined letters of the ancient Roman alphabet.

A key principle of early Venetian typographers, in particular Jenson, was legibility. A generic term for this style of typography is Antiqua, in contrast to the Germanic Black-Letter style. As far as we know, Jenson designed, cut, and founded his own type. No doubt his experience of overseeing working in metal at the French Royal Mint stood him in good stead. According to an advertisement issued by his firm shortly after his death, Jenson's typographic symbols, "do not hinder the reader's eyes, but rather help them and do them good. Moreover, the characters are so intelligently and carefully elaborated that the letters are neither smaller, larger, nor thicker than reason or pleasure demand".

A comparison of a sample of Gutenberg's Black Letter (Figure 1) with Jenson's Venetian Old Style (Figure 2) is instructive. At first glance, the two seem to have almost nothing in common. The letters look as if they might even represent different alphabets. Gutenberg's style is a version of the letters in medieval manuscripts. Jenson's is completely different: to a modern reader, it looks uncannily familiar, because it established typographical principles that are still relevant today. It is astonishingly, even shockingly modern – a design achievement worthy of the 20th-century Bauhaus at its best. It was

the foundation of almost all subsequent type-founding and design in the Roman alphabet down to the present day, with the exception of German Fraktur, which owes more to the tradition of Gutenberg and medieval manuscripts and which, even in 19th-century Germany, was recognized to be unsuitable for printing dictionaries, not least because it is uneconomical in terms of space on the page and its potentially ambiguous when used in a small size. Typefaces based on medieval manuscript lettering are designed to be read slowly and sequentially. Medieval reading was slow. By contrast, the legibility of Jenson's type style enabled fast, non-sequential skimming and dipping, of a kind characteristic of dictionary use.

Figure 1: The Gutenberg Bible (c. 1455): sample from the *Book of Exodus*

Figure 2: Sample of Jenson's typography (from the Wikipedia entry for 'Jenson')

It takes a modern reader all of thirty seconds to become familiar with what can now be seen as the idiosyncrasies of Jenson's Venetian Old Style. These are:

– representation, in certain contexts, of the letters *n* and *m* as a superscript bar over a preceding vowel (suggesting nasalization of the vowel rather a full-quality consonant),

– two forms of the letter *s*, long and short, whose uses are contextually determined,

– two short forms of Latin words meaning 'and': the symbol '&', which is still used today in certain contexts, and 'q:' for the bound morpheme *–que*, which has gone out of use.

In all other respects, Antiqua type styles are recognizably the same as their modern equivalents. Other great type designers and punch-cutters of 16th-century Europe (Graffo, Bembo, Garamond, Baskerville, and others) would design typographical symbols that share most of their fundamental characteristics with those of Jenson, although it has to be said that they do not share the same classic simplicity. Jenson's typographical principles have survived unchanged through the centuries and through various more recent technological revolutions for over 500 years. This is all the more remarkable when we consider the idiosyncrasies of conventional handwriting styles of the Renaissance, which require many hours of training in paleography before they can be read with fluency.

An important aspect, from the point of view of lexicography, of Jenson's contribution was that his typographic principles made it possible for printers to put many more words on the page without sacrificing legibility. This was to be an important contribution to the herculean lexicographic efforts that were to come. In a big text (and Renaissance dictionaries were big), more words on the page means fewer pages, which in turn means a more manageable product.

Fourteen years after Jenson's death, his printing and publishing business in Venice was inherited (in 1494), through marriage, by a man who was to play a pivotal role in the Italian Renaissance. Teobaldo Manucci, better known as Aldus Manutius (1450-1515), was a scholar with a passion for Ancient Greek philosophy and classical literature. Aldus was a man of means as well as scholarship. He devoted himself to using the technology of typesetting and printing to disseminating as many classical works as he could, rescuing them from obscurity, and to preventing further losses. He commissioned the typographer Francesco Griffo to create additional typefaces, including Greek (though the Greek type-faces have many cursive features and are less legible than the Roman ones designed by Jenson). Aldus acquired ancient Greek manuscripts from all over the Levant and the eastern Mediterranean region and employed Greek-speaking editors and compositors to collate and edit these manuscripts and get the texts typeset and printed. Venice was well placed for this activity, as the Venetian Republic during the 15th and 16th century held sway politically over some of the islands of Greece (Naxos, Crete, and the Ionian islands), so he had access to Greek-speaking scholars and workers.

Another important figure must now be briefly mentioned. In 1508 the Humanist scholar Erasmus came to stay in Venice as a guest of Aldus Manutius. Here, he compiled his *Adagia*, a sort of dictionary of quotations from Classical authors. As he readily acknowledged, Erasmus received much help from the scholars and editors in Aldus's workshop, including Aldus himself. The *Adagia* is not merely a collection of quotations and proverbs, but also contains discursive articles on certain selected key words and concepts. It is a precursor of the lexicographical insistence on supporting definitions and explanations with citations.

1.3   Printing in 15th-century England

Printing was introduced to England in the 1470s by William Caxton. Caxton took up printing only towards the end of his life; he was an extremely energetic man with many other business, artistic, and literary interests: a highly respected and successful merchant as well as a writer, translator, printer, and publisher. After a period spent living and working in Bruges and elsewhere, he established himself as an importer of velvet, silk, and other luxurious fabrics, eventually rising to be governor of the Company of Merchant Adventurers of London. It was not until 1475-1476, when he was over sixty years old, that he established the business on which his present-day fame rests. He set up a printing press, at first in Bruges and later in London, in imitation of one that he had observed in Cologne, and began to print books in English. The output of Caxton's press was prolific. Among its most famous publications were Chaucer's *Canterbury Tales* and Mallory's *Morte d'Arthur*.

Caxton and his business partner Wynkyn de Worde (an Alsatian whom he had met in Bruges) did not publish any dictionaries apart from a very modest French-English glossary. The earliest printed dictionary in England was the *Promptorium Parvulorum* ('Young People's Storeroom'), an English-Latin word list, printed in 1499 by Richard Pynson. This work had been compiled about sixty years earlier by Galfridus Anglicus (alias Galfridus Grammaticus 'Geoffrey the Grammarian'), a Dominican friar who lived in Norfolk. Its 10,000 entries (words and phrases) had already been laboriously copied out by hand several times – the only means of dissemination possible until the invention of printing – before Pynson set it in type and printed it (Figure 3). Both Caxton and Pynson used type styles that were based on those of Gutenberg. Neither had been able to learn about or benefit from the streamlined, economical character of contemporary Venetian typography. Indeed, principles of typographical clarity analogous to those of 15th century Venice were not really introduced into England for another 300 years. Over a hundred years after Pynson, Cawdrey's printer still used black-letter type for glosses, and English typography of the 17th and 18th centuries is full of unnecessary flourishes and ligatures. It looks cluttered and fussy compared with the clean lines and legibility of Jenson and Aldus Manutius.

Figure 3: Extract from Pynson's printing of *Promptorium Parvulorum*, 1499

1.4 The Estienne family of Paris and Geneva

If we compare the first printing of *Promptorium Parvulorum* (1499) with the Latin dictionaries compiled, edited, and printed in Paris by Robert Estienne in the 1530s, we see a quantum leap in both technology and scholarship. *Promptorium Parvulorum* is a practical work for students struggling to express themselves in Latin, i.e. it is designed for encoding use. It is printed in heavy black-letter type and even though it was using the

latest technology for printing, its typography makes it look old-fashioned. By contrast, the *Dictionarium, seu Thesaurus Linguae Latinae* (1531) of Robert Estienne is a decoding aid. It is a work for scholarly use by people reading the Latin texts of classical antiquity, many of which were printed by Estienne himself. In his authoritative study of Renaissance lexicography, Considine (2008) argues that preservation of "heritage" was an important part of the goal of Renaissance lexicographers such as Estienne and his sons. Early lexicographers were not merely producing practical tools for language learners or translators; they were contributing to the Renaissance programme of preserving and indeed reviving the classical heritage.

The type of Estienne's *Dictionarium* was designed, cut, and cast by Claude Garamond, one of several type cutters with whom Estienne had a business relationship. Garamond's elegant type style owes more to the Venetian Antiqua school of typography than to Gutenberg, though it is embellished by the occasional flourish which Jenson would surely have regarded as superfluous. Nevertheless, Estienne's *Dictionarium* is both a work of scholarship and a triumph of elegance in the printer's art – an aesthetic pleasure to peruse as well as a scholarly inventory of the vocabulary of classical literature. This is also true, though to a lesser extent, of the *Thesaurus Linguae Graecae* (1572), which was compiled, edited, printed, and published forty years later by Robert's son Henri Estienne II.

If we look at an entry from Estienne's 1531 *Dictionarium* – I have chosen, more or less at random, the entry for *conclamo* (Figure 4) – we can see immediately that this is not a bilingual French-Latin dictionary. It is a monolingual dictionary of Latin, with a French gloss (in this case, "Crier") appended. The rest of the text of the entry is taken up with morphological information, a monolingual gloss in Latin ("simul clamare"), and a great wealth of citations from Latin authors, on the basis of which Estienne offers collocational norms, some of which are glossed or explained in Latin (not French).

The French glosses in Estienne's *Thesaurus Linguae Latinae* play a comparatively minor role. More striking is the large number of citations and references. Estienne was concerned not merely to say what the meaning of each Latin word is, but to record where the word is used in the classical Latin texts that he had available to him. This is in essence very similar to the lexicographical principles adopted for the academy dictionaries of the 17th century to the present day.

Estienne's work is part of the true foundation of European lexicography. Following Starnes (1963), we may regard Robert Estienne's Latin dictionary of 1531 as a seminal work, but this does not mean that it had no predecessors or that he and his team of lexicographers were working in a vacuum. He was an early part of a highly productive accretive continuum of European lexicography. Other Latin dictionaries had appeared even earlier, in particular that of Ambrogio Calepino (1502), of which more later. It is clear that the scholars in Estienne's workshop made use of these works, just as OED built on the foundations laid by Johnson (1755) and other earlier lexicographers.

Among the factors that distinguish Estienne's 1531 dictionary from its predecessors are its meticulous scholarship, the systematic inclusion of citations from works of classical literature (many of which were also printed by Estienne), a concern with semantic differentiation and phraseology, and reliance on readable typography.

Conclámo, conclamas, pen. prod. conclamare, Simul clamare.
Cic. 6. Philip. 105, Cum vos vniuersi vna mente atque voce iterũ
à me conseruatam remp. esse conclamastis. Criii.
Conclamare ad arma, Liu. 3. ab vrbe, 109.
Conclamare victoriam, Cæsar 5. bel. Gal. 138.
¶ Conclamare, pro clamare. Cæs. 1. bel. Gal. 106, Quos cũ apud se
in castris Ariouistus conspexisset, exercitu suo præsente con-
clamauit quid ad se venirent. Plaut. Milit. 5. 11.
- vbi abiit, conclamo,
Heus quid agis tu in tegulis?
Conclamare auctione. Plaut. Menæb. 14. 96, Vis conclamari au-
ctionem fore quidem?
Conclamatum est. Terent. Eun. 1. 3. 56, Desine, iam conclamatum est.
Vbi Donatus, Conclamatum est, transactum ac finitum: vt Con-
clamata corpora, quæ nihil reliqui iam habent ad vitæ officia.
Lucanus, Corpora nondum Conclamata iacent. Aut cõclama-
tum, satis deploratum, satisque vociferatum est: quia dixerat,
O infortunatum senem, si & hic amare cœperit. Táquam dicat,
Iam occisum patrem, tandé scimus, iàmque deflemus. Hæc Do-
natus. Sipontinus autem, Cõclamatum est (inquit) inde dici con-
sueuit, quia cum spiritus vitalis aliquando interclusus fallere
homines soleret, mortui aqua calida ablui consueuerant, & per
interualla conclamari. In quo cum liberari nequirent, post vl-
timam conclamationem quasi rebus desperatis vtebátur, & ci-
neres eorum condebantur sepulchris.
Conclamata vasa. Exercitu castra moturo, conclamari solet vt
vasa colligant. Cæsar, Vasis militari more conclamatis, pro-
gressus milia, &c. Idem 1. bel. ciuil. 144, Quo cognito signũ
dari iubet, & vasa militari more conclamari.
Non conclamatis vasis abire, per translationé dicitur de iis qui
hospite (vt aiunt) insalutato abeũt. Cæsar 3. bel. ciuil. 144, Sci-
pio nostrorum studium alacritatémque pugnandi cũ cogno-
uisset, suspicatus fore vt postero die aut inuitus dimicare co-
geretur, aut magna cum infamia in castris se continere, noctu
neque conclamatis quidem vasis, flumen transiit.
Conclamito conclamitas, conclamitare, frequentatiuum. Plautus
Merc. 3. 51, Conclamitare tota vrbe, & prædicare.

Figure 4: R. Estienne, *Dictionarium*, 1531, entry for 'conclamo'

...quam didicit, quasi à seipso doctus. Od. χ de Phemio ci
tharœdo, Αὐτοδίδακτος δ᾽ εἰμί, Θεὸς δέ μοι ἐν φρεσὶν οἴμας
Παντίας ἐνέφυσεν. dicit se à nullo magistro doctum Dei
afflatu & instinctu canere. Sic Xen. paupertatem
πενίας αὐτοδίδακτον appellauit. sic & in Epig. αὐτοδίδακτος
ὄψις dicitur. ¶ Αὐτοδίδακτος (inquit Bud.) Suapte na
tura aliquid faciens, vt quum lætabundus rubescit.
INDE Αὐτοδιδάκτως idem Bud. exponit Naturæ in-
stinctu.

Figure 5: Extract from H. Estienne's *Thesaurus Graecae Linguae*, 1572

There can be no doubt that Considine (2008) is right that the main purpose of Robert Estienne's 1531 *Dictionarium* was to contribute to the preservation of the heritage of classical literature, and the same is true of the equally ambitious and equally monumental *Thesaurus Graecae Linguae*, published by his son Henri Estienne in 1572 (Figure 5).

Two other important dictionaries of Robert Estienne show a different side of this great lexicographer. As we have seen, his main concern in 1531 was to cater to the needs of scholars and literati by preserving the heritage of the classical Latin language. But he was also sensitive to the needs of more humble students and language learners. The *Dictionnaire francoislatin* of 1539 (Figure 6) is a practical work explicitly aimed at students wishing to express themselves in Latin. A noticeable feature is the large number of idiomatic French phrases for which Latin equivalents are offered. For example, *l'ordre et collocation des mots* is glossed as 'verborum constructio'. Robert Estienne placed considerable emphasis on phraseology and context: it is perhaps not too fanciful to believe that he would have been sympathetic to modern theories of collocation and construction grammar.

Figure 6: R. Estienne, *Dictionnaire francoislatin*, 1539, entry for 'mot'

A complementary (and equally practical) work, published by Robert Estienne in 1552, is the *Dictionarium Latino-Gallicum* (Figure 7). This is not a revised version of his 1531 work. Instead, it is a practical guide whose aim is to help students decode the meanings of Latin words and Latin texts into their native French. As can be seen in Figure 8, there

are many more French glosses on the Latin words and phrases than in the 1531 work (though they are still, by modern standards, sparse). The 'principle parts' of verbs are given at the start of the entry ("conduco, conducis, conduxi, co[n]ductum, conducere"), which is helpful for both decoding and encoding use to anyone studying Latin. Citations from literature have been replaced by short phrases, often with a gloss. The authority of a classical author for phraseology is invoked in abbreviated form, generally without a full citation. Thus, the Latin phrase 'nimium magno conducere' is included on the authority of Cicero and glossed as 'Acheter trop cher', i.e. in English, 'to buy too dear'. This is information of a kind that is particularly useful for students learning to read and understand Latin texts, as opposed to scholars who were already fluent in Latin. It is also, coincidentally, of potential interest to modern scholars studying the cultural persistence of conventional metaphors and idiomatic phrases in European languages going back to classical Latin.

Figure 7: R. Estienne, *Dictionarium Latino-Gallicum*, 1552, entry for 'conduco'

Trench (1858) rightly describes lexicographers as "the inventory clerks of language", but these great Renaissance lexicographers were very much more than mere inventory clerks. They were scholars, compilers, definers, printers, and publishers. The Estienne firm was founded by Henri Estienne (c. 1460-1520), who had married the widow of a printer in 1502 and expanded the business. Three sons and two grandsons became printers. There can be no doubt that Robert Estienne (1503-1559) was the greatest of the family, even though his son Henri II was to successfully tackle the even more challenging task of compiling a scholarly dictionary of classical Greek. Part of the greatness of Robert lies in his evident concern for students as well as scholars and the range of the different dictionaries that he and his staff compiled and published, a range that would have been quite impractical without the recent innovations in the technology of printing and typesetting.

In addition to his remarkable achievements in scholarly and practical lexicography, Robert Estienne also ran a successful printing business, publishing editions of major classical texts and other works. According to his biography (Armstrong 1954), he printed and published on average 18 books a year in Paris, as well as undertaking his massive lexicographic projects. He ran a lively and polyglot workshop. According to his son Henri II, "There sat down to table daily a staff of ten assorted nationalities, together with family and guests, all speaking Latin, including the servants" (Armstrong 1954, 15). She estimates, on the basis of contemporary records, that in its heyday the firm employed a staff of 50 (2 type-founders, 18 compositors, 5 proof-readers, 21 printers, 3 apprentices, and one shop boy), in addition to the master himself and his family. Estienne was on intimate terms with the greatest Parisian scholars and intellectuals of his day. He styled himself "printer to the king" but eventually, as an outspoken Protestant, in or before 1550 he found it prudent to remove himself to Geneva, where his output dropped to about six books a year.

1.5   Polyglot and bilingual dictionaries during the Renaissance

The most important and innovative bilingual dictionary of the early 16th century was compiled in English. It is Palsgrave's ambitious *Lesclaircissement de la langue francoyse* (1530). Palsgrave had been tutor at the English court to Henry VIII's sister Princess Mary, who in 1515 became Queen of France. His guide to the French language is not only a bilingual dictionary but also a grammar. The dictionary part contains 18,890 English-French equivalents. Black Letter type is used for English, Antiqua for French. The arrangement is alphabetical by part of speech; i.e., each part of speech is given a separate "table". The table of substantives consists mostly of single-word equivalents, with disambiguation of polysemous words, e.g. there are two entries for ***meale: meale of corne*** is glossed as 'farine', ***meale of meate*** is glossed as 'repast'. The table of verbs pays more attention to phraseology (see Figure 8). Each sense of each English verb is first embedded in an English phrase (or given an English gloss), and then the target word and/or the phrase as a whole is translated into French.

Palsgrave was a true comparative linguist as well as a pioneering lexicographer. However, rather surprisingly, his fine example was not followed: his work did not serve as a model for other bilingual dictionaries of vernacular languages – at least, not for another

sixty years. Instead, the standard lexicographical tool used for translation during the Renaissance was a polyglot dictionary based on Latin. It is time to examine how this came about.

Figure 8: Extract from Palsgrave's *Lesclaircissement de la langue francoyse* (1530)

By 1490, many cities in Italy, Germany, and the Netherlands had a printing press, many of which produced dictionaries, vocabularies, and word lists of one sort or another – some in thematic order, others in more or less exact alphabetical order. Most of these were monolingual Latin dictionaries, the demand for them reflecting the status of Latin as the language of knowledge, culture, and international communication. The first Greek-Latin lexicon was compiled and published by a Carmelite monk, Giovanni Crastone of Piacenza (1497).

As for vernacular languages, there blossomed in the early 16th century a fine crop of monolingual Italian dictionaries, as described by Alonge (2006). This is a clear indication of the confidence of Italians in their language as a literary medium rivaling Latin, distinguishing it in status from other vernaculars of Renaissance Europe.

Surprisingly, though, there were few bilingual dictionaries of vernacular languages at this stage. Everything was mediated through Latin, which functioned as a sort of inter-lingua. As shown by Kramer (2006) and Schoonheim/Pijnenburg (2006), in the German-speaking lands and the Netherlands, early Latin-German and German-Latin lexicographic works appeared, notably Van der Schueren (1477), Dasypodius (1535-1536), and Maaler (1561). The complex relationships among European languages of this period are well described in Burke (2002).

The seminal work in the development of European multilingual lexicography was the *Dictionarium* of Ambrogio Calepino. Calepino's original edition (1502) was a Latin alpha-betical vocabulary book, with glosses in Latin supported by citations, together with en-cyclopedic entries for the figures of classical mythology. In a second edition, glosses in Italian and French were added. By a process of accretion, the vocabularies of other lan-guages, starting with Greek and Hebrew, were gradually added to successive editions of Calepino's original. In the words of Fried (2007, 231), "it evolved into the first polyglot dictionary". By 1580, a dozen different editions, containing glosses in up to eleven dif-ferent languages, all attributed to Calepino, were in print, published in locations as far apart as Reggio nell'Emilia, Venice, Paris, Strasbourg, Hagenau, Lyon, and Rome. In Paris alone, five competing editions appeared between 1524 and 1541. The 1573 edition printed and published in Venice includes the following comment in its front matter, quoted and translated by Freed:

> *In hac postrema editione, ut hoc dictionarium commodius exteris nationibus inservire possit, singulis vocibus latinis italicas, gallicas, & hispanicas interpretationes inseri curavimus.*
>
> In this latest edition, in order that this dictionary might more fully serve foreign nations, we have taken care to insert Italian, French, and Spanish definitions among the lone Latin entries.

By this time, Ambrogio Calepino himself (1450-1510) was long dead and his book had become common property. Stathi (2006) argues that the popularity of the many 'Calepi-nos' was due, not to its etymologies, but to its explanations of meanings and the inclusion of examples of word use. The extraordinarily complex bibliographical history of this work and its derivatives was traced by Labarre (1975). This shows that multilingual edi-tions really began to take off in the 1550s (Figure 9); by the 1580s it had come to include lexical items in up to 11 languages – not only Latin, Greek, Hebrew, Italian, French, and Spanish, but also 'outlandish' tongues such as German, English, Polish, and Hungarian. By the end of the century, a Latin-Portuguese-Japanese 'Calepino' had appeared, sup-porting the missionary work of the Portuguese Jesuits who were at that time attempting to Christianize Japan.

It has been said (with what justice I know not) that Calepino's work is deficient in schol-arly precision. Moreover, the polyglot works that bore his name are great, cumbersome things, not very suitable for carrying around on a tour of Europe and not particularly user-friendly. Nevertheless, they seem to have been the principal works that served the practical translation needs of Europeans in the 16th century.

Not only did Calepino's work become the common property of Europeans in many different editions; his very surname also passed into the vocabulary of Italian, French, English, and other languages. In the 16th and 17th centuries, Italian **calepino** and Eng-lish **calepin** were used as generic terms for a dictionary. In French, **calepin** was further

extended to mean a notebook or a compilation of rare and unusual linguistic facts, and was used in various colloquial expressions such as *mettez cela sur votre calepin* 'add that to your calepin'. Watson (1908) and Starnes (1955) show that a 'calepin' was a widely available – and widely used – resource in schools and universities throughout England in the 16th century. **Calepine** was also adopted by Edmund Spenser as a proper name for an allegorical character in the *Faerie Queene*, the significance of which is discussed by Fried (2007) in an article that contains a remarkably illuminating account of Renaissance lexicography.

Conclamo, as, are, a.p. Simul clamare, clamare. συμβοᾷν. { Ital. Gridare infieme, gridare. Gall. Crier enfemble, s'efcrier. Hifp. Liamar dando uoces. Germa. Mitfchreien, zûfamen fchreien. } Cic. 6. Phil. Quum uos uniuerfi una mente atᵹ uoce iterum à me conferuatam effe Rempub. conclamaftis. Tacit. lib. Conclamant patres, cor pus ad rogum humeris Senatorum ferêdum. ¶Conclamare, pro fimplici Clamare. Cæfar 1. bel. Gallic. Quos quum apud fe in caftris Ariouiftus confpexiffet, exercitu fuo præfente conclamauit quid ad fe uenirent. ¶Conclamare uafa, militaris eft locutio, pro denunciare inter milites, ut uafa colligant: quôd fiebat quum exercitus moturus erat caftra. ἐναφυγὴν σημαίνειν. Cæfar lib. 1. bell. Ciu. Quo cognito, fignum dari iubet, & uafa militari more concla mari. ¶Côclamatum eft, hoc eft, tranfactum & finitum eft. πέπρα κται. { Ital. Egli è finito. Gall. Tout eft fait. Hifp. Le cofa efta acabda da. Germ. Die fach ift fchon vber hin, es ift fchon befchehen. } Quod locutionis genus ab ijs tractum putatur, qui morientibus affiftût. Solent enim illi cadauerum recens exanimatorum faciem identidem frigida abluere, nomenᵹ eorum clara uoce inclamare, ne fortè fyncope correptos pro mortuis funerarent. Nam qui animi deli quium patiuntur, mortuis fimilimi funt: unde & nonnullos pro mortuis claros, & à rogo relatos accepimus. Quum itaᵹ affufa fub inde frigida, nomineᵹ fæpius inclamato nihil proficeretur, poft ul timam conclamationem re iam defperata comburebâtur. ¶Hinc corpora conclamata dicuntur, hoc eft, deplorata, quibufᵹ extremum hoc conclamationis præftitum eft officium. Luca. corpora nondum Conclamata iacent.

Conclamatio, onis, uerbale καταβοὴ, καταβόησις. { Ital. Crido di piu genti. Gall. Cri de plufieurs gens. Hifpan. Aquello llamado con bo zes. Ger. Das zûfammen fchreien uiler. } Clamor multorum Seneca quum aliqua conclamatio eft quomodo exeat, non quid efferat, quærit.

Figure 9: Extract from a 1550 Basel edition of Calepino

There were some exceptions to all this polyglottalism. As we have already noted, Caxton printed a short, practical French-English vocabulary in 1480, although this is a comparatively minor work. An Italian-German thematic dictionary, *Introito e porta*, was compiled by Adam von Rottweil as early as 1477. It stands at the head of a long tradition, ultimately comprising 89 separate publications between 1477 and 1636.

1.6   Dictionaries in 16th-century England

The Renaissance dictionaries discussed in the preceding sections bore rapid fruit in England, in the first place as a source for the first printed Latin-English dictionary in England, the *Dictionary* of Sir Thomas Elyot (1538) (Figure 10). Unlike *Promptorium Parvulorum*, this was a work for decoding use, as was its most important successor, the *Dictionarium Linguae Latinae et Anglicanae* (1587), compiled and printed by Thomas Thomas, printer to the University of Cambridge (Figure 11). This latter work enjoyed tremendous success for several decades. It is admirably succinct and practical. The English glosses in it are full and informative. As printer to the University of Cambridge, Thomas Thomas was well aware of the needs of students and was at pains to provide them with help in the form of systematic but succinct glosses in their own native tongue.

Typographically, Elyot's work is very obviously indebted to the medieval tradition of Pynson, Caxton, and Gutenberg, whereas Thomas's work of fifty years later is very much more legible. It owes much to the Renaissance typographical tradition of Estienne, Aldus Manutius, and Jenson – though it must be said that it seems sadly debased compared with the beautiful clean lines of Jenson's original Venetian Old Style. Neither the Parisians of the 16th century nor the Elizabethans in England could resist a flourish – literary or typographical.

Finally, in this brief survey of Renaissance lexicography, we must mention the evolution of bilingual dictionaries proper. Two such works are well known to students of Shakespeare: John Florio's *Italian and English Dictionary* of 1598 and John Minsheu's *Dictionarie in Spanish and English* of 1599. Both of these are practical works for the emerging modern world and what was eventually to become the European Enlightenment. They are designed as aids for translation between contemporary vernacular languages, rather than being mediated through Latin. Minsheu's work was an expanded version of an earlier work called *Bibliotheca Hispanica* (1591), compiled by Richard Percyvall. Along with Palsgrave for English-French, these are the precursors of modern bilingual dictionaries.

Minsheu was to go on to compile *The Guide into Tongues* (*Ductor in linguas*, 1617), an ambitious polyglot work in eleven languages. It would no doubt be an interesting research topic to determine the debt of Minsheu to Calepino. This, however, lies outside the period and the scope of the present study.

Despite the efforts of Palsgrave (1530), it was not until the 1590s that the European intelligentsia accepted that it was not necessary to use Latin as an interlingua or reference point in order to translate words and phrases of one vernacular language into those of another. The first French-German/German-French dictionary was published in 1596 by

Levinus Hulsius in Nürnberg. He also compiled the first Italian-German/German-Italian dictionary. Other bilingual dictionaries of vernacular languages were to follow thick and fast during the 17th century.

These dictionaries contributed, albeit somewhat belatedly, to the internationalization of European culture, making the literature and culture of countries such as Italy and France accessible to speakers of remoter northern languages such as English.

A batis, two wordes made of the preposition, a, z the ablatiue case plurel of Batus, whiche in englysshe is a measure, signifieth an officer, that hath the orderyng of measures, as a clerke of the markette
Ab accidentibus, an officer, vnto whome it belongeth to write such thing as do chance.
Abacti, officers deposed, or such as be constrayned to resygne their authoritie.
Abactores, theues that steale cattell.
Abacus, a countyng table, or cupborde.

Figure 10: Extract from the *Dictionarium* of Sir Thomas Elyot, 1538

Cactos, Plin. A kinde of thistle, after some an Artichoke.
Cacubalum, Plin. Herba acinos fereas nigros, quę etiam vocatur Sarichnos, & Strumus.
Căcŭla, æ, m. g. Plaut. A souldiours slaue or boy page.
Caculatum, ti, n. g. Servitium, Fest.
Căcūmen, ĭnis, n. g. Plin. The toppe, heighte, or sharpe ende of a thing : the ridge of an house: also perfectenes.
Căcūmĭnātus, a, um, part. vel nom. d part. Plin. Sharpe topped, with sharpe toppes.
Căcūmĭno, as, Ovid. To make copped or sharpe at the toppes.
Cădāver, ĕris, n. g. A deade bodie, carrion, a carkasse, a corse: also ruine.

Figure 11: Extract from Thomas Thomas' *Dictionarium Linguae Latinae et Anglicanae*, 1589

Figure 12: Extract from Minsheu, 1599

1.7   A 21st century analogy with the Renaissance

In this first part of the paper I have identified three themes in Renaissance lexicography: the preservation and dissemination of the classical heritage; the creation of practical tools for students of Latin and Greek; and the emergence of bilingual dictionaries as practical aids for translation among vernacular languages. None of this would have been possible without the invention of printing technology and the creation of type fonts that make economic and elegant use of space on the page.

I have argued that histories of English lexicography such as Landau (2001) should pay more attention to the formative influences of the great Latin dictionaries of the 16th century. Studies by scholars such as Armstrong, Bately, and Considine provide an important

perspective. A curious fact is that much 16th-century European lexicography used Latin as an interlingua, so that it took several decades for genuine bilingual lexicography to emerge, apart from a few pioneering works such as Palsgrave (1530).

A modern analogue now suggests itself, namely that of the development of computer technology in the second half of the 20th century, which could have (or should have) an impact on present-day lexicography that is as profound as was the development of printing technology in the 15th century. The full possibilities are only just beginning to be worked out. There are at least four aspects:

1. **Evidence.** Just as the Renaissance programme of collecting, printing, and publishing the texts of classical antiquity led to major, technologically innovative dictionaries of Greek and Latin, so the advent of electronic corpora and internet search engines have opened up possibilities for new lexicographic descriptions of phraseology and meaning in contemporary languages.

2. **Resources.** For Renaissance lexicographers, newly printed copies of classical texts served as resources to be quarried for the lexis of Latin and Greek. At present, a plethora of electronic resources, of variable quality and accuracy, for NLP and AI applications are being developed for modern languages. One only needs to look at the Global WordNet Programme, to see an example. It remains to be seen who will be the Robert Estienne of the 21st century and how he or she will present the lexicon of a modern language (presumably English) for a new generation of users, which will include machines as well as humans.

3. **Compilation.** In the 16th century, the index card was invented and used to compile lexicographical information and sort data into alphabetical order. Now, the computer has freed lexicographers from the tyranny of having to work in alphabetical order. Any entry in any part of the alphabet can be compiled, edited, improved, and expanded at any time.

4. **Dissemination.** The invention of printing enabled the rapid reproduction of large numbers of copies of large, complex texts in legible print. This was to be an essential component of lexicography for the ensuing 500 years. At the present time, this whole technology is being superseded by on-line dissemination of information. The waters are muddy and a business model has not yet clearly emerged. But the potential is tremendous. It has hardly begun to be tapped.

2. Lexicography and technology in the 21st century

2.1 Corpus evidence

The single most important technological development in the 20th century, as far as lexicography is concerned, was the development of the electronic corpus. Nowadays, it is commonplace for corpus linguists to stuff billions of words of text into a computer, tokenize them, tag them in various ways (usually starting by assigning a part of speech to each token), lemmatize them, and analyse them in various ways for various purposes. This technology is a central component of all reputable modern lexicographical research. Before publication of the Cobuild Dictionary in 1987, lexicographers were reliant on two source of evidence:

1. vastly slow and expensive reading programmes, such as that which made the OED possible at the end of the 19th century;
2. introspection – consulting one's intuitions in order to invent example of word uses in sense that the lexicographer already believed to exist.

Dictionaries that are not supported by a vast budget cannot afford either the time or the money for a reading programme that can be relied on to collect examples of all uses of all words in a language. As regards reading programmes, the fact is that a reading programme has a built-in tendency to distort. This is because citation readers tend, very naturally, to select citations for rare, unusual, and 'interesting' uses. They do not send in citations for words and uses that may be presumed to have already been collected from other sources. James Murray recognized this problem in 1878, at the very beginning of his work as editor of the work that was to become the OED, when he had just begun to sort the 'slips' on which citations collected by the reading programme were copied out:

> The editor and his assistants have to spend precious hours searching for examples of common everyday words. Thus, in the slips we have 50 examples of **abusion,** but of **abuse** not five. (Murray 1878)

As regards the second source of evidence, introspection, it has gradually become clear from corpus-driven studies of lexis that introspection is a flawed source of evidence. In the first place, it is conducive to self-fulfilling prophecies. The lexicographer or linguist believes that a word has a certain meaning or use, and invents an example to support that belief. In the second place, it encourages linguists and lexicographers alike to ignore facts that even casual corpus analysis can make blindingly obvious. Despite being sanctioned by generative linguists and cognitive linguists alike, reliance on invented examples and introspective judgements of acceptability has led to a situation in which all the research findings based on such examples (that is to say, much of the linguistic research of the past seventy years) must be regarded as suspect. Such findings need to be confirmed, not only by the acceptability judgements of a peer group, but also by empirical evidence of actual usage. Corpus-driven lexicography has a role to play in this wholesale re-evaluation of linguistic research.

One type of finding that is often overlooked by evidence based on citations collected in reading programmes and examples invented by introspection is the relative frequency of different uses or meanings of words.

I will discuss in some detail a very simple example, from which much follows. Some years ago, in conversation with an eminent American linguist who shall be nameless, I was surprised to discover that, in his deep introspection-based analysis of the verb *hazard*, there was no place for the possible meaning 'say tentatively'. If we ask, "How is the English verb *hazard* normally used?" corpus evidence confirms the introspectively obvious answer: in normal usage, "People hazard guesses" accounts for over half the English sentences containing this not very common verb.

As soon as a reader sees this answer, it seems obviously plausible. In this case, the corpus evidence and our intuitions are in accord.

But then counterexamples may start springing to mind. Maybe you can hazard an idea. Is an idea a kind of guess? It also seems plausible that you can hazard your money or

your life – but *money* and *life* are certainly not kinds of guesses. What is going on here? Let us enjoy the new-found luxury of consulting corpus evidence, and look and see what other uses of the verb *hazard* can be found in readily accessible actual data. Looking at the BNC-50 corpus (consisting of 50 million words, half the British National Corpus) here is a summary of what I found.

The part-of-speech tagger used by the Sketch Engine corpus search tool (Kilgarriff et al. 2004) finds 50 verb uses of hazard in this half-corpus. Six of them turn out to be tagging errors: they are nouns or modifiers, not verbs. For example, the expressions *hazard assessment*, *hazard control*, and *hazard studies* are all modifier + noun, not verb + noun. These are mistakes – but they are not unreasonable and not unusual. An error rate of 12% in part-of-speech tagging is about par for the course using the CLAWS tagger.

Of the 44 verb uses in this half-corpus, 23 take the word *guess* or *guesses* as a direct object. In other words, the expression *hazard a guess* is a prototypical norm of English usage, accounting for around 50% of all uses of the verb.

```
1.   loving friendship with Justin. Guesses are hazarded, and are quoted from
2.  experience and technical background. And, I hazard a guess, more logged
3.being a builder, nor an architect, I can only hazard a guess. During construction in
4. Punter. What's it all about? I can only (hap)hazard a guess. BAY CITY ROLLERS The
5.   such mindless, moronic abuse, one can only hazard a guess. Thank goodness Ronnie
6.mining industry? Would my hon. Friend care to hazard a guess about how many fewer
7.  had hair and eyes like her mother. I would hazard a guess and say she would be
8.  all, can result in lost profits. When staff hazard a guess as to the price of
9.  party. No one at this stage is prepared to hazard a guess at the outcome of the
10.        What the connection is we can only hazard a guess at but it confirms all
11. ceived virtually no recognition? I can only hazard a guess at what it must have
12.  replied, `But off the top of my bonce I'll hazard a guess how many have
13.   away from the wall. Stifling a giggle, she hazarded a guess that the wardrobe
14.  rs to them as Part I and Part 2. One might hazard a guess that Part I was
15. UTIONS Where do your art materials live? We hazard a guess that they're lurking
16. ipatory excitement than others, and I would hazard a guess that, even if they've
18.   command line to begin restoring. But I'd hazard a guess that if you restore the
19.   to work OK once you boot from a floppy I'd hazard a guess that the MS-DOS system
20. passengers in those stations' heyday, but I hazard a guess that considerably more
21. raiser to the day's racing. In fact I would hazard a guess that one, if not both
22. me movies age and some movies date. I would hazard the guess that The Graduate
23. ntity of the farmer is not revealed; we may hazard the guess that he was William
24.  agent of society itself. Indeed, one could hazard a further [ ] observation
25.   is becoming proficient. Perhaps we can now hazard an attempt at defining `a go
26.   the North American standards. He does not hazard any opinions on how
27.   6.1 The Phillips curve. Although Phillips hazarded some theoretical conjectures
28. succeeding shapes and colours from which we hazard the inference that a leaping
29.  about his achievement, such as it was, and hazarded the opinion that he might
30.part of the tour's organiser -- and, I would hazard, a severe case of under-
31.   odd name--`Chicken'. `Not Hen Chicken?' I hazarded, as this humorous
32. e demand. Some critics of the NAIRU concept hazarded that, if there was any
33. gh in parliament, and it seemed sensible to hazard that a man of this standing
34. ultimate, supreme grade of evil'. It may be hazarded that it was this inevitable
35. ffect. To take the British example, I would hazard that the ratio of real balance
36. tleman, in his stable costume, would rather hazard his neck four-in-hand, than
37. would have been lost and commandos were not hazarded in foolish risks, although
38. objective, and a principle strong enough to hazard lives for, America cannot hope
39.marvelled at the readiness of the British to hazard so much in company with Franc
40. ted making references to `underwater rocks' hazarding the goodwill of his visit.
41. one may from time to time admire people who hazard their entire company on one
42.  d who wish the King [] well, who would not hazard their estates for him".
43. save my life I know you would, some of you, hazard your own. And yet will not
44. e morals and eternal welfare of numbers are hazarded and ruined for want of an
```

Figure 13: Some uses of the verb *hazard*: concordance from BNC50

But what about the remaining uses of this verb? Lines 24-35 can be classified as exploitations of the most normal use of the verb, whereas 36–44 are examples of a completely different norm for the same verb. Why should we say this? On what facts in the texts is this assertion based?

In 24-29 the direct object has the same semantic type as the prototypical direct object, *guess*. *Conjecture*, for example, is a near synonym of *guess*. Both words denote speech acts or mental events in the mind of the speaker. Likewise, an *observation*, an *attempt at defining*, *opinions*, and *inferences*, are kinds of speech act or thought act. A *that*-clause (as in 32-35) represents a speech act or a proposition underlying a speech act, while direct speech (as in 30-31) is an overt expression of a proposition. The words *I hazarded* are, it seems, deliberately chosen in 31 (rather than the more neutral expression *I said*) in order to imply that what is said is a matter of guesswork rather than a simple factual statement. It can be seen that in 24-35 the semantic values gradually move outward from the prototypical norm *hazard a guess*. Some speakers of American English assert that for them direct speech with the verb *hazard*, as in 30 and 31, is unidiomatic. Be that as it may, even in British English this is a comparatively rare way of exploiting the norm, though arguably all the more effective because of its rarity. By contrast, in 36-44 the direct objects belong to a completely different semantic type and activate a different sense of the verb, one that is a close synonym of *risk*. Further examination of corpus evidence shows that people hazard not only their lives and their companies or business enterprises but also their wealth, their reputation, and other valued entities. Here, what is hazarded is an object of value, and the purpose of hazarding it is to try to obtain some benefit. The two norms are not entirely independent. As a matter of historical fact, the notion of hazarding a speech act developed as an exploitation of the notion of hazarding or gambling money or an object of value in a game of dice. A person's ability to say true things is, in social terms, an object of value. But this is irrelevant to the meaning and use of the word in modern English. Very often, a modern sense of a word develops as an exploitation of an earlier sense. Sometimes the two senses coexist peacefully for centuries. In other cases, the new sense elbows the older one out of existence.

The point of this analysis is not to show that *hazard a guess* is an idiomatic phrase of English, but to illustrate how normal phraseology works. Instead of trying to draw boundaries encompassing all possible uses of a word, lexicographers of the future, with the benefit of corpus evidence at their fingertips, will seek to identify the prototypical phraseological norms with which each meaning of each word is associated. Needless to say, most cases will turn out to be a lot more complex than *hazard*, but careful examination and sorting of very large quantities of data for each word will generally reveal phraseological prototypes around which each sense is grouped. The lexicographer's art is then required to decide where to draw a line on the continuum between normal uses and exploitations of norms.

2.2   Lexical sets and semantic types

Few verbs are associated as strongly as the one we have just discussed (*hazard*) with just one noun as the typical direct object. However, almost all verbs are associated with one or more prototypical groups of nouns, and these different groups often pick out different

senses of the verb. This is explained in more detail in Hanks (2012), using the example of the verb *shower*. That essay also goes on to show how a rather different apparatus is needed for the corpus analysis of nouns, using the noun *spider* as well as *shower* used as as a noun. There is no need to repeat that detailed analysis here. The associations between phraseological norms and meanings of English verbs are being explored by corpus pattern analysis in the AHRC-funded DVC project ("Disambiguation of Verbs by Collocation"), being conducted at the Research Institute of Information and Language Processing at the University of Wolverhampton, UK, with support from corpus linguists at the Faculty of Informatics, Masaryk University, Brno, Czech Republic. The results of the analyses (800 verbs so far analysed) are posted at http://deb.fi.muni.cz/pdev/.

Natural language is a puzzling mixture of logical and analogical processes. Until the availability of large quantities of corpus evidence, the analogical side was largely neglected. Corpus-driven studies of lexis (e.g. Sinclair 1964, 1991, 1998, 2004; Stubbs 2001) have provided ample evidence that the cognitive prototypes identified by Rosch (1973), Geeraerts (2010), and others are matched by phraseological prototypes that shape everyday usage in natural languages such as English.

Unfortunately for the lexical analyst, prototypical phraseology often consists of groups of words associated with other groups of words, rather than individual lexical items. Hanks/Pustejovsky (2005) and Hanks/Jezek (2008) discuss both the merits and some of the problems of grouping words together in lexical sets according to a shared semantic type. A well-known example concerns the idiom *to be shaking in one's shoes* (meaning to be frightened), which can also appear as *shivering in one slippers, quaking in one's boots, quivering in one's sandals*, and various other phraseological permutations and exploitations.

These are examples of the issues with which the empirical lexicographers of the future will grapple, if they are to achieve an accurate representation of the relationship between word use and word meaning.

2.3   Exploiting phraseological norms

The central argument of Hanks (2013) is that a natural language is indeed rule-governed behaviour, as so many thinkers have observed, but that there is not just one gigantic monolithic system of rules governing such behaviour. Instead there are two interlinked rule systems: one governing the normal, conventional, grammatical uses of words; the other governing creative exploitations of normal word use, in metaphors, metonymy, ellipsis, anomalous arguments, etc. The business of lexicography is with the first of these; bad lexicography often results from a failure to make a distinction between normal word use and creative exploitations, as in the case of the American dictionary that defines *newspaper*, not only as a noun, but also as a verb, with the sense "to do newspaper work (as running a newspaper or reporting or editing news)". This is a clear case of failing to recognize a linguistic exploitation for what it is. I argue that such uses do not belong in a dictionary unless there is clear evidence that they have established themselves as conventions of the language. Unfortunately, the dividing line between norms and exploitations is not clear-cut, not least because some imaginative creative uses of

words tend to catch on and become part of the norm. Today's exploitation may become tomorrow's norm. Part of the job of lexicography is to monitor the gradual establishment of particular exploitations as new secondary norms.

Continuing the discussion of the verb *hazard*, let us conclude by looking at a clear case of an exploitation, in order to see how this aspect of meaning in language works.

> 1. I hazarded various Stuartesque destinations like Florida, Bali, Crete and Western Turkey.

Many readers – especially computational linguists and other people with a logical orientation – coming to this sentence out of context like this judge it to be crazy, meaningless, unidiomatic, or uninterpretable. But this fact merely underlines the unnatural nature of what linguists and logicians do in general and what corpus linguists do in particular. No normal reader takes a sentence from the middle of a text and pores over it, asking what it might mean, without reference to what has gone before. Texts have a beginning, a middle, and an end. Example 1 comes from Julian Barnes's 1991 novel *Talking It Over*. Barnes is a writer admired for his stylistic elegance – The Complete Review, for example, when reviewing this novel called him "a very fine stylist" – so our problems with interpreting this sentence are unlikely to be due to infelicity or ignorance on the part of the writer. In fact, when the sentence is put back into context, it makes unremarkable good sense, in a way that can only be explained in terms of exploitations of norms. The extended context is given in 2.

> Stuart needlessly scraped a fetid plastic comb over his cranium. 'Where are you going? You know, just in case I need to get in touch.' 'State secret. Even Gillie doesn't know. Just told her to take light clothes.' He was still smirking, so I presumed that some juvenile guessing game was required of me. I hazarded various Stuartesque destinations like Florida, Bali, Crete and Western Turkey, each of which was greeted by a smug nod of negativity. I essayed all the Disneylands of the world and a selection of tarmacked spice islands; I patronised him with Marbella, applauded him with Zanzibar, tried aiming straight with Santorini. I got nowhere.

At least two kinds of linguistic exploitation are present here. The first is ellipsis. Having just said that "some juvenile guessing game was required," the speaker does not need to repeat the word *guess*. This particular exploitation rule (ellipsis) is then promptly repeated in five subsequent clauses, in each of which a noun denoting a location or type of location (Disneylands, spice islands, Marbella, Zanzibar, Santorini) is (in its particular context) elliptical for a speech act (a guess) referring to that location. Moreover, a secondary exploitation of considerable complexity is involved in "tried aiming straight": normally, you aim a gun straight at something, you aim (or fire) a question at someone; you don't aim straight at a destination. However, it is noteworthy that, once the scenario has been set up, these stylistic complexities do not distract from the comprehensibility of the text. No ordinary human reader puzzles over what was being essayed or aimed at.

The second directly relevant kind of exploitation that enables a reader to understand Barnes's sentence is lexical creativity involving a combining form. Nowhere else in the novel, nowhere else in the BNC, have I been able to find the word *Stuartesque*. It is used occasionally in texts found on the Internet, with the meaning 'characteristic of Stuart', referring in each case to a completely different Stuart. But that does not make it meaningless or (in context) hard to understand. Stuart is the name of a character in the novel, and

in English the combining form *-esque* is regularly affixed to a proper name to form an adjective meaning 'resembling the person or place named, typically in respect of certain noticeable and even eccentric or bizarre characteristics': *Kafkaesque*, *Bergmanesque*, *Monroesque*, *Hollywoodesque*, *Dylanesque*, *Jaggeresque*, *Caravaggesque* are just a few of the examples of such creations attested in the Oxford English Corpus.

There are many other kinds of exploitation, notably freshly created figurative language, that cannot be discussed here. It can be hoped that this brief discussion will be sufficient to illustrate the kinds of challenges that lie in wait for lexicographers of the future, if they decide to take seriously the need to investigate the relationship between phraseology and word meaning.

3.   Acknowledgements

4.   References

4.1   Dictionaries

Calepino, A. (1502): *Dictionarium*. Very many subsequent editions and multilingual elaborations, compiled by various hands and printed in various European cities. Reggio nell'Emilia.

Cawdrey, R. (1604): *A table alphabeticall of hard usual English words...* London.

Dasypodius, P. (Peter Hasenfus) (1535-1536): *Dictionarium latinogermanicum et vice versa germanicolatinum.* Strasbourg.

Elyot, Sir T. (1538): *Dictionary* [Latin-English Dictionary]. London: Printed by Thomas Berthelet.

Estienne, R. (1531): *Dictionarium, seu Thesaurus Linguae Latinae*. Paris.

Estienne, R. (1539): *Dictionnaire francoislatin.* Paris.

Estienne, R. (1552): *Dictionarium Latino-Gallicum.* Geneva.

Estienne, H. II. (1572): *Thesaurus Linguae Graecae.* Paris.

Florio, J. (1598): *A worlde of wordes, or most copious and exact dictionary in Italian and English.* London: printed by Arnold Hatfield for Edw. Blount.

Galfridus Anglicus (alias Galfridus Grammaticalus; Geoffrey the Grammarian) (c1440): *Promptorium Parvulorum* ['The Young Person's Storeroom']. Printed in 1499 by Richard Pynson and subsequently by others.

Hulsius, L. (Levin van Hulse) (1596): *Dictionaire* [sic] *François-Allemand* [French-German dictionary]. Nürnberg.

Maaler, J. (1561): *Die Teütsch Sprach: Dictionarium latinogermanicum novum.* Zürich.

Minsheu, J. (1599): *Dictionarie in Spanish and English.* London.

Palsgrave, J. (1530): *Lesclaircissement de la langue francoyse.* London.

Rottweil, A. v. (1477): *Introito e porta.* [Italian-German thematic dictionary].

Sinclair, J./Patrick, H. (eds.) (1987): *Collins Cobuild English language dictionary.* London/Glasgow: Collins.

Thomas, T. (1587): *Dictionarium linguae Latinae et Anglicanae.* Cambridge.

Torrentius, H. (Hermann van Beck) (1498): *Elucidarius carminum et historiarum vel vocabularius poeticus.* Deventer.

Van der Schueren, G. (1477): *Vocabularius qui intitulatur theuthonista vulgariter dicendo Duytschlender.* [Dutch-Latin, Latin-Dutch glossary]. n.p.

4.2   Other works

[Anon] (1554): *A very profitable book to learn the manner of reading, writing and speaking English and Spanish.* London.

Armstrong, E. (1954): *Robert Estienne, Royal Printer: an historical study of the elder Stephanus.* Cambridge: Cambridge University Press.

Bately, J. (2009): Bilingual and multilingual dictionaries of the Renaissance and early seventeenth century. In: Cowie, A. (ed.): *The Oxford history of English lexicography*, vol. 1. Oxford: Oxford University Press.

Burke, P. (2002): *Languages and communities in early modern Europe.* (= The Wiles Lectures, Queen's University, Belfast). Cambridge: Cambridge University Press.

Considine, J. (2008): *Dictionaries in early modern Europe: Lexicography and the making of heritage.* Cambridge: Cambridge University Press.

Cowie, A (ed.) (2009): *The Oxford history of English lexicography.* Oxford: Oxford University Press.

Eisenstein, E.L. (1979): *The printing press as an agent of change: Communications and cultural transformations in early-modern Europe.* Cambridge: Cambridge University Press.

Fried, D. (2007): Defining courtesy: Spenser, Calepine, and Renaissance lexicography. In: *Review of English Studies* 58, 229-244.

Firth, J.R. (1957): *Papers in Linguistics 1934-1951.* Oxford: Oxford University Press.

Geeraerts, D. (2010): *Theories of Lexical Semantics*. Oxford: Oxford University Press.

Halliday, M.A.K. (1961): Categories of the theory of grammar. In: *Word* 17 (3), 241-292.

Hanks, P. (1994): Linguistic norms and pragmatic explanations, or why lexicographers need prototype theory and vice versa. In: Kiefer, F./Kiss, G./Pajzs, J. (eds): *Papers in Computational Lexicography: Complex '94*. Budapest: Research Institute for Linguistics, Hungarian Academy of Sciences.

Hanks, P. (2004): Corpus pattern analysis. In: Williams, G./Vessier, S. (eds.): *Proceedings of the Eleventh Euralex International Congress*. Lorient: Université de Bretagne-Sud, 87-98.

Hanks, P. (2010): Nine issues in metaphor theory and analysis. In: *International Journal of Corpus Linguistics* 15 (1), 133-150.

Hanks, P. (2013): *Lexical Analysis: Norms and Exploitations*. Cambridge, MA: MIT Press.

Hanks, P./Jezek, E. (2008): Shimmering lexical sets. In Bernal, E./De Cesaris, J. (eds.): *Proceedings of the XIII EURALEX International Congress*. Barcelona: Universitet Pompeu Fabra, 391-402.

Hanks, P./Pustejovsky, J. (2005) A pattern dictionary for natural language processing. In: *Revue Française de Linguistique Appliquée* 10 (2), 63-82.

Hausmann, F.-J. (1984): Das erste französisch-deutsch Wörterbuch: Levinus Hulsius' *Dictionnaire* von 1596-1607. In: *Zeitschrift für romanische Philologie* 100/3-4.

Jezek, E./Hanks, P. (2010): What lexical sets tell us about conceptual categories. In: *Lexis* 4, 7-22. Internet: http://screcherche.univ-lyon3.fr/lexis/IMG/pdf/Lexis_4.pdf (last visited: 17/05/2013).

Kilgarriff, A./Rychly, P./Smrz, P./Tugwell, D. (2004): The Sketch Engine. In: Williams, G./Vessier, S. (eds): *Proceedings of the Eleventh Euralex International Congress*. Lorient: Université de Bretagne-Sud, 105-116.

Korshin, P.J. (1974): Johnson and the Renaissance dictionary. In: *Journal of the History of Ideas* 35: 2, 300-312.

Kramer, U. (2006): German lexicography. In: Brown, K. (ed.): *Encyclopedia of language and linguistics*. 2nd edition. Amsterdam: Elsevier, vol. 5, 45-52.

Labarre, A. (1975): *Bibliographie du Dictionarium d'Ambrogio Calepino*. (= Bibliotheca bibliographica Aureliana 26). Baden-Baden: Koerner.

Lancashire, I. (2006): Computing the lexicons of early modern English. In: Renouf, A./Kehoe, A. (eds): *The changing face of Corpus Linguistics*. (= Language and Computers 55). Amsterdam: Rodopi, 45-62.

Landau, S.I. (2001): *Dictionaries: The art and craft of lexicography*. 2nd edition. Cambridge: Cambridge University Press.

Murray, J.A.H. (1878): *Presidential address*. The Philological Society.

Murray, J.A.H. (1900): *The evolution of English lexicography.* (= The Romanes Lecture). Oxford: Clarendon Press.

Pustejovsky, J. (1995): *The generative lexicon*. Cambridge, MA: MIT Press.

Pustejovsky, J./Hanks, P./Rumshisky, A. (2004): Automated induction of sense in context. In: *Proceedings of the 20th international conference on Computational Linguistics*. Internet: http://acl.ldc.upenn.edu/C/C04/C04-1133.pdf (last visited: 17/05/2013).

Putnam, H. (1970): Is semantics possible? In: *Metaphilosophy* 1 (3), 187-201.

Putnam, H. (1975a): The meaning of "meaning". In: *Minnesota Studies in the Philosophy of Science* 7, 131-193.

Putnam, H. (1975b): *Mind, language, and reality: Philosophical papers.* Cambridge: Cambridge University Press.

Rosch, E. (1973a): Natural categories. In: *Cognitive Psychology* 4 (3), 328-350.

Rosch, E. (1973b): On the internal structure of perceptual and semantic categories. In: Moore, T. (ed.): *Cognitive development and the acquisition of language.* New York: Academic Press, 111-144.

Schoonheim, T.H./Pijnenburg, W.J.J. (2006): Dutch and Flemish lexicography. In: Brown, K. (ed.): *Encyclopedia of language and linguistics*. 2nd edition. Amsterdam: Elsevier, vol. 4, 20-23.

Sinclair, J. (1966): Beginning the study of lexis. In: Bazell, C./Catford, J./Halliday, M./Robins, R. (eds.): *In Memory of J.R. Firth*. London: Longman, 410-431.

Sinclair, J. (ed.) (1987): *Looking up: An account of the COBUILD project in lexical computing*. London: Collins ELT.

Sinclair, J. (1991): *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Sinclair, J. (1998): The lexical item. In: Weigand, E. (ed.): *Contrastive Lexical Semantics.* (= Amsterdam Studies in the Theory and History of Linguistic Science: Ser. 4, Current Issues in Linguistic Theory 171). Amsterdam: John Benjamins, 1-24.

Sinclair, J. (2004) *Trust the text: Language, corpus and discourse.* London: Routledge.

Starnes, DeWitt T. (1963): *Robert Estienne's influence on lexicography*. Austin: Texas University Press.

Starnes, DeWitt T./Noyes, G.E. (1946): *The English dictionary from Cawdrey to Johnson.* Chapel Hill: University of North Carolina Press.

Starnes, DeWitt T./Talbert, E.W. (1955): *Classical myth and legend in Renaissance dictionaries.* Chapel Hill: University of North Carolina Press.

Stathi, E. (2006): Latin lexicography. In: Brown, K. (ed.): *Encyclopedia of language and linguistics*. 2nd edition. Amsterdam: Elsevier, vol. 6, 723-724.

Stubbs, M. (2001): *Words and phrases: corpus studies of lexical semantics.* Oxford: Blackwell.

Tesnière, L. (1959): *Éléments de syntaxe structurale.* Paris: Klincksieck.

Trench, R.C. (1858): On some deficiencies in our English dictionaries. In: *Proceedings of the Philological Society*.

Watson, F. (1908 [1968]): *The English Grammar Schools to 1660: their curriculum and practice.* London: Cass (Reprint 1968).