



Universiteit  
Leiden

The Netherlands

## **These kind of words: number agreement in the species noun phrase in international academic English**

Stenton, A.J.

### **Citation**

Stenton, A. J. (2024, September 3). *These kind of words: number agreement in the species noun phrase in international academic English*. LOT dissertation series. LOT, Amsterdam. Retrieved from <https://hdl.handle.net/1887/4039590>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4039590>

**Note:** To cite this publication please use the final published version (if applicable).

# 5 The corpus analysis

## 5.1 Introduction

In Chapter 2, I showed how modern reference grammars analyse number agreement in the species noun phrase (SNP), and in the clause of which it is a constituent. In Chapter 3, I analysed the exemplifications (proscriptions), recommendations (prescriptions), and explanations of the variant usages of the species noun phrase presented by the usage guide writers between 1770 and 2010. In Chapter 4, I investigated the acceptability judgements of respondents to an attitude survey of academic usage of the species noun phrase. In this chapter, I analyse a corpus of academic writing to determine how the variants of the species noun phrase are actually used in a sample of manuscripts submitted for publication to a number of academic journals. This is therefore the second part of my investigation into the views of the general public, and constitutes the usage survey. I start by describing how the corpus that I have used – the Stenton Corpus, which was also the source of the examples in Chapter 4 – was compiled, and how the various concordance files were extracted from it (§§5.2, 5.3), before presenting my analyses (§§5.4, 5.5). Specifically, I investigate whether the academic authors in the Stenton Corpus tend to avoid the variant *these kind of N2*, as proscribed in the usage guides, or whether, as described by Denison (2002), Keizer (2007), Klockmann (2017a), and others, they in practice treat *these kinds of N2* and *these kind of N2* not as variant usages but as different structures with different functions (see §2.4). In the course of the analysis, I will present a further variant of the species noun phrase, not covered in the usage guides and the grammars – *this N2 kind* – and investigate how this variant is used by the authors to avoid number mismatch in the species noun phrase (§§5.4, 5.5). Also investigated is whether, as mentioned in both the usage guides and the reference grammars, the authors make use of the *N2 of this kind* variant to avoid a number mismatch in the species noun phrase. As part of the investigation, I look at the relative frequencies of the three variant forms (§5.5.2), and at number agreement of the species noun phrase as subject with the verb in the clause (§5.5.3). These investigations also aim to shed light on what the academic authors in practice treat as the head of the species noun phrase, a topic discussed in §2.3.3. The concluding remarks (§5.6) will present an overview of the use of the species noun phrase by the academic authors in the Stenton Corpus.

## 5.2 The Stenton Corpus

The corpus used for this study – the Stenton Corpus – is what McEnery and Hardie (2012, p. 11) term an “*opportunistic corp[us]*”, in that it consists of “nothing more nor less than the data that it was possible to gather for [this] specific task”. In this case, the corpus consists of 1,031 unedited manuscripts (mss) accepted for publication in six academic journals published by Cambridge University Press (CUP), in Cambridge, England, over the period 2006 to 2016. Once submitted and accepted, the copyright of these mss passed to CUP, who in turn granted me permission to use them anonymously for analysis. Whether the fact that the mss were published in England makes it a corpus of British English is discussed below (§5.2.3). This corpus was also the source of the examples used in the survey described in Chapter 4. The mss comprising the corpus are discussed in detail below.

### 5.2.1 The journals

The manuscripts in the corpus were submitted for publication to the following six journals, listed with the dates of mss received, number of mss, and word counts in parentheses:

*Asian Journal of International Law*

(AJL: 2011–2016; 104 mss; 749,000 words)

*Asian Journal of Law and Society*

(ALS: 2013–2014; 21 mss; 179,000 words)

*International Journal of Law in Context*

(IJC: 2007–2015; 219 mss; 1.65 million words)

*Bilingualism: Language and Cognition*

(BLC: 2009–2011; 48 mss; 397,000 words)

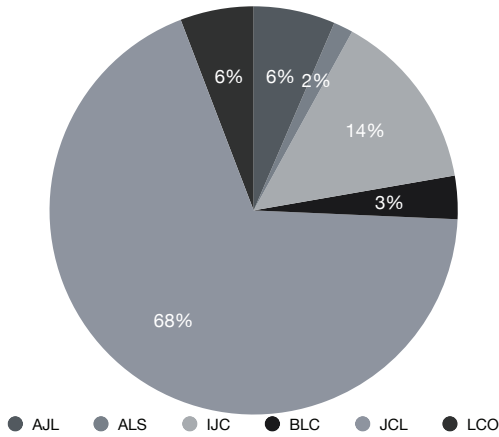
*Journal of Child Language*

(JCL: 2006–2016; 555 mss; 7.93 million words)

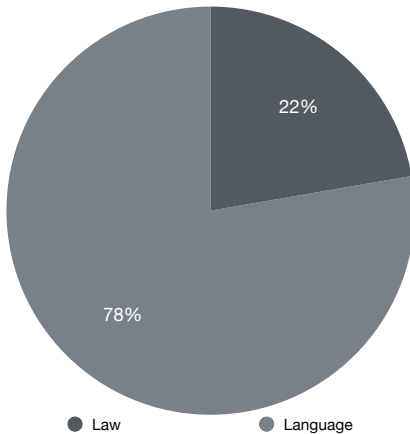
*Language and Cognition*

(LCO: 2013–2016; 84 mss; 674,000 words)

There are thus three Law journals (the Law sub-corpus) and three Language journals (the Language sub-corpus). The total word count of the corpus is 11.58 million. The Law sub-corpus contains 2.58 million words (22% of the whole corpus), and the Language sub-corpus 9 million words (78% of the whole corpus). The breakdown of the corpus by journal is shown in Figure 5.1, whilst the breakdown by sub-corpora is shown in Figure 5.2. What Figures 5.1 and 5.2 make clear is that the Stenton Corpus is weighted quite heavily towards the Language sub-corpus, and also quite heavily



**Figure 5.1 Journals by word count**



**Figure 5.2 Law and Language sub-corpora by word count**

towards one journal – JCL – and these are aspects of this ‘opportunistic’ corpus that will be kept in mind in the analyses that follow.

The corpus figures were all arrived at with the help of Sketch Engine (see §5.3 below).<sup>1</sup> I also loaded the corpus files into two freely available concordance programs: AntConc<sup>2</sup> and CasualConc.<sup>3</sup> AntConc calculated a total of 12.88 million words and CasualConc 13.54 million words. Averaging these three totals yields a word count of 12.7 million, and this is the figure that I will use when discussing the corpus more generally. However, when investigating any sub-corpora differences, because of their different sizes I will use either percentages or normalised frequencies per million words based on the Sketch Engine figures (see McEnery and Hardie, 2012, pp. 48–51). To

1 Sketch Engine generates ‘token’ counts for the six sub-corpora, which total 18,340,688. There is a note on the Sketch Engine Corpus info[rmation] page that lists the six .zip files that make up the Stenton Corpus: “The number of tokens is only an estimate. The exact number depends on the corpus configuration. Lines such as <john> are interpreted either as tokens or as structures based on the corpus configuration. Thus, the sum of the values specified in this column may not be exactly the same as the value specified in the list of corpora.” The value specified in the list of corpora is 11,571,305, which is given as the ‘Word’ count on the Corpus info page, and which is more in line with AntConc and CasualConc (see fnns. 2 and 3). I will use this ‘Word’ figure in my analyses.

2 AntConc 3.4.3m (Macintosh) 2014; Tcl 8.5 & Tk 8.5 (8.5.9) © 1987–2017. Online <<https://www.laurenceanthony.net/software/antconc/>> (last accessed 8 July 2022).

3 CasualConc 2.0.6 (201702025) © 2014–2017. Online <<https://sites.google.com/site/casualconc/download>> (last accessed 8 July 2022).

put these figures into some perspective, the first electronic corpus – the Brown Corpus – compiled in the early 1960s, contained 1 million words (Baker, 2010, p. 59; and see §2.3 for the corpora used by the reference grammars), whilst the GloWbE corpus (Corpus of Global Web-based English) currently contains “about 1.9 billion words”.<sup>4</sup> The most pertinent aspects of the Stenton Corpus for this study are that it consists of manuscripts that have been submitted and accepted for publication, but that have not been copy-edited (see §5.2.2 below), and that it does not reflect a single regional variety of English (see §5.2.3 below).

### 5.2.2 *The manuscripts*

The Brown Corpus (cf. §5.2.1 above) is described as containing “Edited American English” (Baker, 2010, p. 59). The 1,031 manuscripts in the Stenton Corpus are not edited, in the sense that they are not copy-edited. The mss were reviewed by the journals’ editorial boards, sent out for blind peer review, and subsequently revised by the authors and re-submitted. All six journals’ ‘Instructions for Contributors’ (IFCs) include something like the following advice:

Authors, particularly those whose first language is not English, may wish to have their English-language manuscripts checked by a native speaker before submission. This is optional, but may help to ensure that the academic content of the paper is fully understood by the editor and any reviewers.<sup>5</sup>

Beyond that, the journal IFCs remind authors that “Spelling should be consistent – either British English or American English throughout”,<sup>6</sup> but specify no more than that. Once the revised mss were approved by the journal editorial boards, they were sent to Stenton Associates<sup>7</sup> for copy-editing, and for subsequent proof-reading and proof collation. I thus had access to each ms at several different stages in its production, i.e. unedited ms, copy-edited ms, unmarked page proofs and marked page proofs. The versions of the mss used in the Stenton Corpus are the unedited mss as received from CUP, and had not, to our knowledge, been professionally copy-edited. This last

4 Online <<https://www.english-corpora.org/glowbe/>> (last accessed 8 July 2022). In this context, a billion is one thousand million.

5 *Asian Journal of International Law*, Notes for Contributors and House Style, 24 April 2017. Online <<https://www.cambridge.org/core/journals/asian-journal-of-international-law/information/instructions-contributors>> (last accessed 8 July 2022). In October 2020, the journal added this note: “We list a number of third-party services specialising in language editing and/or translation, and suggest that authors contact as appropriate.” This advice was not formally given when the mss for this study were collected.

6 *Language and Cognition*, Instructions for Contributors, 30 January 2017. Online <<https://www.cambridge.org/core/services/aop-file-manager/file/575abd7e32fa8bf21d8bd276/LCO-ifc.pdf>> (last accessed 8 July 2022).

7 Stenton Associates, of which I was a partner, specialised in academic journal copy-editing.

point is also the conclusion that we drew from working on the mss, in terms of both structural and content issues, for example missing references, examples not numbered in sequence, a mix of British and American spelling and idiom, incomplete sentences, etc. (and see §1.1). This absence of copy-editing is a major difference between the Stenton Corpus and many other corpora of written English.<sup>8</sup> The significance of this aspect of the Stenton Corpus is that it avoids what Rawlins and Chapman (2020, p. 10) refer to as “one of the weaknesses of corpus research – many of the texts in the corpora have been edited, thereby giving the attitudes and practices of copy-editors an outsized influence in the published language”.<sup>9</sup>

The mss files in the Stenton Corpus have also not been ‘tidied up’ in any way. This is in line with Sinclair’s “major principle of respecting and trusting the integrity of the complete text as the basis for linguistic description, analysis and theory building”, as reported by Carter (2004, p. 5). McEnery and Hardie (2012) also comment on this:

... the common process of deleting pictures, tables and other non-paragraph material from corpus texts is clearly a much grosser violation of the text than introducing annotation can conceivably be. Yet this is a violation that a great majority of corpora, including the Bank of English, have carried out and will in all likelihood continue to carry out. Given that such deleted items may materially alter the interpretation of the remaining text, it is curious that such a violation has been overlooked.

(McEnery and Hardie, 2012, p. 155)

Following this approach, the mss in the Stenton Corpus still contain, for example, headings, footnotes, references, tables and figures, so any examples of the species noun phrase that occur in these parts of the text will be captured in the corpus search. The issue of introducing annotation, raised by McEnery and Hardie above, is addressed below in the section on Sketch Engine (§5.3).

### 5.2.3 *The authors*

In terms of the overall number of authors in the Stenton Corpus, there are 1,687 listed for the 1,031 manuscripts. Some authors wrote for more than one journal (for example, one IJC author also wrote for ALS, and several JCL authors also wrote for BLC (=20) or for LCO (=7)), so in order to arrive at the number of unique authors in the corpus,

8 Jonathon Owen notes that at Brigham Young University in Utah there is a Faculty Editing Service [now called the Faculty Publishing Service], of which he says: “The purpose of the Faculty Editing Service is to clean up manuscripts before they are submitted to academic journals, so the editing done by the interns and their supervisors at the Faculty Editing Service is probably lighter than what would normally be found in manuscripts edited by publishers” (2020, p. 294). The interns are “probably all students in the editing minor” course at the university (2020, p. 294). The service is now (April 2021) a paid-for service, similar to that referenced in fn. 5 above. I am not aware of this service being offered at other institutions, and the addition to the information in fn. 5 would seem to support this.

9 See Lukač and Stenton (2023) for a survey of copy-editors and proof-readers.

an author who contributed to more than one journal was counted only once. This resulted in a total of 1,657 different authors, with 337 different authors writing for the Law journals and 1,320 different authors writing for the Language journals. Although these absolute numbers look very different, when the figures are normalised there are 131 authors per million words in the Law sub-corpus, and 147 authors per million words in the Language sub-corpus. Looking at these figures slightly differently, each Law author contributed an average of 7,656 words to the sub-corpus, whilst each Language author contributed an average of 6,818 words to the sub-corpus, so the two sub-corpora are broadly comparable in terms of authorship.<sup>10</sup>

As mentioned above (§5.2.2), one aspect of the Brown Corpus is that it presents one regional variety of English: American English. What this typically means is that the texts making up the corpus were published in the United States, or were written by native speakers of American English; similarly, for the LOB Corpus, set up to provide a British English comparison corpus to the Brown Corpus,<sup>11</sup> the texts making up this corpus were published in the United Kingdom, or written by native speakers of British English. Although all six journals in the Stenton Corpus were published in the UK, the language variety of the Stenton Corpus is much more difficult to determine. As noted above (§5.2.2), contributors to the journals used in the Stenton Corpus were advised to use either British or American spelling, but this in itself is not a sufficient basis for allocating the mss to either British or American English. This was a problem recognised by the team at Lancaster University when they were preparing the British National Corpus 2014 in collaboration with Cambridge University Press.<sup>12</sup> The compilers of the corpus noted that, whilst a published text by a single author could be identified as being written by a native speaker of British English, other types of publication were identified as being British English if they were, for example, published in the UK, or if the website source had a .uk ending. The primary concern of the BNC 2014 compilers was in not identifying any source material as British English based on the spelling practice followed in it, as this would lead to circularity, in that patterns of e.g. British spelling would be determined by analysing texts which had been selected on the basis of their British spelling.<sup>13</sup>

10 The log-likelihood value of +3.69 shows the differences to be non-significant (see fn. 26 below).

11 According to Baker (2010, p. 59), the Brown corpus “was followed by a corpus-building project which took place over the 1970s, with the aim of building a British equivalent of the Brown corpus, using texts from 1961 as the creators of Brown had done. This project involved collaboration between Lancaster University, the University of Oslo and the Norwegian Computing Centre for the Humanities at Bergen, and the resulting corpus was therefore known as LOB (Lancaster Oslo-Bergen). LOB followed the same sampling framework as Brown, also having a million words from fifteen genres of writing (consisting of 500 samples each of about 2,000 words).”

12 See <<http://corpora.lancs.ac.uk/bnc2014/>> (last accessed 5 July 2022).

13 Róisín Knight, personal communication, 3 November 2015.

The mss used for the Stenton Corpus were all produced for publication in the UK, but the only information that is available about the authors is their institutional affiliation at the time the ms was submitted, for example “Behavioural Science Institute, Radboud University Nijmegen – the Netherlands”, and this clearly in itself does not imply that the author was Dutch. There is no information about the nationality, age or gender of the authors, and none on their native language(s). To try to off-set this lack of information, I compiled a list of the affiliations by country of all listed authors.<sup>14</sup> The list contains 2,261 affiliations<sup>15</sup> in fifty-nine countries.<sup>16</sup> The details are set out in Appendix F, Table F3. Seven of the countries listed in Table F3 fall within what Trudgill and Hannah (2017, p. 12, Map 1.1) call “[n]ative English-speaking areas” (Australia, Canada, Ireland, New Zealand, South Africa, the UK and the USA), and a further eight countries fall within their “second-language varieties of English” (ESL) (2017, pp. 128–145) (Bangladesh, India, Kenya, Nigeria, Pakistan, the Philippines, Singapore and Sri Lanka). This list does not demonstrate that the authors based in these countries were native or non-native speakers of any particular variety of English, but it does indicate that at that time they were working in an at least partly English-speaking environment. Appendix F gives a detailed break-down of the authors, including numbers of contributing authors per journal (Table F1), numbers of papers contributed per journal (Table F2), and the country of institutional affiliation by journal (Table F3).

For the Stenton Corpus as a whole, four of the seven native English-speaking areas provided the most authors: the United States (686), the United Kingdom (309), Canada (149), and Australia (143). New Zealand provided 11 authors, Ireland 8, and South Africa 6. The eight ESL countries between them provided a total of 52 authors, with Singapore providing the bulk of those at 36. The seven native English-speaking areas thus provided 59% (1322/2261<sup>17</sup>) of the authors by affiliation, with the ESL areas providing a further 2% (52/2261). In addition, several other countries provided more than 1% (23 authors), in some cases many more than the native English-speaking areas: Germany (116), the Netherlands (96), Italy (64), France (57), Israel (57), Spain (49), Belgium (38), China (38), Hong Kong (31), Denmark (30), and Finland (24). These eleven countries between them provided 27% (600/2261) of the authors by affiliation. The remaining thirty-six countries provided fewer than 23 authors each, with thirty-two of them each providing fewer than 10. These thirty-six countries

14 133 authors did not list an affiliated institution. These institutions would have appeared in the published versions of the mss, but not always in the mss as submitted for copy-editing.

15 The number of affiliations is different from the number of (different) authors as the affiliations are based on the mss; some authors contributed to more than one ms, and so appear more than once, and some list more than one affiliation.

16 The countries are listed as the authors chose to present them, and so include, for example, Hong Kong, Palestine and Taiwan.

17 This refers to the number of affiliated authors (see Table F3 in Appendix F).

together provided 12% (270/2261) of the authors by affiliation. The Language sub-corpus includes authors from forty-nine of the fifty-nine countries listed, whilst the Law sub-corpus includes authors from thirty-seven of the fifty-nine.

What these figures demonstrate is that it is simply not possible, on the basis of the information provided, to determine which variety of English these authors represent, nor even, of course, whether they are native or non-native speakers. I have therefore not assigned the mss in the Stenton Corpus to the variety of British English, both because of this lack of detailed information about the authors, and notwithstanding the fact that all of the mss were published in England. Instead, following Trudgill and Hannah (2017), I have chosen to label the language of the Stenton Corpus ‘International Academic English’.<sup>18</sup> One consequence of referring to the language in this way, as pointed out by Mauranen (2012, p. 69), is that “[t]here are no native speakers of *academic* language”, and consequently there is no need to try to investigate any potential differences between native and non-native writers in the Stenton Corpus. International Academic English is therefore being treated in this study as a register of Standard English, used by authors world-wide to exchange information and ideas.

### 5.3 Generating the concordances in Sketch Engine

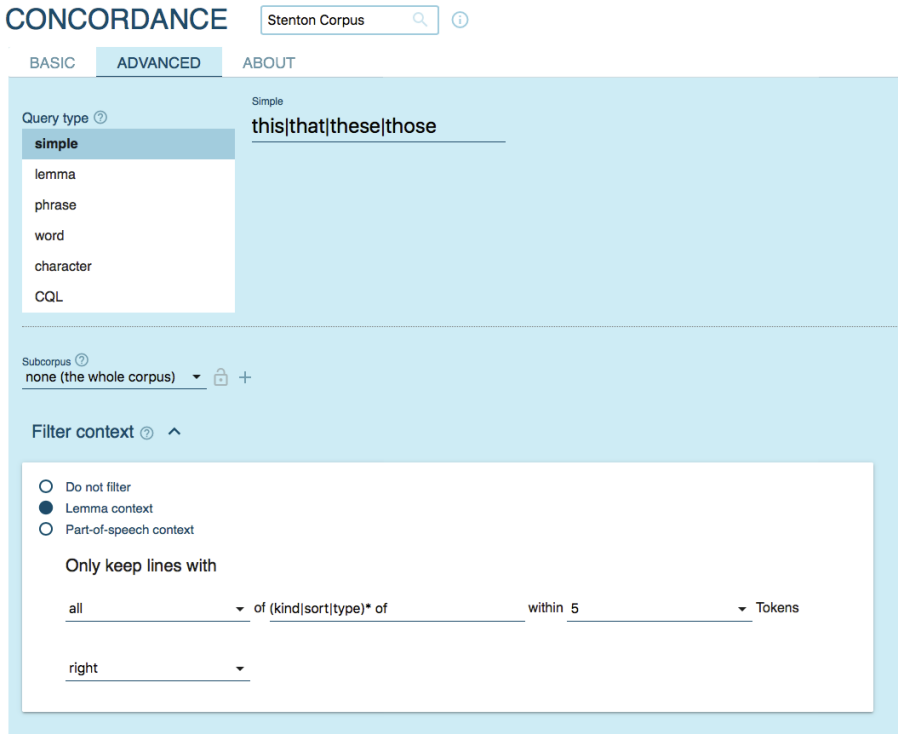
The 1,031 manuscripts comprising the Stenton Corpus were uploaded to the Cambridge English Corpus section of Sketch Engine for analysis.<sup>19</sup> Various codes were automatically added to the texts as they were uploaded, including part-of-speech (PoS) tags. These were not used in the analyses below, as a pilot concordance showed them not to be 100% reliable. In addition, Sinclair (2004, p. 18) has commented on the potential pitfalls of using categories designed for clause analysis in the analysis of corpora. Once the files had been uploaded, a search was made for examples of the species noun phrase, e.g. *THIS TYPE OF*.<sup>20</sup> Figure 5.3 shows the initial search screen.

As I was not looking for just a single word in the corpus but for a string of words of variant forms (e.g. *this type of, these kinds of*), the syntax of the search is important. Although I am looking for a phrase, the initial ‘Query type’ is ‘simple’, as shown in Figure 5.3: ‘*this|that|these|those*’. This is because the phrasal aspect of the species noun phrase is accounted for in the specification of the ‘Filter context’, described below. The use of the pipe symbol, ‘|’, in the search string means that the search will

<sup>18</sup> Mauranen (2012, p. 1) and Crystal (2017, p. 206) prefer “English as a *lingua franca*”, but both of them are writing in the context of spoken academic English. Clearly, I can make no observations about International Academic English as a variety of English on the basis of a study of just one set of usage variants.

<sup>19</sup> Online <<https://www.sketchengine.eu>> (last accessed 8 July 2022).

<sup>20</sup> *THIS TYPE OF* in small capitals will be used throughout this chapter as a shorthand for the variations *this / that / these / those kind(s) / sort(s) / type(s) of; OF THIS TYPE = of this / that / these / those kind(s) / sort(s) / type(s)*.



**Figure 5.3** The Sketch Engine initial search screen for THIS TYPE OF

include *this* or *that* or *these* or *those*. I have selected ‘Subcorpus’/‘none (the whole corpus)’ as at this stage I want to search the whole of the Stenton Corpus, rather than, for example, the Law or the Language sub-corpus. The ‘Filter context’ restricts the selection of ‘this|that|these|those’ to only those strings that also include ‘kind’, ‘sort’ or ‘type’. In this context, presenting this search string within parentheses and separated by the pipe symbol again means that either *kind* or *sort* or *type* will be found (e.g. *this kind*), whilst the asterisk shown in the search string substitutes for any letter immediately following any of the forms within the parentheses. In principle, the asterisk will pick up any letter(s) following *kind*, *sort* and *type*, e.g. *typed*, but in practice the other constraints restrict the search to the plural forms *kinds*, *sorts* and *types* (e.g. *these kinds*, *those types*). The ‘of’ restricts the search to those strings which also include [space +] *of* immediately following the species noun (e.g. *this kind of*, *those sorts of*). The *of* that appears before the parentheses in Figure 5.3 is part of the instruction “all of ... within” and is not part of the search string itself. Finally, setting the ‘Filter context’ to ‘within 5 Tokens right’ will also retrieve those strings where the species noun is separated from the determiner by one or more pre-modifiers (e.g. *two*

*different in these two different types of*). This search string does not include the noun (N2) following the *of*. The inclusion of the N2 is achieved in the specification of the concordance (see below). Figure 5.4 shows the first page of the concordance retrieved by the search string in Figure 5.3.

Figure 5.4 shows a ‘key word in context’ (KWIC) concordance. The search term is centred and highlighted (e.g. *This* in line 1). The filter context terms are also highlighted (e.g. *this type of* in line 2). By setting the ‘context size’ for this search to 60 characters to the left and to the right of the search term (this is not shown in Figure 5.3 as it is specified elsewhere), the concordance includes a line of about 120 characters. ‘Context size’ is defined as “the width of the concordance window in number of characters ... (including positional attributes)” (Thomas, 2016, p. 44); the rider in parentheses in Thomas’s description explains why, if you try to count the context, you will not find exactly 120 characters. In practice, this context is sufficient to include the second noun (N2; e.g. *those types of associations* in line 4), and any of its pre-modifiers (e.g. *this type of express non-justiciable language* in line 13), although it is not always sufficient to include a verb for those examples where the species noun phrase functions as subject (e.g. *this type of legislation can be used* in line 5). In Figure 5.4, the ‘</s><s>’ characters are some of the tags, or HTML codes, generated by Sketch Engine when the files are uploaded, and mark the end/beginning of sentences, respectively. These tags were manually deleted after the concordance was exported from Sketch Engine and imported into a Word file for manual analysis (see below). The ‘file#0’ etc. on the left of each concordance line uniquely identifies the source file, i.e. it is not an example number generated for this concordance. It thus identifies the original file if needed for further reference, for example to check for a verb that is not included in the context to the right of the search term. Given that the items of interest for this study occur largely in the right-hand part of the concordance line, the question might arise as to why I didn’t set the context to, for example, 20 characters to the left and 100 characters to the right, as this would be more likely to include any verbs that would follow a species noun phrase functioning as subject. The reason that I didn’t do so is that the longer context to the left is often useful in confirming that a string is actually an example of a species noun phrase (see below), and in providing an antecedent text referent for the species noun phrase.

The concordance in Figure 5.4 nevertheless does include some strings that are not examples of the species noun phrase, in what seems to be a duplication of some lines. In line 3, for example, *this* is not a determiner but a pronoun, and is not part of a species noun phrase: the concordancing program included pronominal *this* in line 3 because of its proximity to *types of*. Line 3 does, however, include a species noun phrase – *those types of associations* – which is included again in the concordance as line 4. Lines 5 and 6 show a similar relationship, with *that* in line 5 being not a



determiner but a subordinating conjunction, itself introducing a species noun phrase – *this type of legislation* – which is included as line 6. Because of this inclusion in the concordance of lines which are not examples of the species noun phrase, the first task was to remove those examples. This was achieved by exporting the Sketch Engine concordance as a text file, then importing it into Word and manually deleting those examples, e.g. line 7 in Figure 5.4: ... *a reminder that different types of norms* ... .

Another type of duplication occurs because of the nature of academic journal papers. It is quite possible that the same or a similar sentence could occur in the Abstract, the Introduction, the body of the text and the Conclusion. For example, the following two near-identical examples (see the underlined phrases) from the same file would seem to come from an Introduction and a Conclusion, respectively:

- (1) ... Second, we predict that the overall use of lead-in labels should be predictive of low vocabulary, as **these types of labels** require an additional effort on the part of the child to search the environment for what the mother is referring to. ... [file#940|JCL]
- (2) ... we predicted that the overall use of lead-in labels should be predictive of low vocabulary, as **these types of labels** require an additional effort on the part of the child to search the environment for what the mother is referring to. ... [file#940|JCL]

In fact, in this case, the first example comes from an untitled introductory section, whilst the second comes from the Discussion section. Both of these, and other similar examples, were included in the corpus, again following the practice of Sinclair (2004, p. 191) regarding the integrity of the text.

The layout of the Word file was set to mirror that of the on-screen concordance, and this Word file was then used for a manual analysis, i.e. reading through the concordance line by line. The advantage of this type of presentation and manual analysis, instead of asking Sketch Engine to search for and flag any statistically significant patterns, is that unexpected patterns may come to light, or at least patterns that question a conventional analysis (see Sinclair, 1991, p. 44), and in fact one such pattern did emerge (see §5.4.1 below). One disadvantage of this approach is that those patterns might not be statistically significant; another is that the analysis can take a very long time, especially if there are hundreds, or even thousands, of examples. However, unexpected and lower-frequency items are often interesting and sometimes revealing, as will be illustrated in the following section.

## 5.4 Analysing the concordances: some preliminaries

This presentation of the results of the corpus analyses starts with some examples of the different species noun phrase variants. The analyses presented in Chapter 2 provided three canonical examples of number agreement (Quirk et al., 1985, p. 765; and see §2.3):

(3a) Those kinds of parties *are* dangerous.

(3b) That kind of party *is* dangerous.

(3c) Parties of that kind *are* dangerous.

Similar examples are also recommended (prescribed) in many of the usage guides surveyed in Chapter 3 (e.g. Allen, 1999, p. 363; and see §3.4.2):

(4a) this kind of house

(4b) these kinds of [houses]

(4c) [houses] of this kind

The canonical (3a) and (3b) were also the most approved of example types in the usage survey in Chapter 4 (cf. §4.4.2). The search string shown in Figure 5.3 would pick up examples (3a–b) and (4a–b), but not (3c) and (4c), so I also ran a second search for these variants. I therefore worked with two separate concordances.

As I have shown in Chapters 2 and 3, both the grammars and the usage guides offer advice on how to avoid the proscribed *these kind of*. Allen (1999, p. 363) gives examples (4b) and (4c) as “[a]lternatives”, whilst Quirk et al. (1985, pp. 764–765) refer to *These/Those sort/kind/type of parties* as an “idiomatic anomaly”, and suggest that rephrasing, as shown in examples (3a)–(3c) above, can “avoid the anomaly”. The ‘problem’ of number agreement in the species noun phrase is seen as that, in practice, writers, and especially speakers, seem to vary the number marking on the determiner (*this*), the species noun (*kind*), the second noun (*car*) and the verb (*is*), for example using *these kind of cars are unpopular*, with singular *kind*, but plural *these, cars* and *are*. One of the aims of this chapter is to determine the extent to which the academic authors in the Stenton Corpus follow this usage guide advice, either knowingly or unwittingly.

#### 5.4.1 THIS N2 TYPE *as an additional variant of the species noun phrase*

What I was looking for in the concordances were examples of both the prescribed (e.g. *these kinds of, of this kind*) and the proscribed (e.g. *these kind of*)<sup>21</sup> variants. As expected, I found examples of both THIS TYPE OF and OF THIS TYPE, as shown in (5a) and (5b), respectively. I have used examples featuring the N2 *error(s)* in this section to make it easier to identify the other varying forms:

(5a) **This type of error** resembles ... [file#171|BLC]<sup>22</sup>

(5b) **Errors of this kind** are ... [file#162|BLC]

However, in manually checking the concordances, I identified another potential pattern:

(6) ..., **this error type** was included ... [file#572|JCL]

Should (6) also be treated as a variant of the species noun phrase? It includes a determiner (*this*), a species noun (*type*) and an N2 (*error*), but not the preposition *of*. Example (6), repeated below as (7a), can easily be transformed into a canonical species noun phrase, like (5a) and (5b). These transformed examples are shown in (7b) and (7c):

(7a) ..., **this error type** was included ... [file#572|JCL]

(7b) ..., **this type of error** was included ...

(7c) ..., **errors of this type** were included

Examples like (6) occur relatively frequently in the Stenton Corpus, with 105 similar instances, including plural examples in which the number of the verb matches that of the determiner and the species noun, but not that of the N2 *error* (see §5.5.3 below):<sup>23</sup>

(8a) ..., **these error types** were ... [file#507|JCL; file#572|JCL]

(8b) ... **those error types** that lead ... [file#572|JCL]

21 There are no specifically proscribed variants of the OF THIS KIND species noun phrase, as this is generally offered as a prescribed alternative to *these kind of*, for use with a plural N2 (see §3.4.1).

22 The text in brackets shows the file number (as shown in Figure 5.4), plus the name of the journal (as shown in §5.2.1).

23 Note that *error* in *this error type* is still being labelled N2, even though it is positioned before the species noun.

Also, the N2 can be pre-modified (see Appendix G5), with the same agreement pattern as above, i.e. plural determiner, species noun and verb but singular N2:

(9) ... **those two error types** are recognized. [file#734|JCL]

It would seem that in (9) it is the species noun (*types*) that is being pre-modified by *two*, as *error* remains unmarked for number. Quirk et al. (1972, p. 914; 1985, pp. 1331–1332, §17.105) provide some support for this position in their discussion of pre-modification in the noun phrase, where “certain postmodifying *of*-phrases correspond ... to noun premodifiers” (1972, p. 914).<sup>24</sup> They use the examples shown in (10):

(10) The question of partition ~ The partition question

Quirk et al. continue: “plural nouns [i.e. in the *of*-phrase] usually become singular” (1972, p. 914). Is it therefore possible to have a variant of the species noun phrase that does not include the preposition *of*? Keizer (2007, p. 174; see §2.4.2), in analysing what she calls the ‘SKT-construction’, noted that “[it] is clearly not possible for N2 to be omitted”, and also that “leaving out both *of* and N2 does not seem to be an option either”. However, in her corpus she did find examples such as the one in (11):

(11) They won’t last long, mate, *these type* never do.  
(Keizer, 2007, p. 174 (56))

This led her to conclude that the *of* in the species noun phrase is not part of what she terms a “complex postdeterminer” (i.e. *type of*), but is in fact a “separate linking element”, for which there is simply no need “when N2 is left unexpressed” (2007, p. 175). If, instead of being unexpressed, the N2 is positioned before the species noun, then similarly there is no need for the ‘linking element’ *of* (and see Sinclair, 2004, p. 18; 1991, pp. 81–98, for a similar proposition).

More recently, Fontaine and Schönthal (2019, 2020), following Fawcett (2000, pp. 304, 306), analyse *of* in the nominal group as what they call a “selector element” (e.g. 2020, p. 200); similarly, Halliday and Matthiessen (2014, p. 394) treat *of* as a “generalized marker of a structural relationship between nominals”. In these analyses, the inclusion or omission of *of* permits the writer/speaker to adjust which noun they wish to position as head of the noun phrase, i.e. it allows the writer/speaker to distinguish between a referential head (N2 *error* in (5a) above) and a syntactic head (species noun *type* in (6) above). It also potentially allows the writer/speaker to adjust the number marking of the different nouns.

<sup>24</sup> Quirk et al. discuss this in the context of prepositional phrase post-modifiers generally, but I have restricted the analysis to *of*-phrases in the context of my discussion of the species noun phrase.

It could, however, be suggested that e.g. *these sentence types* is not a variant of the species noun phrase at all, but simply a compound noun (*sentence types*) preceded by a determiner (*these*), albeit a compound of which the second noun is always *type(s)*, and weight might be given to this argument in that the examples are restricted to the Language sub-corpus. There is certainly an argument to be made for this analysis. Both Biber et al. (1999, pp. 589–591) and Huddleston and Pullum (2002, pp. 448–451) note that for a compound noun (or ‘composite nominal’ [Huddleston and Pullum] or ‘noun modifier + noun head’ [Biber et al.]), the primary stress would be on the first noun, and that would indeed be the case with e.g. *these sentence types*. Huddleston and Pullum also provide diagnostic tests for the composite nominal, in which the two nouns “can enter separately into relations of coordination and modification” (2002, p. 449), so that *two London colleges* can be coordinated to become *various London and Oxford colleges*, or *various London schools and colleges*, and it can be modified to become *two south London colleges* or *two London theological colleges*, thereby demonstrating the separability of the composite noun structure.

There are no such examples of coordination within this structure in the Stenton Corpus, but there are twenty examples which include modification, with only seven different modifiers: *two, three, four, different, same, initial* and *particular*. All bar *initial* are defined in *COBUILD* in terms of number, and all of those bar *particular* are defined in terms of plural number. The example from the Stenton Corpus with *particular* is the only one to also include a singular determiner and species noun: *this particular sentence type*. The N2, i.e. the noun other than the species noun, is always singular. This would suggest that the plural determiner in all the other examples is modifying either the species noun, or the compound noun: e.g. *these two word types*. This would seem to lend support to the analysis of e.g. *these sentence types* as simply a determiner plus compound noun, rather than as a species noun phrase. However, Biber et al. (1999, pp. 589–590) do note that “the division between a noun compound and a sequence of noun modifier + noun head is in actuality a cline”, and they identify an example – *family member* – with a partitive relation where “[*member*] identifies parts of [*family*]” (1999, p. 591). In terms of the examples of interest for this study, e.g. *these sentence types*, the determiner + species noun identifies a sub-set of the N2 *sentence*. For the current study, therefore, whilst acknowledging that e.g. *these sentence types* may not be unequivocally analysed as a variant of the species noun phrase, it is at least an alternative to the species noun phrase in the sense of providing an option for different number marking, i.e. on the determiner, the species noun and the N2. With that caveat, I have decided to include this construction in my study. The inclusion of this additional construction as a variant thus necessitated the generation

of a third concordance, to ensure that I gathered all the examples in the corpus, rather than just those that were picked up in the generation of the first two concordances.<sup>25</sup>

In the usage guides consulted for this study, Garner (1998, p. 664; 2022, p. 1120) is one of only two to highlight the use of the species noun phrase without *of*, as shown in example (12):

- (12) Councilman Mike Tassin also opposed the project, saying this *type person* [...] does not match others already in the area.

In this case Garner advises revision to *this type of person*. Evans and Evans (1957, p. 525) also list examples of this structure, again only with *type*:

- (13) this type car, that type person

They comment that “[t]his construction does not appear in written English but is too widespread in speech to be called anything but standard” (1957, p. 525), i.e. standard in American English. There are no examples of this structure in the Stenton Corpus.

There are, however, two further possible variants of the species noun phrase, highlighted in bold in (14) and (15):

- (14) ... these were the most frequent types of utterances, and close to 80% of all **tokens** of single-word transitive verbs in the input ***were of this type***. ... [file#810|JCL]
- (15) ... The second type of territorial legislation provides a concurrent alternative to the personal laws. **Of this type**, the most far reaching ***one is*** optional civil marriage and divorce laws, which exists alongside the religious personal laws. ... [file#353|IJC]

In (14) the canonical structure *tokens of this type were* is transformed to place the verb before the species noun, whilst (15) can be seen as a transformation of *this type of one is*. However, there are only two examples like (14) and only one like (15) in the Stenton Corpus, and there are no examples of a pronoun (e.g. *one*) as N2 in the other variants. These potential variants were therefore not considered any further here, but should be noted for inclusion in a subsequent study with a larger or different corpus.

On the basis of the arguments given above, I prepared concordances for three variants of the species noun phrase: THIS TYPE OF N2, N2 OF THIS TYPE and THIS N2 TYPE. These will be described in the following section.

<sup>25</sup> This potential additional variant of the species noun phrase also emphasises the usefulness of a manual concordance analysis (see §5.3).

## 5.5 The corpus data

### 5.5.1 *Frequency of the species noun phrase in the corpus*

The Stenton Corpus of International Academic English contains a total of 1,145 examples of a species noun phrase variant distributed in 501 of its 1,031 manuscripts, so about half of the mss in the corpus contain at least one example. There are 321 examples in the Law sub-corpus and 824 examples in the Language sub-corpus. The corpus as a whole thus contains 99 examples of the species noun phrase per million words (pmw). However, the Law sub-corpus is much smaller than the Language sub-corpus, at 2.58 million to 9 million words (see Figure 5.2 in §5.2.1 above). When these frequency figures for the number of examples are normalised per million words to accommodate the different sizes of the two sub-corpora, the Law sub-corpus is shown to include 124 examples per million words, whilst the Language sub-corpus includes 92 examples per million words, yielding a ratio of 1.3:1, i.e. for each time the species noun phrase occurs in the Language sub-corpus, it occurs 1.3 times in the Law sub-corpus. Put slightly differently, despite there being many more examples in the Language sub-corpus, the species noun phrase is 1.3 times more likely to occur in the Law sub-corpus. This difference can be tested for statistical significance using the online log-likelihood (LL) calculator hosted by the University of Lancaster Department of Linguistics and Modern English Language.<sup>26</sup> The result, i.e. +20.73, shows that the difference in sub-corpora frequency is indeed significant. The following section will examine in more detail the figures for the three main variants of the species noun phrase: *THIS TYPE OF N2*, *N2 OF THIS TYPE* and *THIS N2 TYPE*.

### 5.5.2 *Relative frequencies of the three variants*

As noted above, in the corpus as a whole, there are 1,145 examples of the species noun phrase. These are made up of 847 examples of *THIS TYPE OF N2*, 193 examples of *N2 OF THIS TYPE*, and 105 examples of *THIS N2 TYPE* phrases. These are shown in Figure 5.5 as percentages of the 1,145 total. For the Law sub-corpus, there are 266 examples of *THIS TYPE OF N2* and 55 examples of *N2 OF THIS TYPE*. There are no examples of the *THIS N2 TYPE* in the Law sub-corpus. These numbers are shown in Figure 5.6 as percentages of the 321 examples in the Law sub-corpus. For the Language journals, there are 581 examples of *THIS TYPE OF N2*, 138 examples of *N2 OF THIS TYPE*, and 105 examples of *THIS N2 TYPE*, i.e. all the examples of *THIS N2 TYPE* are in the Language sub-corpus.

<sup>26</sup> This online calculator is hosted at <<https://ucrel.lancs.ac.uk/llwizard.html>>. See McEnergy and Hardie (2012, pp. 48–53) on this use of the log-likelihood test. In this study, I am following the practice of Lukač (2018a, p. 125) in setting the minimum key value to 15.13, corresponding to  $p < 0.0001$  (the 0.01% level, or 99.99th percentile).

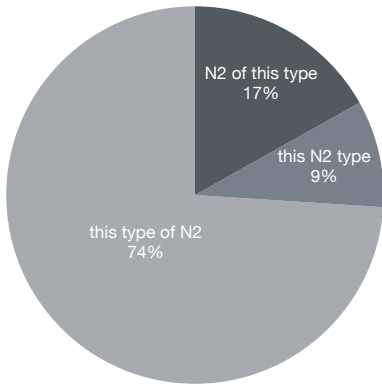


Figure 5.5 The Stenton Corpus

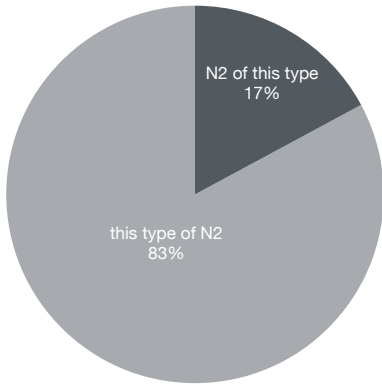


Figure 5.6 The Law sub-corpus

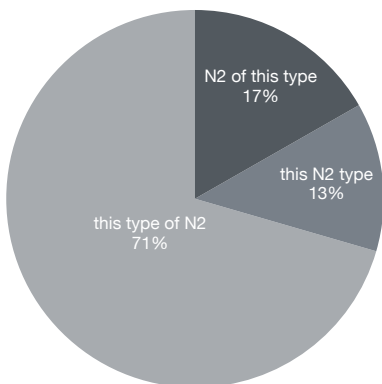


Figure 5.7 The Language sub-corpus

These numbers are shown in Figure 5.7 as percentages of the 824 examples in the Language sub-corpus. When these figures are normalised per million words, we find a ratio in favour of the Law sub-corpus, 1.6:1 for the THIS TYPE OF N2 examples, and 1.4:1 for the N2 OF THIS TYPE examples. These ratios remain unchanged if the THIS N2 TYPE examples are removed from the calculations. These numbers thus show that, despite their lower frequencies, these two variants are more likely to occur in the Law sub-corpus than in the Language sub-corpus, although only the THIS TYPE OF N2 variant is significantly so.

### THIS TYPE OF N2

As can be seen from Figure 5.5, THIS TYPE OF N2 is by far the most frequent of the species noun phrase variants, accounting for 847 of the 1,145 examples (74%). This is also the variant most commonly commented on in the usage guides (see §3.4.1), with their general exemplification (proscription) of *these kind of* as an error, and their recommendation (prescription) of *these kinds of* instead. It is also this variant that was the basis for the attitude survey in Chapter 4. Of these 847 examples, 604 included a singular species noun, and 243 included a plural species noun. I therefore created two sub-concordances to analyse these examples, one listing singular *kind/sort/type*, and one listing plural *kinds/sorts/types*, in order to more easily identify examples of number conflict between the determiner and the species noun, e.g. *these kind of*, as it is this number conflict which is the focus of the usage guide proscriptions. Of the 604 examples of THIS TYPE OF with singular

*kind/sort/type*, only six include a plural determiner, and all six occur with a plural N2, as shown in (16), which also includes a pre-modifier (*representative*) of the N2. This pattern was also noted in the usage guide of Peters (2004, p. 508; and see §3.4.1):

- (16) ... However, legislation surrounding **these type of representative arrangements** must also be constructed in a way that respects the rights in the CRPD and ensures that the individual ... [file#174|IJC]<sup>27</sup>

This is clearly a minority use, but it should be noted that four of the six files which included these instances also contained other examples of the species noun phrase, generally of the same variant, whilst one of them, file#870, contained examples of all three species noun phrase variants, one of only nine files in the whole corpus to do so. What is particularly notable about these three examples in file#870 is that the three species noun phrase variants are used in three different clause functions. Example (17) shows THIS TYPE OF N2 as the object of a preposition (O-P: *for*), (18) shows N2 OF THIS TYPE as the object of a verb (O-V: *could use*), and (19) shows THIS N2 TYPE as subject of the verb (V: *are*):

- (17) ... Prosodic information is often assumed to be a useful basis **for this type of discrimination**, although no systematic studies of the prosodic cues available to infants have been reported. ... [file#870|JCL]
- (18) ... However, there is reason to doubt that infants **could use distributional information of this type** as the initial source of information for discriminating sentences. ... [file#870|JCL]
- (19) ... then, perhaps, provide a foundation for distinguishing wh- questions from declaratives on distributional grounds, as we have found that **those sentence types are** prosodically similar in infant-directed speech, as they are in adult-directed speech. ... [file#870|JCL]

Hoey (2005, p. 43) refers to this type of patterning, where syntactic variants are used in different clause functions, as ‘colligation’, which he traces back via Halliday (1959, p. 46) to Firth (1957, p. 13),<sup>28</sup> and Tieken-Boon van Ostade and Kostadinova (2015) also report colligational patterns in the use of *have went* in American English. However, only nine files in the Stenton Corpus include all three variants, and eight of those do not show such a three-way split of variant and function, so this is not something that can be pursued in the current study.

<sup>27</sup> See also §4.4.4, fn. 27 on this example.

<sup>28</sup> And see Gries (2015, p. 507; also Stefanowitsch and Gries, 2003) on a comparable collexeme analysis as part of a collocation analysis within Construction Grammar.

There were no examples of singular *this/that* with plural *kinds/sorts/types*. File#479 initially looks as though it contains such an example, i.e. (20), but it features a compound determiner linked by *and (this and other)*, and so plural *types* would be expected to follow this:

- (20) ... Further longitudinal study with additional children will be needed to more fully explore the extent of ***this and other types of individual variation***. ...  
[file#479|JCL]

On the basis of this analysis, it would seem that the proscriptions of the usage guide writers against the use of e.g. *these kind of N2* have been successful, in that there are only six examples of it in the whole of the Stenton Corpus of International Academic English (see above), with both sub-corpora almost exclusively featuring number agreement between the determiner and the species noun. However, it is not possible to say from the current study whether this reflects the academic authors following the strictures of the usage guide writers (and the advice in the reference grammars cited in §2.3), or the usage guide writers representing the practices of these academic authors (amongst others), or indeed both of these. This topic will be discussed further in §5.6 below. *These kind of N2* was also the least favoured variant in the attitude survey, with up to 90% of respondents finding it unacceptable.

Unfortunately, this lack of examples of e.g. *these type of N2* vs. *these types of N2* also means that I cannot test the Denison/Keizer Construction Grammar post-determiner hypothesis or Klockmann's Minimalist Program derivational hypothesis, both of which view *these types of* and *these type of* not as variant structures, nor as grammatical/ungrammatical structures, but as different structures, which in Denison's case reflect different semantic and discourse functions (see §2.4.1), and which for Klockmann reflect different syntactic derivations which themselves reflect different features in the lexicon and therefore represent different meanings (see §2.4.4). As noted in §2.4.2, Keizer's corpus didn't contain any examples of e.g. *these kind of* either; the examples she gives are invented.

## **N2 OF THIS TYPE**

Of the 193 examples of the N2 OF THIS TYPE variant species noun phrase, all show number agreement between the determiner and the species noun, whether the N2 is singular (21)<sup>29</sup> or plural (22):

<sup>29</sup> In (21) the N2 is labelled singular rather than unmarked for number as it is preceded by the indefinite article *a*.

- (21) ... **A warning of this sort** can boost the ICC's deterrent effect, as demonstrated by the example from the DRC discussed in the previous section. ... [file#29|AJL]
- (22) ... **Participants of this kind** do not meet the threshold of the mental elements required for accessorial liability—but they would be liable for a number of offences ... [file#6|AJL]

Only 10 of the 193 examples show a plural determiner and species noun (5%), 8 of which are in the Language journals:

- (23) ... they referred to objects that were in the presence of the mother and infant at the time of the labeling utterance, since only **references of these types** allow for the infant to focus on the labeled object. ... [file#940|JCL]

This plural variant is not mentioned in the reference grammars (see §2.3), and Ayto (1995, p. 171; see also §3.4.2) is the only usage guide to list it. As with the THIS TYPE OF N2 variant, it would seem that the academic authors in both sub-corpora again satisfy the prescriptions of the usage guides for number agreement between the determiner and the species noun, but again we don't know who is following whom here.

### **THIS N2 TYPE**

The 105 examples of the THIS N2 TYPE variant are all from the Language sub-corpus, as shown in Figure 5.7 above, and all show number agreement between the determiner and the species noun, either singular (24) or plural (25); the N2s are always singular or unmarked for number:

- (24) ... **This control type** was carried over to the non-cognate set in order to maintain consistency. ... [file#125|BLC]
- (25) ... Mayan children can acquire **these complement types** by analyzing the adult speech they hear. ... [file#400|JCL]

This variant of the species noun phrase does not feature in the usage guides, but it should be noted that the authors again observe number agreement between the determiner and the species noun.

### **Summary**

In total, then, of all 1,145 examples of the three variant structures, only 6 do not follow number agreement between the determiner and the species noun. This being the case, the question might arise as to why the feature is discussed at such length, or indeed at all, in the usage guides. I will return to this topic briefly in Chapter 6.

### 5.5.3 Number agreement with the verb

Previous sections have concentrated on number agreement within the species noun phrase, i.e. between the determiner, the species noun and the N2. Here, I investigate number agreement with a clause element outside the species noun phrase: the verb. Of the three clause functions filled by the species noun phrase – O-P, O-V and SUBJ (see §5.5.2 above) – only that of subject is relevant to the investigation of number agreement with the verb, as it is the number of the subject that either determines or co-varies with the number of the verb (see Quirk et al., 1985, p. 755, and §2.3.2 above). In this section, therefore, I again treat the THIS TYPE OF examples and the THESE TYPES OF examples as separate concordances, as I expect to find that they would typically co-vary with singular and with plural verbs, respectively.

#### THIS TYPE OF N2

In the THIS TYPE OF N2 concordance, there are 280 examples of the species noun phrase as subject. These include 92 examples in the Law sub-corpus and 188 examples in the Language sub-corpus. The normalised frequencies (cf. §§5.2.1, 5.5.1 above) are 36 per million words (pmw) for Law and 21 pmw for Language, and the log-likelihood calculation of +16.48 shows that THIS TYPE OF N2 as SUBJ is statistically more likely to appear in the Law than in the Language sub-corpus. These 280 examples overwhelmingly show singular number agreement throughout (272/280, or 97%), i.e. on the determiner, the species noun, the N2 and the verb, as shown in example (26):

- (26) ... **This type of argument was advanced** by Ukraine against Romania in the Black Sea case. ... [file#0|AJL]

There are only two examples with a plural determiner, eight with a plural N2, and two with a plural verb, and two of those examples include all three, hence only eight examples which show plural. The two examples with plural agreement throughout, apart from the species noun, are shown in (27) and (28):

- (27) ... **These kind of planning requirements make** it virtually impossible for home occupations (sex services) to apply for, let alone receive, development ... [file#309|IJC]
- (28) ... As **these kind of utterances were** the most frequent ones in the data, and, in addition, other single-word sentence frames also had high percentages of ... [file#810|JCL]

These are also the only examples where the species noun does not co-vary with the verb, and so it would seem that here the authors are presenting the N2 as the head of

the species noun phrase. This follows Quirk et al.'s (1985, p. 1238) statement that the head "dictates concord with other parts of the sentence", including the number of the verb in the verb phrase (VP) (1985, p. 755; and see §2.3.2 above), and they may also support the analysis that the determiner can be seen as modifying the N2 rather than the species noun (cf. §2.4). Of the other examples which include a plural N2, four have a verb marked for singular, as in (29):

- (29) ... but rather to be adjectival passives. *This type of passives has also been reported* to be unproblematic ... [file#619|JCL]

and it would seem that here the author is presenting the singular species noun as the head of the phrase, with its anaphoric reference to the single type of *adjectival passives* in the preceding sentence. The final two examples with a plural N2 include a verb which is unmarked for number, so it is simply not possible to assign headship. What these examples show is that the authors in the Stenton Corpus prefer number agreement throughout the species noun phrase functioning as subject, but that in those few cases where they do not, they have the option of marking either the species noun or the N2 as the head of the phrase by adjusting the marking on the verb and/or the N2.

## THESE TYPES OF N2

In the THESE TYPES OF N2 concordance, there are 99 examples of the species noun phrase as subject. These include 27 examples in the Law sub-corpus and 72 examples in the Language sub-corpus. The normalised frequencies (cf. §§5.2.1, 5.5.1 above) are 11 pmw for Law and 8 pmw for Language, and the log-likelihood calculation of +1.36 shows that this difference is not significant. In this concordance, the norm is plural number throughout, i.e. on the determiner, the species noun, the N2 and the verb, as shown in example (30).

- (30) ... somewhat differing patterns of conceptual categorization and construal, and that, in the case of bilinguals and second language learners, **these types of conceptualization differences have** the potential to transfer across languages – or, more precisely, the conceptual distinctions ... [file#131|BLC]

Where this is not the case, it is either because the N2 is either singular or unmarked for number, as in (31), or because the verb is unmarked for number, as in (32) where it functions as a modal auxiliary:

- (31) ... I will argue throughout the article that **these two sorts of representation are** harmful because of the exclusions and inequalities they sustain both for the individuals (actual applicants ... [file#184|IJC]

- (32) General and TPP-specific criticisms of ISDS often focus on its potential negatives without taking into account the potential benefits that **these types of provisions can provide** for investors and, in turn, for governments seeking to encourage inbound foreign investment and ... [file#59|AJL]

There are no examples with a singular determiner, no examples with a singular N2, and none with a singular verb. In terms of identifying the head of the species noun phrase, note that example (31) also includes the species noun pre-modifier *two* (see Appendix G5), which contributes to the plurality of the species noun phrase as a whole. As well as this use of *two*, the determiner (*these*) and the species noun (*sorts*) are both plural, whilst the N2 (*representation*) is unmarked, but the verb (*are*) is also plural, so the number of the verb is showing agreement with the determiner and the species noun, and not with the N2. There are 20 such examples in the THESE TYPES OF concordance, and all show the same number marking. There are no distinctions in the use of the THIS TYPE OF and the THESE TYPES OF variants between the Law and Language sub-corpora.

## N2 OF THIS TYPE

In the N2 OF THIS TYPE concordance, there are 86 examples of the species noun phrase as subject. These include 28 examples in the Law sub-corpus and 58 examples in the Language sub-corpus. The normalised frequencies (cf. §§5.2.1, 5.5.1 above) are 11 pmw for Law and 6 pmw for Language, and the log-likelihood calculation of +4.789 shows that this difference is not significant. Only three of the examples include a plural determiner (all *these*), and those three also include a plural N2, a plural species noun and a plural verb, so the number is consistent throughout. The remaining 83 examples include a singular determiner and species noun, but the number of the N2 is variable, either singular, plural or unmarked, as shown in (33), (34), and (35), respectively:

- (33) ... **a gesture of this type is classified** as an observer viewpoint gesture, because of the hand representing a whole entity. ... [file#1029|LCO]
- (34) ... **Participants of this kind do not meet** the threshold of the mental elements required for accessorial liability—but they would be liable for a number of offences ... [file#6|AJL]
- (35) ... **Understanding of this sort**, however, **develops** only gradually. ... [file#495|JCL]

With this variant, the singular N2 examples, as in (33), all featured the indefinite determiner *a/an*. This is the only variant that includes this. When the N2 is singular, the verb is either singular (=10), as (33), or unmarked (=10) for number. In the plural

N2 examples, as in (34), the verb is either plural (=31) or unmarked (=16). In the unmarked examples, as in (35), the verb is either singular (=13) or unmarked (=7). There are thus no examples which show a conflict of number.

In terms of the headedness of this variant, when the verb is plural the N2 is also plural or unmarked and, apart from the three examples mentioned above, the determiner and species noun in these examples are singular, so for the majority of them the N2 is taken as the head. In those examples where the N2 is singular or unmarked for number, the verb was also either singular or unmarked, as were the determiners and species nouns, so headship remains unclear.

This seems to be the opposite of those examples of the *THESE TYPES OF* variant, where it was the DET + SN that was being taken as the head. However, what both variants have in common is that it is the N which is not part of the prepositional phrase – *of representation* in (35) and *of this kind* in (36) – that functions as head. This will be discussed further in §5.6.

### THIS N2 TYPE

The final variant is the *THIS N2 TYPE*. Of the 105 examples in the Stenton Corpus, 35 function as subject (33%). All of the N2s in this variant are unmarked for number (cf. §5.5.2). The determiners, species nouns and verbs, however, all show variable number. When the species noun is singular (=16), the determiner is singular and the verb is either singular (=11) or unmarked for number (=5). When the species noun is plural (=18), the determiner is plural and the verb is either plural (=14) or unmarked for number (=4). The following examples show singular determiner + species noun with singular verb (36) and with an unmarked verb (37), and plural determiner and species noun with plural verb (38) and unmarked verb (39):

- (36) ... **This control type was carried over** to the non-cognate set in order to maintain consistency. ... [file#125|BLC]
- (37) ... **This event type allowed** the investigation of potential differences in the distribution of attention allocated to the agent and the action in the ... [file#964|LCO]
- (38) ... However, **these error types were** also relatively rare, and **were observed** in only half the children with SLI. ... [file#507|JCL]
- (39) ... **These word types demonstrated** varying levels of phonological and conceptual similarity between L1 and L2. ... [file#470|JCL]

With this variant, then, when the determiner, species noun, N2 and verb are all either singular or unmarked for number, it is not possible to determine whether it is the

species noun or the N2 that is functioning as the head of the clause. However, when the determiner, the species noun and the verb are all marked for plural (=14), then the authors are presenting the species noun as the head of the clause.

## Summary

In all the variants presented above, there was an overwhelming tendency to avoid number mismatch between the species noun phrase and the verb, either by marking all of the elements to agree in number, or by the verb not being marked for number, thereby avoiding number conflict. In those cases where there is not number agreement throughout, there tends to be agreement between the N2 and the verb, perhaps as a result of the authors wanting to mark the N2 as the head of the species noun phrase. In these cases too, however, number conflict is still avoided.

### 5.5.4 kind vs. sort vs. type

I noted in §3.4.2 that the usage guide writers tend to focus on the species nouns *kind* and *sort*. Gilman (1989, p. 577) comments that “although it is seldom mentioned by the handbooks, *type* has fallen into the same sort of pattern: *And in America we don't do those type of things*”. Allen (1999, p. 363) notes that “[t]he issues raised here seem to occur less often in practice with the third alternative, *type of*”. Sayce (2006, pp. 61–63) adds the comment: “When you do use them, note that ‘type’ has a more formal, technical meaning than ‘kind’ and ‘sort’, which are more informal terms.” Sayce’s observation was based on thirty years’ work as an editor, particularly of scientific materials (cf. §3.4.2, fn. 19), and although it is not directly relevant to the topic of number agreement, in this section I address whether Sayce’s comment applies to the writing of the academic authors in the Stenton Corpus.

The figures for the corpus as a whole, shown in Table 5.1 overleaf, do indeed show a preference for *type* (=721), over *kind* (=309), over *sort* (=115): a ratio of 63:27:10%. These figures are based on the number of species noun phrases in the whole corpus: 1,145. A similar pattern appears in both the Law sub-corpus, with its 321 examples: *type* (=138), *kind* (=125), *sort* (=58) (a ratio of 43:39:18%), and in the Language sub-corpus, with its 824 examples: *type* (=576), *kind* (=190), *sort* (=58) (a ratio of 70:23:7%). Sayce’s comment does therefore seem to apply to the use of the species noun phrase in the Stenton Corpus, and more strongly in the Language than in the Law sub-corpus. The normalised frequencies<sup>30</sup> are also shown in the table.

---

30 These frequencies were again calculated using the online log-likelihood (LL) calculator hosted by the University of Lancaster Department of Linguistics and Modern English Language (cf. §5.5.1).

**Table 5.1 Relative frequencies of *kind*, *sort* and *type***

Corpus	Species noun phrases	Species nouns		
		<i>type</i>	<i>kind</i>	<i>sort</i>
Stenton Corpus i.e the whole corpus (11.58 mw)	1145 (99 pmw)	721 (62 pmw)	309 (27 pmw)	115 (10 pmw)
Law sub- corpus (2.58 mw)	321 (124 pmw)	138 (54 pmw)	125 (48 pmw)	58 (23 pmw)
Language sub- corpus (9.00 mw)	824 (92 pmw)	576 (64 pmw)	190 (21 pmw)	58 (6 pmw)
	Log-likelihood	type vs. kind	type vs. sort	kind vs. sort
Stenton Corpus		169.50	489.29	168.70
Law sub- corpus		0.64 (ns)	33.63	25.11
Language sub- corpus		203.72	490.96	74.02
Law vs. Language		type	kind	sort
		3.72 (ns)	47.98	42.60

The normalised frequency figures for the whole corpus are *type* = 62 pmw, *kind* = 27 pmw and *sort* = 10 pmw. For the Law sub-corpus, the normalised frequencies are: *type* = 54 pmw, *kind* = 48 pmw, *sort* = 23 pmw. For the Language sub-corpus, the normalised frequencies are: *type* = 64 pmw, *kind* = 21 pmw, *sort* = 6 pmw. What these figures again show is that, overall, *type* is more frequent than *kind*, which is more frequent than *sort*, and the log-likelihood calculations show these frequency differences to be significant. In the sub-corpora, *type* is 1.2 times more frequent in the Language than in the Law sub-corpus; *kind* is 2.3 times more frequent in the Law than in the Language sub-corpus; and *sort* is 3.8 times more frequent in the Law than in the Language sub-corpus. Log-likelihood calculations show that the differences in the frequency of *type* are not significant, but the differences in frequency of both *kind* and *sort* are significant. In the Stenton Corpus as a whole, then, *type* is used significantly more often than both *sort* and *kind*, whilst *kind* is used significantly more often than *sort*. In the Law sub-corpus, there is no significant difference in the use of *type* and *kind*, but both are used significantly more often than *sort*. In the Language sub-corpus, *type* is used significantly more often than both *sort* and *kind*, whilst *kind* is used

significantly more often than *sort*. The main difference then between the Law and the Language sub-corpora is that in the Language sub-corpus there is a clear distinction between the frequency of use of *type*, then *kind*, then *sort*, as in the corpus as a whole. In the Law sub-corpus, however, there is no significant difference in the frequency of use of *type* and *kind*, although *type* is the more frequent. Comparing the use of *type*, *kind* and *sort* between the two corpora, we find that there is no significant difference in the frequency of use of *type* between the Law and Language sub-corpora. Both *kind* and *sort* are significantly more frequent in the Law than in the Language sub-corpus.

Whilst the usage guide writers tend to focus on *kind* and *sort*, Biber et al. (1999, pp. 256–257) found a preference for *type* in the academic writing in their corpus. This latter preference is strongly echoed in the Stenton Corpus, with a general preference for the use of *type* over *kind* over *sort*. The only exception to this pattern is in the Law sub-corpus, with the OF THIS TYPE examples favouring *kind* over *sort* over *type*. This choice of species noun clearly warrants further investigation.

Further data on the relative frequencies of the constituents of the species noun phrase is given in Appendix G, including the relative frequencies of the species nouns in the three variants of the species noun phrase (Appendix G1), the relative frequencies of the singular vs. plural species nouns (Appendix G2), the relative frequencies of the determiners (Appendix G3), and the number of the N2 (Appendix G4).

## 5.6 Concluding remarks

The data analysed in this chapter comes from the Stenton Corpus (§5.2), a corpus of about 12.5 million words (cf. §5.2.1), based on a collection of 1,031 manuscripts on the subjects of Law and Language by 1,657 different authors, submitted for publication in academic journals published by Cambridge University Press (§§5.2.1–5.2.3). The corpus was searched and the concordances generated in Sketch Engine (§5.3). The corpus analysis introduced some innovations with respect to the use of unedited text in a formal register, that of written International Academic English (§5.2.2), and the identification of a further variant of the species noun phrase – THIS N2 TYPE (§5.4) – a variant which is not treated in either the usage guides or the modern reference grammars, although its inclusion can be questioned (§5.4.1). The use of unedited text is significant as the texts more directly represent the choices of the 1,657 authors, rather than, as in the case of the Stenton Corpus, the choices of the two copy-editors responsible for editing them. The identification of a further variant of the species noun phrase, in addition to the two established variants – THIS TYPE OF N2 and N2 OF THIS TYPE – is important because it further increases the range of options available to those authors for number marking in the species noun phrase. In fact, other potential variants were also identified, but there were too few examples of them to be included

in this study. This lack of examples limits a number of potentially interesting aspects of the analyses, and it seems that a much bigger corpus would be needed to investigate these properly.

The main aim of the analysis was to investigate whether these academic authors in practice followed the advice of the usage guides, as described in Chapter 3, and avoided the use of the proscribed variant *these kind of*, i.e. the use of a plural determiner (e.g. *these*) with a singular species noun (e.g. *kind*). The main findings are summarised below, first those on the corpus as a whole and then those on some sub-corpora differences. Discussion of these results then follows.

- [i] Examples of the species noun phrase were found in 501 of the 1,031 manuscripts. There were 1,145 examples in total. There were more examples in the Language than in the Law sub-corpus (824 vs. 321), but they were statistically more frequent in the Law than in the Language sub-corpus.
- [ii] THIS TYPE OF N2 was the most frequent variant, followed by N2 OF THIS TYPE and THIS N2 TYPE (847 vs. 193 vs. 105). There were more examples of each variant in the Language sub-corpus (THIS N2 TYPE appeared only in the Language sub-corpus), but THIS TYPE OF N2 was statistically more frequent in the Law sub-corpus.
- [iii] The THIS KIND OF N2 examples overwhelmingly showed number agreement throughout the species noun phrase, with only six examples of *these kind of N2*.
- [iv] All the N2 OF THIS TYPE examples showed number agreement between the determiner and the species noun.
- [v] All the THIS N2 TYPE examples showed number agreement between the determiner and the species noun.
- [vi] The THIS TYPE OF N2 variant as subject overwhelmingly showed number agreement throughout, i.e. including the verb (272/280).
- [vii] The THESE TYPES OF N2 variant as subject largely showed number agreement throughout (79/99). In the twenty examples where there was not full agreement, the N2 or the verb was unmarked for number, so there was no number conflict.
- [viii] In the N2 OF THIS TYPE variant as subject there were no examples of a conflict of number.
- [ix] In the THIS N2 TYPE variant as subject, all of the N2s were unmarked for number, and there were no examples of a conflict of number.

There were relatively few significant sub-corpora differences:

- [x] The main difference was that all the THIS N2 TYPE examples came from the Language sub-corpus.
- [xi] Despite this, the species noun phrase was statistically more frequent in the Law sub-corpus.
- [xii] The THIS TYPE OF N2 variant was statistically more frequent in the Law sub-corpus.
- [xiii] The THIS TYPE OF N2 variant as subject was also statistically more frequent in the Law sub-corpus.
- [xiv] On the use of *type* vs. *sort* vs. *kind*, in the corpus as a whole, *type* was used significantly more often than both *sort* and *kind*, whilst *kind* was used significantly more often than *sort*.
- [xv] In the Law sub-corpus, there was no significant difference in the use of *type* and *kind*, but both were used significantly more often than *sort*.
- [xvi] In the Language sub-corpus, *type* was used significantly more often than both *sort* and *kind*, whilst *kind* was used significantly more often than *sort*.
- [xvii] There was no significant difference in the frequency of use of *type* between the Law and Language sub-corpora, but both *kind* and *sort* were significantly more frequent in the Law than in the Language sub-corpus.

It would, I believe, take a much larger corpus, and perhaps one with more subject areas than Law and Language, to determine if any of the differences noted above might be genre differences.

Given that the number of examples like the proscribed *these kind of* was so small as to be negligible, it would seem that the academic authors in the Stenton Corpus did follow the proscriptions of the usage guides, and indeed the guidance of the modern reference grammars described in Chapter 2. However, it is of course simply not possible on the basis of a corpus analysis to say that the academic writers avoided the proscribed uses *because of* the advice in the usage guides. For that we would need to undertake a survey of the authors on their use of reference materials. Further, it is equally likely that the usage guide writers themselves based their proscriptions/prescriptions on the usage of groups such as these academic writers. An analysis of the citation sources used in the seventy-seven usage guides in HUGE as examples of both 'good' and 'bad' usage would be a useful and revealing topic, but once again this was beyond the scope of the current study.

However, to say that the academic authors avoided the overt number conflict of e.g. *these type of* does not mean that they did not have options in how to present different number choices in the species noun phrase. Example (40) shows one such choice, with plural N2 (*errors*) and verb (*were found*) but singular determiner and species noun (*this kind*):

- (40) ... Only **6 errors of this kind**, however, **were found** in the combined corpora of the 7 children Pine & Lieven (1997) analyzed. ... [file#924|JCL]

The THIS N2 TYPE variant, which was used exclusively in the Language sub-corpus and exclusively with the species noun *type*, also permits differences in number marking without number conflict, with singular/unmarked *complement* and plural *these ... types appear* in example (41):

- (41) ... Table 4 lists the contexts where **these complement types appear** in seven Mayan languages. ... [file#400|JCL]

The authors may also emphasise the number of a species noun phrase in other ways, for example with a number pre-modifier of the species noun, e.g. *two* in (42) (and see Appendix G5 for more examples of this):

- (42) ... This matching allowed us to reduce the effects of factors other than familiarity when **these two verb types were compared**. ... [file#736|JCL]

Equally, authors had the option of emphasising the singularity of an N2 using the OF THIS TYPE variant with the indefinite article *a*, as in (43):

- (43) ... **A warning of this sort can boost** the ICC's deterrent effect, as demonstrated by the example from the DRC discussed in the previous section. ... [file#29|AJL]

This option was not available in any of the other species noun phrase variants. It would therefore seem that authors had a number of options available to them, not just to avoid overt number conflict but also to specify number, options which these academic authors did make use of, and these options will be considered further in Chapter 6.

These are the main findings of this chapter, but frequency data on some other aspects of the species noun phrase can be found in Appendix G. These include: the relative frequencies of the species nouns in the three variants (G1); the relative frequencies of the singular and plural species nouns (G2); the relative frequencies of the determiners (G3); the number of the N2 in the three variants (G4); the pre-modifiers used with the species nouns (G5); and parenthetical specifications of the species noun phrase (G6).

As noted above, some of my initial findings warrant further investigation with a larger or different corpus. This would include verifying THIS N2 TYPE as a species noun phrase variant, as well as investigating patterns such as “close to 80% of all *tokens* ... *were of this type*” (§5.4.1, example (14)) and “*Of this type*, the most far-reaching *one is* ...” (§5.4.1, example (15)). There is also scope for further investigation into the use of a pronoun as N2; the co-variance of the N2 with the same (or similar) word in context; which of the two nouns – the species noun or the N2 – typically functions as the head of the species noun phrase; colligation restrictions on species noun phrase variants; and (*these*) *type of* as a post-determiner. However, in order to maintain the authenticity of the corpus, it would need to be compiled from original manuscripts as submitted for publication, rather than those which have been copy-edited. This would necessarily involve the co-operation of a large publishing house, as with the current study.