HISTORICAL ENGLISH PHRASEOLOGY AND THE EXTENDER TAG

Abstract

EXTENDER TAGS are phrases such as *and so forth*, which have also been called SET-MARKING TAGS, VAGUE CATEGORY IDENTIFIERS, LIST EXTENDERS, GENERAL EXTENDERS and LIST COMPLETERS. Overstreet, on the basis of a single token from 1818, asserts, "There is little doubt that ... extenders have been part of the English language for many years" (1990: 6). In fact, extender tags can be found throughout the history of English. This paper surveys previous literature, and supplies new information about these multifunctional phrases, describing their semantic and syntactic behavioral characteristics as PHRASEOLOGISMS and COLLOSTRUCTS. The paper thus provides a framework for their study within the context of phraseology, which allows the consideration of all variants, whether general or specific.

Keywords: extender tags, general extenders, phraseology, history of English, corpus linguistics, fixed expressions, lexicalization, collostructions, collocations.

Resumen

Las EXTENDER TAGS, también llamadas SET-MARKING TAGS, VAGUE CATEGORY IDENTIFIERS, LIST EXTENDERS, GENERAL EXTENDERS y LIST COMPLETERS, son construcciones como *and so forth*. Overstreet (1990: 6), basándose en un ejemplo de 1818, afirma que hay pocas dudas de que tales extensores han sido parte de la lengua inglesa desde hace muchos años. De hecho, éstos pueden encontrarse a lo largo de la historia del inglés. Este artículo repasa la literatura previa sobre el tema y añade nueva información sobre estas construcciones multifuncionales, describiendo sus características de comportamiento semántico y sintáctico como FRASEOLOGISMOS y COLOSTRUCTOS. El artículo de este modo proporciona un marco de trabajo para su estudio dentro del contexto de la fraseología que permite considerar todas sus variantes, sean éstas generales o específicas.

Palabras clave: coletillas extensoras, extensores generales, fraseología, historia del inglés, lingüística de corpus, expresiones fijas, lexicalización, colostrucciones, colocaciones.

1 INTRODUCTION[1]

The study of English historical phraseology can be considered both a very old pursuit (Knappe (2004) reviews scholarly work from as early as the fifteenth century) and an up-and-coming one (Markus urges "that we *start* discovering the ... patterns that were part of [Middle English] speakers' competence" (2008, emphasis added)). As a contribution to this (re-)emerging field, the present paper follows the research agenda set forth by Sinclair (1996/2004) and Stubbs (2002),

[1] This paper draws on presentations made at *Collocations and Idioms 1: The First Nordic Conference on Syntactic Freezes* (2006) and the *20th SELIM Conference* (2008). I thank both audiences and two anonymous reviewers for their comments.

surveying the historical behavior of the extender tag, and establishing a framework for its further study.

The EXTENDER TAG is known by many different names, having been studied by over a dozen linguists, working from varying perspectives, using varied terminology and sometimes quite different definitions. However, in my experience of speaking to both linguists and non-academics about my work, the prototypical members of the class (*or whatever, and so forth*) are easily recognized and acknowledged.[2] Formally, prototypical extender tags consist of a coordinating conjunction (*and* or *or*) followed by a noun phrase which typically includes a semantically empty head (*thing*) and/or a modifier which extends the denotation of the noun (*other*). Functionally, extender tags may be used after a noun phrase to denote a larger semantic category of which the noun is one member (as in example 1, from Old English), or in a list to indicate that other items not already mentioned are nonetheless included (as in example 2, from late Middle English).

(1) *ða gesinhiwan mon sceal manian, <u>& eac gehwelcne mon</u>*
then married-couples one ought to-exhort, and also every man[3] (850–950 HC COCURA 397)[4]

(2) *Crownets of Bayes, of Gold, of Myrtill, <u>or some other thing</u>*
coronets of laurel, of gold, of myrtle, or some other thing (1590 *MEMEM* s. v. *be*)

[2] Usually the prototypical tokens allow the listener to correctly identify the set of linguistic items to which I am referring (EXTENDER TAGS). Occasionally, however, a larger (and partly overlapping) set is understood: VAGUE LANGUAGE MARKERS. As will be noted below, vague language is one important context within which extender tags have been studied.

[3] For clarity, I have given a word-for-word gloss for Old and Middle English examples. I have tended toward cognates and single lexeme glosses where possible.

[4] In all examples I have added underlining of the extender tag for clarity. Most examples, their datings, and the form of abbreviation of the work's title are taken from one of the following sources: The *Middle English Dictionary* (*MED*), the *Middle English Compendium* (*MEC*), *the corpus of Middle English Medical Texts* (*MEMT*), *Michigan Early Modern English Materials (MEMEM), the Helsinki Corpus of English Texts (HC), the Corpus of Early English Correspondence Sampler* (*CEECS*), or *Early English Books On-line* (*EEBO*). Obviously it would be preferable if all examples were from recent scholarly editions rather than mediated through dictionaries and corpora, but such electronic resources are invaluable for allowing searches for specific lexis and word strings. Concerns about the reliability of mediating sources are acknowledged and will be addressed below.

The first part of this paper introduces extender tags. Section 2 reviews the basic literature on them, and section 3 surveys their use in the history of English. The second part of the paper turns to the broader topic of diachronic English phraseology, or the idiomatic principle in language, with section 4 introducing the topic by means of a very brief survey of terminology and foundational literature.

Section 5 points out that extender tags can be studied as a class or alternatively as individual sub-categories of that class. As will become clear, the basic meaning or function of an extender tag is carried not (or not only) by the individual words comprising it, but by the very presence of the tag, whatever its form. Many extender tags are functionally interchangeable. In this sense, it is worth studying extender tags as a COLLOSTRUCT,[5] a whole class of syntactically similar items not linked to an individual lexeme. Although individual extender tags can be seen as functionally interchangeable, they do also have their own peculiarities, such as register restrictions. Individual extender tags can be studied as different phraseologisms, or as sub-categories of a single collostruct.

This latter perspective overlaps with the budding field of corpus semantics, popularized by Stubbs's 2002 textbook. Section 6 demonstrates that all four of the behavior patterns in Stubbs's model can be fruitfully studied with respect to extender tags. Section 7, however, goes on to show that these four are not the only behavior patterns worth observing. Section 8 concludes the paper.

2 EXTENDER TAGS IN SECONDARY LITERATURE

As noted above, the formal characteristics of a prototypical extender tag are that it begins with a coordinator, *and* or *or,* which is followed by an indefinite pronoun or semantically empty noun: *stuff, thing,* or *somebody*, for example. This general word may be modified: *other stuff, somebody like that;* and in some cases even elided: *and all that* [*stuff*] or even *and all* [*that stuff*]. Thus far, most scholars are in agreement, but they part company on further details. Is the main function of such tags to mark a

[5] The term COLLOSTRUCT (Stefanowitsch and Gries 2003) overlaps with the Construction Grammar use of the word CONSTRUCTION (although *collostruct* is more narrowly defined), but is less potentially ambiguous (*construction* is very widely used with other meanings outside Construction Grammar).

list as incomplete? Are they used primarily to signal that a reader should understand a list or single cited item to be representative of a larger set? Or are they better understood as having an interpersonal function, such as marking shared knowledge? As is so often the case in language, there is no one-to-one relationship between this form and its function.

Neither are scholars in agreement over how best to describe either the grammatical or the semantic characteristics of extender tags. Overstreet is one of the few whose study encompasses not only extenders as described above, but also SPECIFIC EXTENDERS (to use her term). Specific extenders, unsurprisingly, are distinguished from general extenders by containing more specific lexis: adjectival modifiers or a relative clause restrict their scope (Overstreet 1999: 12, 51–52), or the head noun may be more specific than *things* or *stuff*. One of my favorite examples of hers is 3:

(3) or any of that whiny crap that's so often mistaken for what makes photography like this "important," or, worse, "courageous."
(1996 *New York Press*, cited by Overstreet 1999: 52)

Specific extenders are abundant in earlier periods of English as well; one example is given in 4:

(4) *& oþer deuoute seyinges of holi kirke*
and other devout sayings of [the] holy church
(?a1475 (a1396) *MED* *Hilton *SP* s. v. *seiing(e* (ger.))

This article therefore encompasses both general and specific extenders; in fact often no distinction is made between them. (This is discussed further in section 5.)

The earliest dedicated study on extenders was published in 1978, a semantic analysis by Ball and Ariel which addresses the need for contextualization in interpreting just what an extender does or does not encompass. Two years later Dines published a sociolinguistic analysis of thirteen extenders in the speech of Australian women. She found that extender tags were more common in working-class speech, and that they were stigmatized for being vague, although Dines pointed out that this was not inherently true, but rather a mythical stereotype (1980: 30).

In 1985 Macaulay noted the use of extenders as an idiosyncratic feature in the speech of a particular Scottish coal miner, who used them much more often than Macaulay's other informants. In the same year, Aijmer

published a very thorough corpus-based study of extenders, including both grammatical and prosodic features in her analysis. Most importantly for my purposes, Aijmer's paper shows an early explicitness about the lexical variation possible in these extenders, such as her statement that in phrases like *and that sort of thing,* "[the word] *Kind* can be exchanged for *sort* without any change in the function of the tag." (Aijmer 1985: 371).

In 1990, Jefferson published a paper which she had first presented orally in 1973, studying lists in natural conversation from the perspective of conversation analysis. The paper observes a strong tendency for lists to be made up of three parts, and analyses extenders as having at least two possible functions in lists —either to provide a third item in a list for which you have only thought of two items, or to abbreviate a list of more than three items down to three. In both cases the result is a list of two items plus an extender, which Jefferson calls a GENERALIZED LIST COMPLETER.

Subsequent studies have usually built on at least one of the works mentioned. Ward & Birner published a semantic/pragmatic study in 1993, sociolinguistic studies were carried out by Youssef (1993) and Stenström et al. (2002), and Overstreet has published widely within the frameworks of pragmatics and discourse analysis (1999, 2005), often in collaboration with Yule (1997a, 1997b, 1999, 2002). Various other linguists have included extenders within larger studies of vagueness in language, including Crystal and Davy (1975: 112 ff.), Meyerhoff (1992), and Channell (1994). Extender tags have also been researched in languages other than English: Dutch (Graman 1998), Montréal French (Dubois 1991), German (Overstreet et al. 2008), Japanese (Lauwereyns 2002), Spanish (Cortés Rodríguez 2006), and Swedish (Norrby 2002).[6]

Table 1 illustrates the wide range of terminology that researchers of extender tags have used. Such a variety exists that this list is restricted to those works making particular reference to extender tags; grammars, for example, are not included here.

[6] Since varied terminology makes it difficult to search for previous literature, I include for reference other relevant studies not cited above: Cucchi 2007, Drave 2002, Evison et al. 2007, Jucker et al. 2003, Lauwereyns 2002, Mittman 2004, Norrby and Winter 2002, O'Keeffe 2003, Sánchez Ayala 2003, Stubbe and Holmes 1995, Terraschke 2007, and Winter and Norrby 2000.

Table 1: Varied terminology for extender tags[7]

Dines 1980 – SET-MARKING TAGS
Macaulay 1985 – TERMINAL TAGS
Aijmer 1985 – UTTERANCE-FINAL TAGS INTRODUCED BY *AND* AND *OR*
Jefferson 1990 – GENERALIZED LIST COMPLETERS
Meyerhoff 1992 – POST-NOUN HEDGES
Dubois 1993 – EXTENSION PARTICLES
Channell 1994 – VAGUE CATEGORY IDENTIFIERS
Overstreet 1999 – GENERAL EXTENDERS
Stenström et al. 2002 – SET MARKERS
Cortés Rodríguez 2006 – ELEMENTOS DE FINAL DE SERIE

I prefer to call them EXTENDER TAGS, but in citing other scholars I may sometimes use their terms, because they do not always overlap exactly. Jefferson's term, *generalized list completers,* reflects her functional definition of what she is interested in, whereas Aijmer's "utterance-final tags introduced by *and* & *or*" is very specifically form-based, as is typical of corpus linguistics. Example 5 meets neither Jefferson's nor Aijmer's criteria, since it is not terminating a three-part list and is not utterance-final, whereas it does meet Overstreet's definition of a general extender:

(5) *like will there still be like, electrical transmissions or whatever going on?* (*MICASE* SEM475JU084)

Often *tag* is reserved for a structure which is found clause-finally (perhaps the most widely studied such structure being the TAG QUESTION). My own use of the term *tags* in *extender tags* should be understood more broadly, as a sequence which follows either a clause or a phrase (compare Biber et al.'s usage of the term COORDINATION TAGS for something which is "added after clauses as well as after noun phrases" (1999: 116), although elsewhere they define a *tag* as a structure "added at the end of the clause" (1999: 139)). The present study encompasses a similarly broad definition of extender tags, in order to provide a framework within which future narrower studies may situate themselves.

[7] Many of these terms are used by more than one researcher and/or in more than one publication, but for the sake of simplicity only one publication is listed per term (and only one term per researcher).

3 Extender tags in early English

All previous published work on extenders has been synchronic, relying on data from the present day (I will use the abbreviation PDE to refer to present-day English).[8] Overstreet cites a quotation in her book from Jane Austen's *Persuasion* (published 1818), which contains two tokens of extenders:

(6) *... they happened to say, that her brother, Captain Wentworth, is just returned to England, or paid off, or something, and is coming to see them almost directly; and most unluckily it came into mamma's head, when they were gone, that Wentworth, or something very like it, was the name of poor Richard's captain ...*
(Austen 1818/1961: 52, cited in Overstreet 1990: 117)

On the basis of this single passage, alongside the 1951 novel *The Catcher in the Rye* (she cites Graman 1998, which identified 635 general extenders in that work alone), she concludes, "There is little doubt that general extenders have been part of the English language for many years" (Overstreet 1990: 6).

Overstreet is right. Both general and specific extenders are attested as far back as Old English, as was shown already in example 1. Examples 7–12 give an indication of the wide range of extenders used in the medieval period, and into the early modern age:

(7) *seinte Sare ... & monie ohre swucche*
Saint Sarah ... and many other such
(c1230(?a1200) *MED Ancr.*, s. v. *swich* (adj.) 5d)

(8) *Tac a lutel radel ant grynt to thin asise ... ant so vorth, as I seyde er.*
Take a little red ochre and grind into thin sizing ... and so forth, as I said earlier
(c1325 *MED Recipe Painting(1)*, s. v. *sō* (adv.) 5(c))

(9) *or he may passe to Ieen or Venice or sum oher*
or he may pass to Genoa or Venice or some other
(?a1425 Hamelius, P. ed. 1919 *Mandeville's Travels* 214/27–28)

[8] My first conference presentation on this subject using historical material was in 2005 (published as Carroll 2007). I am aware of two other scholars now working on these topics with historical data: Ortega Barrera (who delivered a conference paper building on my research in 2008) and Sánchez Barreiro (working within "The Coruña Corpus of English Scientific Writing" project).

(10) *xxi. or xxij. or more*
21, or 22, or more
(1485 EEBO *Here begynnys a schort* ... STC 161:01)

(11) *from Hamborg, and Dansk, Lubeck, &c.*
from Hamburg, and Gdańsk, Lübeck, etc.
(1586 CEECS LEYCESTE 200)

(12) *the said yeerely rent or 200 l., or any part thereof*
the said yearly rent or 200 pounds, or any part thereof
(1666 CEECS COSIN 149)

In 2007 I published a short inventory of the forms of general extenders found in the Corpus of Early English Correspondence Sampler (CEECS), and addressed the problematic nature of the general-specific division, demonstrating it to be more of a cline than a dichotomy. That inventory followed a very strict interpretation of Overstreet's dichotomy, and therefore resulted a very limited set of very short extender tags, eight in total, ranging in length from *&c* to *or such like* (Carroll 2007: 48). However, there is a tendency for shorter tags, and general ones, to occur more frequently than longer or more specific ones, and in that sense the 2007 list remains a useful inventory, since it focuses on the forms which are likely to be more widely represented in the data.

In studying extender tags (those identified in 2007 and others, including longer, more specific ones as well), I have found evidence that supports some of the claims made by previous researchers, but also some evidence that suggests that certain previous claims may be too narrow. The syntactic constraints on extenders are not as restrictive as has sometimes been suggested, and their functions are broader than some researchers have recognized.

Several researchers have asserted that extenders are found only in clause-final position (Overstreet and Yule 2002), or have restricted their studies to clause-final tokens (Aijmer 1985). Although it is true that extenders are common in clause-final position, they are also found elsewhere:[9]

(13) *That no maner of Hostyler, ne other man, horbereth ne reteyn in his hous, man ne woman of yevel nam.*

[9] In 14 the extender ends a list but not a clause. Examples can be found (although they are rare) in which a list contains an extender which does not end the list (one such is cited in Carroll 2007).

that no manner of innkeeper, nor other man, harbor nor retain in his house, man nor woman of evil name
(?a1450 *MED *Procl.Camb* s. v. *herberwen* (v.) 1(c))

(14) *Iob, Iosep, and mony ohere suche weren riche of pite.*
Job, Joseph, and many other such were rich in pity
(c1400 *MED Bk.Mother* s. v. *rīche* (adj.) 1(c))

I will return to the distribution of extender tags in section 6, below.

The second broadening insight offered by the study of historical data is that the form of the extender tag was used in a wider range of functions than some researchers acknowledge. Jucker at al. describe the function of the vague category identifier in this way: "The vague category identifier asks the hearer to identify ... a larger concept that [the speaker] cannot name, either because she does not know a name or because she cannot recall one at the moment" (2003: 1749). Insofar as a VAGUE CATEGORY IDENTIFIER (VCI) is a functional label, this is true. However, the same form which can be used as a VCI can also be used with other functions not accounted for by Jucker et al.

Example 15 shows an extender tag used as a VCI. Lacking a single lexical item to denote "body parts that move by themselves", the writer illustrates the concept by means of an example followed by a VCI:

(15) *ohere membris þat mouen hem bi hemsilf as þe lippis and ohere siche.*
other members that move them by themselves, as the lips and other such
(c1475 *MED *Mandeville* (Wel 564) s. v. *swich* (adj.) 5d (b))

However, in 16, the same extender tag (or one which is identical apart from being disjunctive instead of conjunctive) is used in a context where the set ("evil temptations") has been named with a single lexeme, *fondyng*:

(16) *When þat I fele any fondyng, as ire or wraþ ... or oher siche.*
When that I feel any evil-temptation, as ire or wrath ... or other such
(a1450 *MED PNoster R.Hermit* s. v. ōther (pron.) 4b)

Jucker et al. were building on Channell's work on vagueness, and so are perhaps not interested in examples like 16, where the presence of a lexical name for the set precludes vagueness, but for my purposes it is worth

noting that vagueness, or the filling of a lexical gap, is in any case an incomplete description of the functions of extender tags.

Sánchez Ayala addresses the lexicalization of generalized list completers (relevant to phraseology, see section 4, below), and suggests that the lexicalization came about in the following way: a speaker would search for a term for use in a list (not, as in Jucker's example, an overarching category's name, but a specific word) and not find it. Sánchez's suggestion is that while the speaker is still mentally searching for the term, she or he trots out a semantically empty phrase in order to buy time (*and things like that*), but that in a number of cases no further example is produced, and so over time this phrase comes to signal the end of an incomplete list (2003: 339).

Finding written evidence for any model based on some sort of lexical retrieval problem on the part of the speaker is of course problematic, since a writer can easily put down the pen and think for a moment without "losing the floor", but even so it is strikingly difficult to find support for this scenario. What there is better evidence for, in contrast to the inability to remember what comes next, is speaker fatigue (or, more accurately for my texts, writer fatigue), a lack of inclination to list what comes next. Indeed in example 17, this motivation of speaker fatigue is made explicit:[10]

(17) *And aboute þis ile er many oþer iles and diuerse cuntreez, and diuerse maners of men, of whilk it ware to mykill to speke of all.*
And about this isle are many other isles, and diverse countries, and diverse manners of men, of which it were too much to speak of all
(?a1425 Warner, F. ed. 1889 *The Buke of John Maundeuill* 93/19–20)

This abbreviating function is particularly true of the extenders *and so forth* and *et cetera*:

(18) *othere colours, as is whit and blew or whit and blak or blak and reed and so forth.*
other colors, as is white and blue or white and black or black and red and so forth
((c1390) *MED* Chaucer *CT.Pars.* s. v. *rēd* (n.(2)) (a))

[10] Another motivation for not giving complete lists would be the desire or need to save writing materials.

(19) *and bedding, hangings, &c.; to the Parson of Tasley*
and bedding, hangings, etc.; to the parson of Tasley
(1418 *MEC, Fifty earliest English wills*)

Both of these extenders are also used to abbreviate texts and quotations as well as lists (Carroll 2007), a function which has not been particularly highlighted in the literature on PDE.

Thus we see that the extender tag has indeed been a feature of English throughout its history. A data-driven historical survey of its use shows that its syntactic distribution was wider than previously thought, and its functions more diverse.

As noted already in footnote 4, these examples come from various sources, including historical dictionaries and corpora. Although the search facilities afforded by such resources are invaluable, they must always be treated with caution. For example, the *MED* abbreviates quotations in ways which can be misleading. Example 20, from an early volume of the dictionary, appears to demonstrate a token of *and so forth:*

(20) *the accidentes of brede or wyne, that is to seie, the colour, the sauour, and so forth ... mowe not be, but in the substaunce of breed or wyne ...*
the accidents [outward characteristics] of bread or wine, that is to say, the color, the taste, and so forth ... may not be, but in the substance [essential nature] of bread or wine ...
(c1430(a1410 *MED* Love *Mirror* s. v. *accident* (n.) 2(a))

However, the same quotation from a much more recent entry in 21 reveals that *and so forth* is in fact not the complete form of the extender, but rather the larger variant, *and so forth of other,* which is also found elsewhere, as seen in 22.

(21) *the accidentes of brede or wyne, that is to seie, the colour, the sauour, and so forth of other, mowe not be but in the substaunce of breed or wyne after her kynde*
the accidents [outward characteristics] of bread or wine, that is to say, the color, the taste, and so forth of other, may not be, but in the substance [essential nature] of bread or wine, after their kind
(c1430(a1410 *MED* Love *Mirror* s. v. *substaunce* (n.) 2(d))

(22) *as is fier to ascende, erþe to falle doun, watir to kele, and so forþ of oþere.*

as is fire to ascend, earth to fall down, water to cool, and so forth of other.

(c1443 *MED* Pecock *Rule* s. v. *unliven* (v.))

Chaucer uses an even longer variant:

(23) *Certes wikkednesse shal be warisshed by goodnesse, discord by accord, werre by pees, and so forth of othere thynges*
Certainly wickedness shall be healed by goodness, discord by accord, war by peace, and so forth of other things
(c1490 *MED* Chaucer *CT.Mel.* s. v. *vanishen* (v.) 5(b))

Should Chaucer's *and so forth of other things* be considered a mere variant of the tag in 21? If not, is it better analyzed as being related to *and other things*? These questions relate to the nature of the lexicon in general, which can be understood as a collection of idiosyncrasies and exceptions, but also exhibits patterned behavior. Corpus linguistics is transforming the study of semantics and of the interrelatedness of grammar and the lexicon. Section 4 contextualizes this with a very brief introduction to phraseological studies.

4 Diachronic phraseology

It can be difficult to find the appropriate balance of focus between two tendencies in language: the idiomatic (a reliance upon prefabricated chunks of text, fixed phrases such as *and so forth*) and the creative (the freedom we have as speakers to construct ad hoc new phrases which fit our particular situation, such as 24).

(24) *ympes and herbes and oþer feele thinges That growed on þat gardyn.*
shoots and herbs and other excellent things that grew in that garden
(c1450(c1405) *MED Mum & S.(2)*, s. v. *impe* (n.) 1(d)).

Regrettably, some linguists veer to one extreme or the other with almost a religious fervor (as it were, the Stubbsians vs. the Chomskians). However, a phraseological approach to the data allows recognition of the variability of forms (such as *and other excellent things* and *and other things that grew in that garden*), while enabling a focus on the extent to which they share their behavior and characteristics with the shorter variants, *and other things* or even *and things*.

Much as extender tags have been studied under many names, which partly but not entirely overlap in the way they are defined, so the term PHRASEOLOGISM overlaps with terms such as (SEMI-)FIXED PHRASE, LEXICAL PHRASE, FORMULAIC PHRASE, HOLOPHRASE, LEXICAL CHUNK, COLLOCATION, MULTI-WORD UNIT, or PREFABRICATED UNIT, and subcategories such as IDIOMS, PHRASAL VERBS, and COMPOUNDS. I will follow Gries in the choice and definition of the term *phraseologism:* "the co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance" (Gries 2008: 6).

Phraseological studies are undergoing an upsurge at the moment.[11] As noted at the outset of this paper, it can be considered both an old field and a relatively recent one. The earliest known Biblical concordance listing the contexts in which words appear in the Bible was produced between 1250–1252 (Bromiley 1979: 757).[12] However, most histories of linguistic phraseology focus on the twentieth century, primarily the latter half of it. One early work on English collocation was Palmer 1933. Other commonly cited early researchers are Firth (e.g. 1951) and Sinclair (e.g. 1966).[13] For the origins of English diachronic phraseology, one might point to the application of Parry and Lord's oral formulaic theory to English literature (Magoun 1953). But works cited by linguists of English as foundational, even for synchronic studies, tend to be much more recent, such as Pawley and Syder 1983 and Peters 1983.[14]

[11] This is evident even within recent SELIM conferences, for example in the presentations by Rodríguez Redondo and Contreras Domingo on phraseologisms built around head and other body parts.

[12] This is the concordance of Hugo St. Caro, originally completed in 1230 but as a lexical index, listing references but no contexts. Contextual phrases were added between 1250–1252 by others (Bromiley 1979: 757).

[13] This brief survey is restricted to English linguistics, but phraseological studies are more well-developed in other traditions. Cowie points to the particularly influential Russian work on idioms and phrases, focusing on the late 1940s through the 1960s (1998: 4).

[14] "The emergence of phraseology as a field in its own right is even more recent [than discourse analysis]. It was only in the 80s and 90s that the field established itself firmly in theoretical and applied linguistics (Pawley & Syder 1983; Peters 1983; Sinclair 1991; Cowie 1998)." (Cosme 2005).

There have been no explicitly phraseological studies of extender tags, although as already noted, Aijmer (1985, 2002) acknowledges the variable nature of extenders, and Overstreet and Yule have explored the use of extender tags within larger formulaic structures, or "constructions" in the Construction Grammar sense, *(I mean) X and everything, but Y and not X or anything, but Y* (but not as semi-lexicalized phrases themselves) (2001, 2002). Even within diachronic studies more broadly, as recently as 2008 Markus can refer to "the deficiency in the research of historical English phraseology" (2008: 181).[15]

5 Extender tags as phraseologisms and collostructs

The list of extenders identified in Carroll 2007 was a very limited set of extenders, all of them very short. Even some of the more recent articles on PDE general extenders, for example, Cheshire 2007, focus on a similarly short list of similarly short forms. There are advantages to restricting one's study to short forms like these. It is very easy to search for such invariant strings in a corpus or indeed any computer-readable text, without needing to parse or tag it (although there is the slight danger of identifying forms which look identical but have other functions). Moreover, these short, fixed phrases tend to be the forms for which one gets the highest token counts; Cheshire points out that her own list of seven forms, six of which are invariant, accounts for just over 75% of all the forms in her data (2007: 165).

However, the last quarter of Cheshire's data likely included forms as varied (and yet similar to one another) as the Middle English phrases listed here:

[15] Measuring the true extent of previous literature is nearly impossible. Even acknowledging the wide range of variant terminology mentioned above, there is much research on individual phraseologisms or collostructs which gives little to no prominence to any synonyms for *phraseology*. To demonstrate with one example, in the same volume as my 2007 inventory is a paper on early English complex predicates, which does mention "idiomaticity", "collocations", and "constructions" in its abstract (Bello-Piñón and Méndez-Souto 2007), but which I had initially overlooked because my current area of interest is not verb phrases.

> *and alle thyng, and all other thyng, and thise othere thinges, and alle ðo þing, and all-kynne thyng, or ony such thing, and swiche othere thinges, and othere notefull thinges, and other thinges hereafter folowyng ...*[16]

Should this list be considered nine separate tags; or eight variants of the single tag *and things,* plus one *or things;* or perhaps three variants on *and/or all thing(s),* three of *and/or other thing(s),* one of *and/or such thing(s),*and one each of the mixed forms *and/or all other thing(s)* and *and/or such other thing(s)*? The problem is more difficult if we also include the even wider range of more specific tokens such as the one already cited as 24 (*and oþer feele thinges That growed on þat gardyn*). Medievalists are likely familiar with the wide dialectal and orthographic variability of Middle English, but tag variants are not unique to the medieval period. As Aijmer says of her PDE data, "The tags are not simply sequences of words which can be accounted for compositionally but they are lexicalized phrases which are partly fixed and partly flexible" (Aijmer 2002: 224).

By this definition, then, I will classify extender tags by their head noun (the "lexical item" of Gries's definition)[17] or, if there is no head noun, by the word immediately preceding a slot where one might have been expected (e.g. by *that* in an example cited above, *and all that* [*stuff*]). It is possible also, and desirable, to reference the entire class of extender tags, not by the term *phraseologism* (as there is no single lexeme which unites the class, not even *and*), but as a COLLOSTRUCT (Stefanowitsch and Gries 2003), a syntactic frame upon which individual extender tags can be built.

The distinction between general and specific tags is a perfect example of the need to find a balance of focus. To concentrate exclusively on general tags simplifies the data, it is true. It is also the case *ipso facto* that any general tag can be expected to occur more frequently than a specific tag. However, my strict adherence to Overstreet's definitions of general extenders in my study of the CEECS data led to an extremely short inventory of forms (2007: 48), which might be seen as underestimating the linguistic importance of other partially-lexicalized phrases, neither

[16] All of these are attested examples, but for the sake of space I have not given references for each. A search of the *Middle English Compendium* will reveal citations for most of these.

[17] Gries's definition of the *phraseologism* was cited above, in section 4.

completely frozen nor completely predictable.[18] I prefer to see the partially-creative phrases as flexible variants of certain of the shorter, even fully-fixed, phrases. The language user has the freedom to create variants falling anywhere on the cline between general and specific, as the occasion demands. At the far end of the cline, the most creative and specific extender tags, especially those built around non-general head nouns, may not be phraseologisms, yet they remain tokens of the extender tag collostruct.

6 Some behavior patterns of extender tags

Until now, my focus has been on the collostruct, that is, the broad category of extender tags. In this section, I change focus and offer some observations on the behavior of certain phraseologisms, certain categories of extender tag within the wider collostruct. What (if any) difference is there between one extender tag and another? Why does the language have such a variety of them? How does a language user select an extender tag for use? Previous synchronic research has shown that certain tags may be marked, whether socially, stylistically, or (as in Macaulay 1985) idiolectally. Jefferson even proposes a poetic motivation, suggesting that choice may be motivated by alliteration, assonance, and rhyme (1990: 69–73).

There seems to be a widespread assumption that all extender tags should be stigmatized as "vague" language, but extender tags are found across a wide range of genres and registers, although with different frequencies and with certain individual tags being very restricted. For example, Biber et al. found for PDE that *or something* is forty times more common in conversation than *or so*, and that *etc.* is more than forty times more frequent in academic texts than it is in fiction (1999: 116).[19]

Stubbs 2002, building on the work of Sinclair (1996/2004), categorizes the phraseological behaviors of LEXICAL ITEMS (words or phraseologisms) into four types. The first three are all ways of describing "the company a word keeps", to use Firth's famous phrase. COLLOCATION is the attraction

[18] Within the context of that publication my concerns about the distinction were made clear. It is only when that list or table is taken out of that context that it evokes concern.

[19] The normalized frequencies they cite are 400 tokens per million words for *or something* in conversation, 10/million for *or so* in conversation, 200/million for *etc.* in academic texts, and less than 5/million for *etc.* in fiction.

of a word or phraseologism to other words or phrases. SEMANTIC PREFERENCE is broader than collocation, referring to an attraction not just to a particular word but to the larger semantic field. COLLIGATION (a name which dates back to Firth, but unfortunately is easily confused with or misheard as *collocation*) is the attraction to a particular grammatical category or larger construction or grammatical frame.

The fourth behavior type is unlike the first three in that it is not a matter of the context in which a word or phrase is found, although that context may be a signal of it. DISCOURSE PROSODY (which Sinclair called *semantic prosody*) refers to the connotations carried by the phraseologism (or collostruct). The distinction between *semantic preference* and *discourse prosody* may be more blurred than clear (Stubbs 2002: 66, 106), but *semantic preference* is more strictly semantic (about denotations), while *discourse prosody* addresses pragmatics or connotations. Typical functions of *discourse prosody* are to mark speaker attitude or to mark discourse function (Stubbs 2002: 88).

While *collocation* or *colligation* can be described either for individual words or for larger units, *discourse prosody* applies only to the larger units (in part because for single words the term *connotation* is already available and preferred). *Discourse prosody* refers specifically to a connotation which cannot be traced to any single lexeme in the construction, but only to their combination. Sinclair used the example of the phrasal verb *set in,* which is used of negative things (such as *rot* or *bitterness*) more often than of neutral things (*trend*), and rarely if ever of good things (he suggests that the sentence *Good times set in* would sound marked or even humorous) (Sinclair 1987: 155–6). Neither the verb *set* nor the preposition *in* has negative connotations on its own. This spreading of a semantic feature beyond word boundaries is thought of as analogous to the way in which phonological features can spread beyond word boundaries in the more common linguistic use of the term *prosody.*

Here follow some new observations on the behavior of historical English extender tags, described in terms of collocation, colligation, semantic preference, and discourse prosody. Although brief, each section will illustrate the nature of the insights such research can yield. The primary source of material for this section was the CEECS (Corpus of Early English Correspondence Sampler; see Nurmi 1998), made up of 23

collections of letters, some personal, some business letters, from the years 1418–1680. It includes 194 informants and comes to 450,000 words.

A. Collocation: the company a word (or phrase) keeps

The fixed phrase or word pairing *and diverse* occurs 25 times in the corpus.[20] In 13 of these 25 tokens, *and diverse* is followed immediately by *other(s)*. Thus, 52% of *and diverse* tokens collocate with *other(s)*.[21] Using the notation of Stubbs (2002), this can be represented as follows:

and diverse 22 <*other(s)* 52%>

Collocational patterns can also be shown by means of a positional frequency table. The asterisked column represents the phrase *and diverse*. Each row begins with a collocate of the phrase, followed by the number of times that collocate is found within four words of the phrase. The remaining columns show the positions relative to the target phrase in which the collocates were found (L4 is the position four words to the left of *and diverse,* while R1 is the position immediately following *and diverse*).

Table 2: Positional frequencies for and diverse *(CEECS)*

	#	L4	L3	L2	L1	*	R1	R2	R3	R4
other(s)	13	0	0	0	0	*	13	0	0	0
of	9	0	2	2	0	*	3	0	0	2
and	8	2	1	0	0	*	0	3	0	2
the	8	1	2	2	0	*	0	0	1	2
his	5	1	1	2	0	*	0	1	0	0

This collocational pattern means that as soon as the word *diverse* was seen to follow *and,* the Early Modern English reader was primed to expect it was likely to be an extender tag which would continue with the lexeme

[20] Numbers in this section include spelling variants. I made use of the software WordSmith Tools 5.0 in compiling lists of collocates and positional frequency tables.

[21] The collocation span I have used is 4 words to the left or right of the target phrase.

other. The Helsinki Corpus shows an extremely similar result for the same time period.[22]

and diverse 23 <*other(s)* 52%>

Table 3: Positional frequencies for and diverse *(Helsinki Corpus)*

	#	L4	L3	L2	L1	*	R1	R2	R3	R4
other(s)	16	0	0	1	0	*	14	0	0	1
the	11	6	1	0	0	*	0	1	1	2
and	9	1	3	1	0	*	0	2	1	1
of	8	0	1	4	0	*	1	2	0	0
in	5	0	2	2	0	*	0	0	1	0

The similarity between the two tables indicates that this was a characteristic of the language more widely, and not merely of letters.

B. Colligation: the syntactic environment

Colligation refers to the grammatical environment in which a lexical item is found.[23] Consider again this word *other,* which was so prevalent with *and diverse*. It occurs in extender tags of varying length, with or without indefinite quantifiers and adjectival modification, as shown below:

(25) *These <u>and many othir thingis</u>*
These and many other things
((a1464) *MED* Capgr. *Chron.* s. v. *halwe-messe* (n.))

(26) *garlek, leke, oynouns, pepir <u>and such obere scherpe þingis</u>*
garlic, leek, onions, pepper and such other sharp things
((c1443) *MED* Pecock *Rule* s. v. *kēne* (adj.) 4(a))

(27) *eggs, alom, gummes, <u>and obere horrible & unholsome thinges</u>*
eggs, alum, gums, and other horrible and unwholesome things
((1457) *MED Let.Bk.Lond.K* s. v. *tarāğe* (n.))

[22] As noted above, CEECS covers the years 1418–1680 in 450,000 words. Sections ME IV, EModE I, and EModE II of the Helsinki Corpus cover a similar time span, 1420–1640, with just under 600,000 words.

[23] It is also possible to study TEXTUAL COLLIGATION, the textual environment in which a lexical item is found. The observations that about a quarter of NP-lists in CEECS end with an extender, and that extenders are less common in short lists than in long ones (Carroll 2007), fall under the heading of textual colligation.

(28) *the Manoirs, Londes, Tenementes, Fermes ... and other thinges*
the manors, lands, holdings, leases ... and other things
((1450) *MED RParl.* s. v. *occupīŏur* (n.))

Although, for clarity's sake, all the above examples were variants on ***and other things, other*** is also found in a similar range of tags with varying head nouns (as well as without any head noun at all):

(29) *hostiary, reder, benette, accolette, and oþer*
ostiary, reader, benet, acolyte, and other
(?a1475(?a1425), *MED, Higd.(2)* s. v. *rēder(e* (n.(1)))

(30) *catel or oþer goodys*
cattle or other goods
(*MED,* (1440), *PParv* s. v. *avēr* (n.(2)) 2(b))

(31) *wilde boores, wolues and oþer byteng beestis*
wild boars, wolves, and other biting beasts
((c1410), *MED,* York *MGame* s. v. *bēst(e* (n.) 3(c))

Although there is much variation, it is possible to categorize the grammatical structure of these extender tags as follows:[24]

CONJ (+ INDEF QUANTIFIER) (+ ADJS) (+ *other*) (+ ADJS) (+ NOUN)

What is both stimulating and frustrating is that almost every element is optional. Notice that even *other* is optional (although the frequency with which it is found makes it a helpful aid in searching for extender tags):

(32) *Mete and drink and mani þing elles*
food and drink and many things else
(c1330(?a1300), *MED, Arth.&M.* s. v. *pal* (n.) 2(a))

(33) *slaughtre of children and swich maner thyng*
slaughter of children and such manner [of] thing
((c1390), *MED,* Chaucer *CT.Pars.* s. v. *Hōli Chirche* (n.) (b))

The only obligatory element would seem to be the conjunction, and even there the speaker has a choice of which conjunction to choose. This is why the broad category of extender tags is considered not a phraseologism

[24] Even this description of the colligation remains incomplete, because it does not take account of possible post-modification of the head noun in specific extenders, seen in examples such as 44, below.

(by Gries's definition), but a collostruct (terms introduced in sections 4 and 5, above).[25]

C. Semantic preference: the company a word (or phrase) keeps

The phrase *and all* shows a clear preference in CEECS for occurring with second-person formal/plural pronouns and personal possessive pronouns of any person: *and all* 247 <*your, my, yours, his, our, her, their, thy* 62%> <*your, you, yours, ye* 70%>.[26] These would ordinarily be categorized as colligational patterns, but instead I wish to highlight the frequency of one particular string, *you and all your(s),* which is found 31 times in the corpus. Both this string in particular, and many of the other tokens of the colligation of *and all* with the pronouns listed above, can be understood as related to certain semantic preferences of *and all.*

The semantic fields of family and household members collocate with the phrase *and all* (that is, they are part of its semantic preference): <*children, sister(s), brother(s), brethren, kinsmen, wife, daughters, son* 11%>, <*household, servants, lord(ship), lady(ship)('s)* 14%>.[27] In each list, the most frequent collocates are listed first. The word *son*, for example, listed last in the family members category, is only found collocating with *and all* in the CEECS corpus one time, but when taken together with the tokens of its semantic associates such as *children*, that single token is seen as part of a larger pattern. The combined semantic field of relations and household members, along with friends, has members collocating with almost a quarter of the total tokens of *and all* (24.7%).

[25] More accurately, the class of extender tags consists probably of at least two collostructs with overlapping functions, since *and so forth* and its variants follow a somewhat different colligational structure.

[26] The phrase *and all* is found 249 times; 153 of those 249 tokens (62%) collocate with a possessive personal pronoun; 173 (70%) with any second-person formal pronoun. These numbers reflect normalized spellings and modernized word-boundaries (*almighty* is sometimes spelled as two words in the corpus, but such "tokens" of *all* have not been included here).

[27] Tokens of the word *lord* referring to people have been manually distinguished from those referring to God.

How does this relate to the pronoun colligations discussed above? It is true but perhaps not terribly interesting that family members are often the objects of the possessive pronoun *your* in the phrase *you and your.* Less obvious is to notice that *and all* also collocates with verbs of keeping <*have, preserve, keep* 11%>, and that the string *preserve you and all* is found 9 times (and the longer string *preserve you and all yours* 8 times). The subject of the keeping verb is often God; names for the Deity <*God, Lord, Jesus, Almighty, heaven*>[28] collocate with 12% of the tokens of *and all.*

These semantic preferences (along with the pronominal colligations) reflect the common usage in letters of benedictions such as the following:

(34) *and God in hevn preserve you and all yours*
(1620 CEECS MBOURCHIER 68)

Other versions of this formula allow insertions within the string, or other variations on it:

(35) *And almigthty God preserve you, my lady, and all your housholde*
(1479 CEECS HUNTON II, 92)

The collocation *and all* is also found independently of such benedictions, but in letters of this period, the longer phraseologism *you and all yours* is a common one, often intercollocating with the phraseologism *God preserve you.*

D. Discourse prosody: the connotations of a phrase

Overstreet (1999) has observed that one of the functions of an extender is to mark shared knowledge. In CEECS extender tags containing *and such* also have the function of marking shared values, especially shared values of disapproval:

(36) *treacherous weapons, as knifes and such like*
(1639 CEECS BHARLEY 51)

(37) *certen horstealers, cutpurses and such lyke*
(1585 CEECS WIFLEETWOOD 297)

This is particularly true of the more specific extender tags:

[28] *Heaven* is included here because if the phrase *God in heaven* is used, then the word *God* may fall outside the 4×4 search range.

(38) *all these Treators and such other wicked people*
(1586 CEECS WIFLEETWOOD 308)

(39) *common receivers of seminaries and such like bad persons*
(1597 CEECS WCECIL 115)

With only 16 tokens of *and/or such-like* in the corpus, this remains a tentative finding, but one worth investigating further in the future. The hypothesis is that although neither *such* nor *and* had negative connotations individually, the collocation of the two in extender tags did carry negative connotations, and was therefore the phraseologism selected for use with other negative words or concepts, such as *treacherous, stealers,* and *treators.* Examples 38 and 39 are of particular interest since they represent in-group expressions of disapproval (while theft is viewed negatively by most people, political and religious values are more varied).

7 Other observations

Collocation, colligation, semantic preference and discourse prosody do not exhaust the observations that can be made about extender tags. Other observations have applications beyond obvious phraseological concerns. Space does not permit more than a few illustrative examples, but three will be given here. For lexical semantics, the study of extender tags may facilitate the identification of hyponym-superordinate clusters; it is also useful for lexicography, in the identification of vague nouns; and for discourse analysis, extender tags aid in the identification of quotable extracts (from a canon or from an inventory of recognized formulae).

That superordinates and their hyponyms collocate is unsurprising. However, it is useful for lexical semantics to recognize that this frequently happens within the context of specific extenders. A common pattern for specific extenders is one in which the noun(s) preceding the extender are hyponyms of the extender's head noun. This is illustrated by four examples below:

(40) *Percely, clarey, and eke sage And all other herbage*
Parsley, clary, and also sage And all other herbs
(a1500(?a1450) *MED Treat.Garden* s. v. *clarē* (n.(2)))

(41) *with jakkes, salettes, bowes, arrowes, glaybes, gissarnes, longdebibes, and other armour defensive.*

with quilted-tunics, helmets, bows, arrows, lances, halberds, pole-arms and other defensive armor
((1450) *MED Complaint* s. v. *gisarme* (n.))

(42) *þe sunne and other sterres*
the sun and other stars
(14-- *MEMT* Rupescissa, *Quintessence* 108)[29]

(43) *fyges and reysyns and othere swete mettes and drynkes*
figs and raisins and other sweet foods and drinks
(1477 *MEMT Diet and Bloodletting* 251)

Secondly, extender tags should be noted by lexicographers, in order not to over-specify the denotations of lexemes in their definitions. The phraseologism *and stuff* is used in Middle English to mark a variety of different cognitive categories. The *Middle English Dictionary* groups each of the two cited below under a different sense, implying that the word *stuffe* was ambiguous between "military supplies" (44), "embroidery materials" (45), and other meanings as varied as "movable goods belonging to a chapel," "quilted material worn under chain mail" and "the meat contained in a crab shell" (s. v. *stuffe*), rather than defining it as a vague or general noun.[30]

(44) *men of Armes & Archers, And moche other stuffe þat longeth to werre*
men of arms & archers, and much other stuff that pertains to war
(a1500 *MED Brut-1419* s. v. *stuffe* (n.) 1(a))

(45) *gold silke and other stuff*
gold silk and other stuff
((1448) *MED Pet.Hen.VI* s. v. *stuffe* (n.) 3(b))

Finally, extender tags serve to mark texts and formulae which are so familiar to the reader that they need not be quoted in full. This discourse-organizational function was already mentioned in section 3 (see also Carroll 2007 for supplementary examples). Quotations and formulaic

[29] Since *MEMT* offers no date for the text, this dating is taken from the University of Glasgow Manuscripts Catalogue.

[30] Similar lexicographic mistakes have been noted by others; see, for example, Channell 1994: 37. The *MED* does use the word *general* in defining two of the eighteen sub-senses of the word.

phrases which are assumed to be shared knowledge may be abbreviated with an extender tag. This is a function which has been largely ignored by extender tag studies of PDE. A corpus search for such tags will answer questions for literary scholars, theologians, and classics scholars: Which texts were assumed as shared knowledge between an author and his readers? How much of a text or formula had to be quoted before an author could be confident that the audience would recognize it?

(46) *saienge to ham: 'Dredeth nouȝt! ȝe secheth Jesu!' and so forthe, as the gospell telleth.*
saying to them, "Fear not! You seek Jesus!" and so forth, as the Gospel tells.
(c1430(a1410) *MED* Love *Mirror* s. v. *and* (conj.) 1c (b))

(47) *A charme for to stawnchyn blood ... 'In nomine patris et cetera ... I conjure the, blood ...*
A charm for stanching blood ... *In nomine patris* et cetera ... I conjure thee, blood ...
(?c1450 *MED Stockh.PRecipes* s. v. *charme* (n.) (a))

8 Conclusion

This paper has introduced a specific dataset, the extender tags of early English, and a linguistic subfield, diachronic phraseology. Either can be studied independently, but both deserve wider attention, and both can profitably be studied together. For example, the acknowledgement that most phraseologisms are not frozen forms but allow for variation motivates a decision to concentrate primarily on general extenders without ignoring the existence (and importance) of specific extenders.

Extender tags have been shown to be multifunctional, with their functions including the covering of lexical gaps, the abbreviation of long lists, and the highlighting of shared knowledge (whether that be shared familiarity with a text or a shared evaluation of denoted items). Stubbs's four behavior types, collocation, colligation, semantic preference and discourse prosody, were shown to be worth studying in regard to extender tags, although it was observed that the colligational patterns are particularly flexible for this class of items.

The strictures of corpus semantics are in some ways ideal for diachronic linguistics. The reliance on attested data is shared by both fields. On the other hand, spelling variation can create difficulties for corpus searches

of early English material, but I hope I have shown that the challenge is worthwhile. With a framework for study thus laid out, and previous bibliography summarized, I look forward to reading (and myself pursuing) more detailed diachronic phraseological studies in the future.

Ruth Carroll
University of Turku

References

Aijmer, K. 1985: What Happens at the End of our Utterances? In O. Togeby ed. *Papers from the Eighth Scandinavian Conference of Linguistics.* Vol 2. Copenhagen, Københavns Universitet: 366–389.

Aijmer, K. 2002: Particles with Vague Reference: 'And That Sort Of Thing.' Chapter 6 of *English Discourse Particles: Evidence from a Corpus.* Amsterdam, John Benjamins: 211–249.

Ball, C. N. & M. Ariel 1978: 'Or Something', Etc. *Penn Review of Linguistics* 3: 35–45.

Bello-Piñón, N. & D. E. Méndez-Souto 2007: Complex Predicates in Early Scientific Writing. In I. Moskowich-Spiegel Fandiño & B. Crespo García eds. *Bells Chiming from the Past: Cultural and Linguistic Studies on Early English.* Amsterdam, Rodopi: 169–178.

Biber, D., S. Johansson, G. Leech, S. Conrad, & E. Finegan 1999: *Longman Grammar of Spoken and Written English.* Harlow, Pearson Education Limited.

Bromiley, G. W. 1979: *The International Standard Bible Encyclopedia,* 2nd edn. Grand Rapids, MI, Eerdmans Publishing.

Carroll, R. 2007: Lists in Letters: Elided NP-Lists and General Extenders in Early English Correspondence. In I. Moskowich-Spiegel Fandiño & B. Crespo García eds. *Bells Chiming from the Past: Cultural and Linguistic Studies on Early English.* Amsterdam, Rodopi: 37–53.

CEECS = 1999. Corpus of Early English Correspondence Sampler. In K. Hofland, A. Lindebjerg & J. Thunestvedt eds. *ICAME Collection of English*

Language Corpora. University of Bergen: The HIT Centre. http://icame.uib.no/ceecs/index.htm.

Cosme, C. 2005/03: Why Study Phraseology and Discourse? *Foreign Language Learning: Phraseology and Discourse.* Centre for English Corpus Linguistics, Université Catholique de Louvain. http://cecl.fltr.ucl.ac.be/PhraseologyARC/welcome.html.

Channell, J. 1994: *Vague Language.* Oxford, Oxford University Press.

Cheshire, J. 2007: Discourse Variation, Grammaticalisation and Stuff Like That. *Journal of Sociolinguistics* 11: 155–193.

Cortés Rodríguez, L. 2006: Los elementos de final de serie enumerativa del tipo *y todo eso, o cosas así, y tal,* etc. *Boletín de Lingüística* 18: 102–129.

Cowie, A. P. 1998: *Phraseology: Theory, Analysis, and Applications.* Oxford, Clarendon Press.

Crystal, D. & D. Davy 1975: *Advanced Conversational English.* London, Longman.

Cucchi, C. 2007: An Investigation of General Extenders in a Corpus of EU Parliamentary Debates. In M. Davies, P. Rayson, S. Hunston & P. Danielsson eds. *Proceedings of the Corpus Linguistics Conference CL2007.* http://ucrel.lancs.ac.uk/publications/CL2007/paper/242_Paper.pdf.

Dines, E. R. 1980: Variation in Discourse – "And Stuff Like That". *Language in Society* 9: 13–31.

Drave, N. 2002: Vaguely Speaking: A Corpus Approach to Vague Language in Intercultural Conversations. In P. Collins, P. Peters & A. Smith eds. *New Frontiers of Corpus Research.* Amsterdam, Rodopi: 25–39.

Dubois, S. 1992: Extension Particles, Etc. *Language, Variation and Change* 4: 179–203.

EEBO = Early English Books Online. http://eebo.chadwyck.com.

Evison, J., M. McCarthy & A. O'Keefe 2007: *Looking out for love and all the rest of it:* Vague Category Markers as Shared Social Space. In J. Cutting ed. *Vague Language Explored.* Basingstoke, Palgrave: 138–157.

Firth, J. R. 1951: Modes of Meaning. *Essays and Studies of the English Association,* n.s. 4: 123–149.

Graman, P. 1998: *General Extenders in English and Dutch: A Comparison of* The Catcher in the Rye *and Three Translations.* Thesis, Nijmegen, Department of English, Katholieke Universiteit.

Gries, S. Th. 2008: Phraseology and Linguistic Theory: A Brief Survey. In S. Granger & F. Meunier eds. *Phraseology: An interdisciplinary perspective.* Amsterdam, John Benjamins: 3–25.

Hamelius, P. ed. 1919: *Mandeville's Travels.* Vol I. London, Oxford University Press.

HC = 1991/1999. Helsinki Corpus of English Texts: Diachronic Part. In K. Hofland, A. Lindebjerg & J. Thunestvedt eds. *ICAME Collection of English Language Corpora.* University of Bergen: The HIT Centre. http://icame.uib.no/ceecs/index.htm.

Jefferson, G. 1990: List-Construction as a Task and Resource. In G. Psathas ed. *Interaction Competence.* Washington, DC, The International Institute for Ethnomethodology and Conversation Analysis: 63–92.

Jucker, A., S. W. Smith, & T. Lüdge 2003: Interactive Aspects of Vagueness in Conversation. *Journal of Pragmatics* 35: 1737–1769.

Knappe, G. 2004: *Idioms and Fixed Expressions in English Language Study before 1800. A Contribution to English Historical Phraseology.* Frankfurt, Peter Lang.

Kytö, M. 1996. *Manual to the Diachronic Part of the Helsinki Corpus of English Texts.* Helsinki, Department of English, University of Helsinki. http://khnt.hit.uib.no/icame/manuals/HC/.

Lauwereyns, S. 2002: *Toka* 'or Something' in Japanese Discourse. In Y. Shirai, H. Kobayashi, S. Miyata, K. Nakamura, T. Ogura & H. Sirai eds. *Studies in Language Sciences (2): Papers from the Second Annual Conference of the Japanese Society for Language Sciences.* Tokyo, Kurosio.

Macaulay, R. 1985: The Narrative Skills of a Scottish Coal Miner. In M. Görlach ed. *Varieties of English Around the World: Focus On Scotland.* Amsterdam, John Benjamins: 101–124.

Maclagan, M., B. Davis & R. Lunsford 2008: Fixed Expressions, Extenders and Metonymy in the Speech of People with Alzheimer's Disease. In S. Granger & F. Meunier eds. *Phraseology: An interdisciplinary perspective.* Amsterdam, John Benjamins: 175–187.

Magoun, F. P. 1953: The Oral-Formulaic Character of Anglo-Saxon Narrative Poetry. *Speculum* 28: 446–467.

Markus, M. 2008: Germanic vs. French Fixed Expressions in Middle English Prose: Towards a Corpus-based Historical English Phraseology. In R. Dury, M. Gotti & M. Dossena eds. *Lexical and Semantic Change*. Amsterdam, John Benjamins: 159–184.

MEC = McSparran, F. ed. 2006: *Middle English Compendium: Corpus of Middle English Prose and Verse*. http://quod.lib.umich.edu/c/cme/.

MED = Kurath, H., S. M. Kuhn, J. Reidy, & R. E. Lewis eds. 1952–2001: *Middle English Dictionary*. Ann Arbor: University of Michigan Press. http://quod.lib.umich.edu/m/med/.

MEMEM = Bailey, R. W., J. L. Robinson, J. W. Downer & P. V. Lehman 1975/1994. *Michigan Early Modern English Materials*. Ann Arbor: Xerox University Microfilms with University of Michigan Press. http://quod.lib.umich.edu/m/mememe/.

MEMT = Taavitsainen, I., P. Pahta & M. Mäkinen 2005: *Middle English Medical Texts*. Amsterdam: John Benjamins.

Meyerhoff, M. 1992: 'A Sort of Something:' Hedging Strategies on Nouns. *Working Papers on Language, Gender & Sexism* 2: 59–73.

Mittmann, B. 2004: *Mehrwort-Cluster in der englischen Alltagskonversation*. Tübingen: Gunter Narr Verlag.

Norrby, C. & J. Winter. 2001: Affiliation in Adolescents' Use of Discourse Extenders. In Allen, C. ed. *Proceedings of the 2001 Conference of the Australian Linguistic Society*: http://au.geocities.com/austlingsoc/proceedings/als2001/winter_norrby.pdf.

Norrby, C. 2002: Svenska påhängsuttryck av typen *å så* och *eller nåt*. En diskussion av deras förekomst och funktion(er) i ett samtida ungdomsmaterial. *Språk & Stil* 11: 183–210.

Nurmi, A. 1998: *Manual for the Corpus of Early English Correspondence Sampler*. Helsinki, Department of English, University of Helsinki. http://icame.uib.no/ceecs/

O'Keefe, A. 2003: 'Like the wise virgins and all that jazz' – Using a Corpus to Examine Vague Language and Shared Knowledge. In U. Connor and

T.A. Upton eds. *Applied Corpus Linguistics: A Multidimensional Perspective.* Amsterdam, Rodopi: 1–20.

Ortega Barrera, I. 2008: A Diachronic Account on Extenders in English Recipes. Conference paper, *International Conference on English Historical Linguistics,* Munich.

Overstreet, M. 1999: *Whales, Candlelight, and Stuff Like That: General Extenders in English Discourse.* New York, Oxford University Press.

Overstreet, M. 2005: And stuff *und so:* Investigating Pragmatic Expressions in English and German. *Journal of Pragmatics* 37.11: 1845–1864.

Overstreet, M., J. Tran & S. Zietze 2008: Increasing Pragmatic Awareness: Die Vagheit der Sprache *und so. Die Unterrichtspraxis / Teaching German* 39.1–2: 24–29.

Overstreet, M. & G. Yule 1997a: Locally Contingent Categorization in Discourse. *Discourse Processes* 23: 83–97.

Overstreet, M. & G Yule 1997b: On Being Inexplicit and Stuff in Contemporary American English. *Journal of English Linguistics* 25: 250–258.

Overstreet, M. & G Yule 1999: Fostering Pragmatic Awareness. *Applied Language Learning* 10: 1–13.

Overstreet, M. & G Yule 2001: Formulaic Disclaimers. *Journal of Pragmatics* 33: 45–60.

Overstreet, M. & G. Yule 2002: The Metapragmatics of *And Everything. Journal of Pragmatics* 34: 785–794.

Pawley, A. & S. F. Snyder 1983: Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency. In J. C. Richards & R. W. Schmidt eds., *Language and Communication.* New York, Longman: 191–226.

Peters, A. M. 1983: *The Units of Language Acquisition.* Cambridge, Cambridge University Press.

Sánchez Ayala, I. 2003: Constructions as Resources for Interaction: Lists in English and Spanish Conversation. *Discourse Studies* 5: 323–349.

Sinclair, J. 1966: Beginning the Study of Lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins eds. *In Memory of J.R. Firth.* London, Longman: 410–30.

Sinclair, J. M. ed. 1987: *Looking Up: An Account of the COBUILD Project in Lexical Computing.* London, Collins.

Sinclair, J. 1996/2004: The Search for Units of Meaning. In *Trust the Text.* London, Routledge: 24–48. [Originally published in *Textus IX* in 1996.]

Stefanowitsch, A., & S. Th. Gries 2003: Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics* 8.2: 209–243.

Stenström, A.-B., G. Andersen, & I. K. Hasund 2002: *Trends in Teenage Talk: Corpus Compilation, Analysis, and Findings.* Amsterdam, John Benjamins.

Stubbe, M. & J. Holmes 1995: *You Know, eh* and Other 'Exasperating Expressions': An Analysis of Social and Stylistic Variation in the Use of Pragmatic Devices in a Sample of New Zealand English. *Language & Communication* 15.1: 63–88.

Stubbs, M. 2002: *Words and Phrases: Corpus Studies of Lexical Semantics.* Oxford, Blackwell.

Stubbs, M., M. Hoey, M. Mahlberg, & W. Teubert 2007: *Text, Discourse and Corpora.* London, Continuum.

Terraschke, A. 2007: Use of General Extenders by German Non-Native Speakers of English. *International Review of Applied Linguistics in Language Teaching* 45.2: 141–160.

Ward, G. & B. Birner 1993: The Semantics and Pragmatics of *And Everything. Journal of Pragmatics* 19: 205–214.

Warner, G. F. ed. 1889: *The Buke of John Maundeuill.* Westminster, The Roxburghe Club.

Winter, J. & Norrby C. 2000: Set Marking Tags – 'and Stuff'. In J. Henderson ed. *ALS Conference Proceedings*: http://www.arts.uwa.edu.au/LingWWW/als99/winter&norrby.pdf.

Youssef, V. 1993: Marking Solidarity across the Trinidad Speech Community: The Use of *an Ting* in Medical Counselling to Break Down Power Differentials. *Discourse & Society* 4: 291–306.

Received 15 Apr 2009; *revision received* 02 Jun 2009; *accepted* 14 Jun 2009