1 Introduction

1.1 The Jespersen Cycle and Theories of Syntactic Change

1.1.1 The Jespersen Cycle

The Jespersen Cycle (Jespersen, 1917) has long been regarded as a textbook case of grammaticalisation, found in many languages. The English Jespersen Cycle manifests itself as a sequence of overlapping stages, each formally distinct from the preceding one as in (1)–(3).

(1) Stage One: Sentential negation is marked by *ne* alone (c.1150–1300)
   a. we **ne** moten halden Moses e lichamlice
      we NEG might observe Moses' law bodily
      'we might not observe Moses law literally'
      (CMLAMBX1,89.735)
   b. we **ne** mugen þat don
      we NEG can that do
      'We cannot do that'
      (CMTRINIT,103.1370)

(2) Stage Two: The sentential negator *not* co-occurs with *ne*. Sentential negation comprises two parts (c. 1150–1400)
   a. ac of hem **ne** speke ic **noht**
      but of them NEG spoke I not
      'but I did not speak of them'
      (CMTRINIT,95.1272)
   b. I **ne** may **nat** denye it
      I NEG may not deny it
      'I may not deny it'
      (CMBOETH,435.C1.262)

(3) Stage Three: Sentential negation is marked by *not* alone. (c.1350–1500)
   a. Thou shalt **not** do so
      You ought not do so
      'You ought not do so'
      (CMROLLTR,41.855)

b. I know **nat** the cause
I know not the cause
'I do not know the cause'
(CMMALORY,627.3549)

The changes result in the grammaticalisation of *not* as a marker of sentential scope negation, and take place during a period from the ninth century to the fifteenth century. What changes constitute the Jespersen Cycle?

Much has been written about the process by which new negative markers are grammaticalised – there are many syntactic, semantic and pragmatic accounts of the processes involved in many languages, each of which decompose the change into different stages, and hence conceptualise the mechanisms and diachrony of the Jespersen Cycle in different ways. Existing theoretical accounts differ fundamentally in their conceptions of the cycle – how many stages are involved, what the syntactic elements and configurations are at each stage, what formal linguistic mechanisms are invoked to explain the change, whether the cycle involves morphological, syntactic or functional change at each stage and how these various factors might interact.

Given that formal and functional accounts structure the cycle in different ways and invoke different mechanisms of change, what evidence do we have to decide between them? Each account makes different predictions about how the three stages *ne, ne...not* and *not* relate to each other, what changes are involved in the cycle and how those changes progress over time. An account of the Jespersen Cycle must not only formalise each diachronic stage, but also allow us to model the changes in a way that predicts the distribution of *ne, ne...not* and *not* observed in diachronic data as the changes progress. Formal and functional linguistic analyses should also structure the Jespersen Cycle in ways that are consistent with what we know about how morphosyntactic changes progress over time.

In this book, I argue that – viewed in this way – quantitative models based on data from diachronic corpora provide crucial empirical evidence to inform formal accounts of the Jespersen Cycle, and to establish what mechanisms of change are involved. By placing strict empirical constraints on which changes are plausible and which changes are not, quantitative models of change provide a solid foundation upon which to build formal and functional analyses of the Jespersen Cycle, and to ascertain how different formal and functional changes relate to each other within the cycle.

This empirical approach provides new evidence that *ne* undergoes morphosyntactic weakening prior to its loss. This idea is first proposed by Jespersen (1917), but has proven problematic to formalise, and is often not captured in syntactic analyses of the Jespersen Cycle such as Haegeman (1995), Roberts and Roussou (2003) and Zeijlstra (2004). Corpus evidence enables us to formalise the steps in this weakening process very precisely and to identify their empirical effects. The morphosyntactic weakening of *ne* is

not only crucial to the concomitant grammaticalisation of *not*, its effects go far beyond the Jespersen Cycle. When properly formalised, the weakening of *ne* provides a coherent explanation of several interlinked changes to negation during the Old English (c.800–1100) and Middle English (c.1100–1500) periods. These include changing patterns of redundant negation and negative inversion, changes in the form and availability of negative concord and the relationship between the loss of negative concord and the Jespersen Cycle.

1.1.2 Morphosyntactic Change and the Jespersen Cycle

How can we analyse the Jespersen Cycle in a way that is consistent with formal approaches to morphosyntactic variation and change? The notion of syntactic parameter is central to generative models of syntactic variability, in order to explain relationships between languages, varieties or diachronic stages within the same language variety; to constrain possible axes of variation across languages, varieties or diachronic stages; and to establish pathways of morphosyntactic change.

Typically, in parameter-based approaches to morphosyntactic change, change proceeds through competition between two formally distinct but functionally equivalent syntactic options. These options constitute different ways to realise or set a particular parameter. For example Pintzuk (1999) argues that Old English VPs may either be head-initial or head-final. In sixteenth century English, Kroch (1989) argues that the tense head may be lexicalised by V to T movement or by insertion of DO in T^0. Variation or change in a single parameter may explain the presence or absence of several surface syntactic phenomena that are the reflexes of that parameter. For example Bobaljik and Thrainsson (1998) argue that whether or not a language projects Agreement (Agr) projections has several consequences including the availability of distinct markers for tense and agreement on finite verbs, object shift and transitive expletive constructions. Kroch (1989) proposes that all reflexes of a particular parametric shift should change at a constant rate over time, providing an empirical basis to identify which changes are manifestations or reflexes of a single parameter.

However, an approach to grammatical competition in terms of binary parameters is problematic for the Jespersen Cycle because, at least on the surface, it appears too restrictive. The Jespersen Cycle does not involve straightforward substitution of one form (*ne*) for another (*not*). Stage two of the Jespersen Cycle is particularly problematic. The Middle English Jespersen Cycle comprises at least three stages (1)–(3) given above, with apparent redundancy in negative marking in the second stage (the co-occurrence of *ne* and *not*). Deriving three stages requires a sequence of at least two parametric changes, raising the question of what these two parametric changes are and how they relate to each other. Notions of functional equivalence and mutual exclusivity pose difficulties too. Clearly,

ne and *not* are not mutually exclusive at stage two, therefore they are not functionally equivalent. Many accounts of the Jespersen Cycle such as Frisch (1997), Roberts and Roussou (2003) and Zeijlstra (2004) regard the grammaticalisation of *not* and the loss of *ne* as independent but intersecting changes, each involving a different parameter. Under this approach, the co-occurrence of *ne* and *not* at stage two is epiphenomenal. However, we will see that *ne...not* is distributionally independent of the preceding and following stages. Modelling the distribution of *ne...not* therefore requires a fundamentally different account of the parameters involved.

The status and place of parameters has been reappraised in recent formal (Minimalist) syntactic frameworks. Recent Minimalist approaches such as Chomsky (1999, 2001), and accounts within this paradigm that seek to account for variation, such as Roberts and Roussou (2003), Adger (2006) and Adger and Trousdale (2007), locate parametric variation in the lexicon, operationalising Borer's (1984) insight. By making the morphosyntactic features associated with lexical items the triggers for certain syntactic operations like agreement or movement, these accounts place very tight constraints on what constitutes a parameter and what values or settings a parameter may have. This constrains our approach to morphosyntactic competition. Competition is not between entire grammars or grammatical subsystems, but a choice between lexical items with mutually exclusive feature specifications all present in the lexicon of an individual speaker. I demonstrate that the Jespersen Cycle can be analysed as a sequence of parametric changes in these terms, and furthermore that such an account receives empirical support from patterns of variation and change in diachronic corpora. It also has the conventional advantage associated with parametric accounts of variation – the ability to explain several changes in the grammar of negation as reflexes of a single parametric shift.

1.2 The Jespersen Cycle and Other Aspects of the Grammar of Negation

Several negation phenomena in early English undergo change. I argue that these changes are not only linked to the parametric changes underlying the Jespersen Cycle, but that examination of these changes informs an account of the cycle. They include changes to redundant negation, negative inversion and negative concord.

Of these, the relationship between the Jespersen Cycle and negative concord is the most widely discussed (see for example Haegeman (1995); Haegeman and Zanuttini (1996); Rowlett (1998); Zeijlstra (2004, 2008)). Languages are either typically[1] negative concord languages or double

[1] See Larrivée (2015) for discussion of double negation readings in negative concord languages.

negation languages. In negative concord, as the Middle English example in (4) illustrates, a clause contains more than one negative word but it receives a reading of sentential scope negation irrespective of how many negative words it contains.

(4) but he was so hard, þat **no** begger might gete **no** good of hym by **no** maner wyse
but he was so hard that no beggar might get no good of him in no manner way
'But he was so hard-hearted that no beggar might get any good of him in any kind of way'
(15th century; MIRK,104.2825)

Contrast that with standard Present-day English, in which each negative form contributes negation to the clause. Thus, Present-day standard English (5) with two negative forms receives an affirmative reading – each negative form is interpreted as negative so their effect is to cancel each other out.

(5) No-one said nothing at the meeting = Everyone said something

Many non-standard varieties of Present-day English (PDE) exhibit negative concord but Standard English does not. Nevalainen (1996) argues that this change is, in part, linked to the standardisation of English in the sixteenth century. However, as I argue in Chapters 7–9, this does not entirely explain the decline in negative concord, which begins much earlier in the thirteenth and fourteenth centuries. I argue that this earlier decline in negative concord is a consequence of the same parametric changes that drive the Jespersen Cycle. Furthermore, I argue that the distinction between not-negation (6a) and no-negation (6b) observed by Tottie (1991a) and the distribution of these two variants in both historical and Present-day English falls out of this account.

(6) a. I didn't see anyone
b. I saw no-one

The other two phenomena I focus on – redundant negation and negative inversion – are less well studied. A redundant negative is a negative word that appears in the complement clause of certain verbs such as *deny, forbid or prohibit*, but does not contribute negative force to the clause, as illustrated by (7).

(7) You may deny that you were **not** the meane of my Lord Hastings late imprisonment
You may deny that you were not the cause of my Lord Hastings late imprisonment

> 'You may deny you were the cause of my Lord Hastings's recent imprisonment'
> (Richard 3, I.iii.502–503, van der Wurff 1999, 301, ex.14)

Present-day English negative inversion is described by Klima (1964) and is central to the analyses of negation proposed by Haegeman (1995, 2001). The term describes the pattern seen in example (8).

(8) Never have I been so reviled
(BNC, Margret Forster Lady's Maid, c.1990)

When a clause-initial negative takes sentential scope, we find subject-operator inversion. The canonical order of subject–verb found in English declaratives is reversed. Van Kemenade (2001) identifies a similar pattern in Old English as early as the ninth century.

The distributions of both redundant negation and negative inversion change over time. Their availability and forms differ at successive stages of the Jespersen Cycle in ways that are problematic to previous accounts of redundant negation such as van der Wurff (1999b), or negative inversion such as Nevalainen (1997) and Ingham (2007), but in ways that inform the analysis of each stage of the Jespersen Cycle in the present account.

1.3 Methodological Approach

1.3.1 Modelling Morphosyntactic Change Using Diachronic Corpora

Previous accounts of the Jespersen Cycle have sought their evidence in historical data, but taken a largely qualitative view of those data. By contrast, the present analysis develops a mixed methodological approach to corpus data which is not commonly practiced in analyses of diachronic change.

On an empirical level, any theoretical account should be testable against naturally occurring linguistic data, such as data from corpora of historical texts. However, there are various ways of using corpora as evidence in historical linguistics. The most common is as a source of qualitative data. Thus the existence of certain forms or patterns in the data at certain periods of time are the phenomena that a theoretical analysis must account for. This is a synchronic approach to data from successive periods of time. Such accounts are focused on the inputs to and products of change. Processes of change are inferred from qualitative data. This kind of analysis will inform the analysis of *ne* and *not* in Chapters 3–5 and the analysis of negative concord in Chapter 8. However, I will argue it alone is not sufficient to characterise the Jespersen Cycle.

A growing body of recent work in historical linguistics has become informed by the quantitative methods involved in variationist sociolinguistics. This work uses corpora not only as a source of synchronic data,

but also as a basis from which to model diachronic change in progress. For example, see the work of Kroch and his colleagues on morphosyntactic change, beginning with Kroch (1989); or the work of Nevalainen and her colleagues on historical sociolinguistics, for example Nevalainen and Ramoulin-Brunberg (2003). These accounts focus on diachronic change as a process and aim to account for the distribution of forms undergoing change.

Change proceeds through competition between mutually exclusive and functionally equivalent forms which represent competing ways to set a syntactic parameter. Morphosyntactic change proceeds along a logistic curve, therefore we can model changes in progress using logistic regression and compare constraints on the distribution of competing forms at different points in time. Each formal account of the stages involved in the Jespersen Cycle implicitly or explicitly provides a basis for modelling the diachrony of the Jespersen Cycle: how we structure the parametric shifts within the Jespersen Cycle makes predictions about the distribution of competing forms as change progresses in corpus data. Kroch (1989, 235, fn.29) suggests:

> Once the principle that contexts change together when they are surface reflexes of a single grammatical competition becomes firmly established, it may be possible, on occasion, to choose among grammars proposed on the basis of synchronic analysis by the predictions they make as to which contexts should change together.

Syntactically independent forms have independent distributions in corpus data, whereas the distributions of forms that are reflexes of a single syntactic parameter will pattern together in diachronic corpus data. Changes that are reflexes of a single parameter should observe the Constant Rate Effect throughout the course of the change. As syntactic analyses structure the parameters involved in the Jespersen Cycle in different ways, so they make different predictions about what competes with what, how variation and change in the distribution of forms is structured, and how it will pattern over time. These predictions are tested against observed patterns of variation and change in corpus data using logistic regression models. This places empirical constraints on possible analyses: only analyses which accurately predict the distributions of *ne* and *not* we observe in the diachronic corpora can be considered appropriate analyses of the Jespersen Cycle.

For example, if the Jespersen Cycle involves a single form of *ne*, *ne* should be subject to the same distributional constraints throughout the Jespersen Cycle, and these constraints should be constant over time. On the other hand, if as I hypothesise, the Jespersen Cycle involves syntactically different forms of *ne* at stage one and stage two, then their distributions in the corpus data will be independent. This provides a very precise empirical basis to evaluate different formal hypotheses of change.

1.3.2 *Modelling Functional Change*

In Chapter 6, I argue that this methodological approach also provides an empirical basis to ascertain the role of functional change within the Jespersen Cycle. Parametric accounts of change face the issue of redundancy. Why would a language have two ways of expressing the same thing, particularly if those ways are functionally equivalent and differ only in form? In one sense, locating parametric variation in the lexicon lessens this problem – lexical synonyms provide some evidence for redundancy. However synonyms may exhibit subtle differences in meaning or differ in their pragmatics. By extension, we might ask on what linguistic levels competing grammatical forms are equivalent and on what levels they are distinct.

Recent accounts of the Jespersen Cycle in languages other than English, principally for French (Detges and Waltereit, 2002; van der Auwera, 2009; Hansen, 2009; Hansen and Visconti, 2009), make functional change fundamental to the grammaticalisation of new negative markers. New negative markers emerge in pragmatically marked contexts, and gradually come into competition with the established negative marker as they spread from pragmatically marked to pragmatically unmarked environments. These accounts imply that functional constraints on new forms weaken as their frequency increases. This challenges Kroch's observation that the the constraints on an innovative form, when estimated probabilistically, will remain constant over the course of the change. Thus accounts invoking functional change make predictions about the diachrony of the Jespersen Cycle, which can be modelled statistically and tested for fit against diachronic corpus data. By testing the distributions of *ne, ne...not* and *not* predicted by these statistical models against the observed distributions of *ne, ne...not* and *not* in corpora, not only can we establish the number of stages required in a model of the Jespersen Cycle and their appropriate syntactic analysis, we can also identify functional shifts involved in the cycle, and model the interaction between the syntactic changes and these functional changes.

1.3.3 *Diachronic Corpora*

Given the mixed methodological approach I adopt, corpora are crucial to this work, not only as sources of qualitative data, but also as sources of quantitative data. The way the distribution of forms patterns over time can only be examined using large scale diachronic corpora that cover a long timespan. Thus the use of such corpora is essential to test the fit of models of morphosyntactic and functional change.

For such work, the corpora must meet a number of criteria. As in any corpus study, the corpus must be a representative sample of the language varieties under examination. However, particularly in studies of change, it is important that the corpus design is such that when the data are subdivided into periods for diachronic analysis, the data from each of those successive

sub-periods are comparable. This ensures that changes in the frequency of a form across successive sub-periods is in fact diachronic change rather than variation resulting from inconsistencies in the composition of the corpus from one period to the next. This problem is exacerbated in historical linguistic research because of the paucity of sources available to us and the narrow demographic of individuals who produced written texts. Therefore there are concerns about whose language they can be said to represent. Despite this, the corpora must be balanced as far as possible from one sub-period to the next. Syntactically parsed corpora are preferred for the investigation of morphosyntactic variation and change, simply because syntactic structures are the object of study and such corpora make those structures easier to interrogate. Finally, as much of the analysis is concerned with testing the fit of models of diachronic change to corpora, in quantitative and statistical terms, the corpora need to be large enough to allow statistical analysis to be performed. On this basis, the corpora selected are the Penn and York historical corpora as follows:

Date	Corpus	Word-count
800–1100	York Helsinki Parsed Corpus of Old English Prose (YCOE) (Taylor et al., 2002)	1.5 million
1100–1500	Penn Helsinki Parsed Corpus of Middle English 2nd edition (PPCME2) (Kroch and Taylor, 2000)	1.2 million
1500–1700	Parsed Corpus of Early English Correspondence (PCEEC, Taylor et al. (2006))	2.2 million
1500–1700	Penn Helsinki Parsed Corpus of Early Modern English (PPCEME) (Kroch et al., 2004)	1.7 million
		6.6 million

For Present-day English, the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) are used. The BNC is a 100 million word corpus of spoken and written British English collected and compiled during the 1980s and early 1990s. COCA is a 520 million word corpus of spoken and written American English dating from 1990 to 2015. Both are grammatically tagged, but not syntactically parsed.

1.4 Outline of the Study

I hypothesise that bipartite *ne…not* is split into two syntactically distinct and competing *ne…not* stages. In the first of these, *ne* is the negative marker, and in the second the negative marker is *not*. The bipartite *ne…not* construction involves (at least) two syntactically and functionally distinct competing forms of *not* and two syntactically and functionally distinct competing forms of *ne*. The analysis therefore needs to establish where *ne* and *not* are negative markers and then what *ne* and *not* are when they are not negative markers.

10 Introduction

I argue that formalising the syntactic differences between the two types of *ne* and the two types of *not* requires a morphosyntactic feature based account, making use of the distinction between semantically interpretable and semantically uninterpretable formal features invoked in Minimalist frameworks. Competition between semantically interpretable and semantically uninterpretable negative items derives changing patterns of redundant negation, negative inversion and – in combination with a quantifier cycle – the changing availability and form of negative concord in Middle English.

The argument proceeds as follows, beginning with the Jespersen Cycle. Chapter 2 presents evidence from quantitative data that demonstrate there are two competing forms of *ne*. Chapter 3 presents evidence to characterise each form of *ne*. Chapter 4 presents evidence for two distinct forms of *not* and evidence to characterise them syntactically and functionally. Chapter 5 then presents a formal analysis of the Jespersen Cycle, focusing on how the relationship between *ne* and *not* at successive stages is formalised in terms of features and dependencies. Chapter 6 examines functional change within the Jespersen Cycle. The patterning of functional constraints over time provides empirical evidence to identify the place of functional change within the cycle and argue for a particular relationship between formal and functional change.

Chapters 7 and 8 turn to negative concord, identifying and formalising links between the Jespersen Cycle, changes to negative concord, and the availability of negative inversion. Chapter 9 extends this analysis, arguing that variation between no-negation and not-negation is a reflex of the Jespersen Cycle that persists into Present-day English. Chapter 10 concludes the study with a model of the Jespersen Cycle, and discusses its implications for theories of morphosyntactic change.