

# The Loss of Bare Singular Noun Phrases in the History of English

Kenneth Hanson, Cristina Schmitt, and Alan Munn

Michigan State University

16th Diachronic Generative Syntax Conference

Budapest, July 3-5, 2014

## Overview

### Background

- There are well-known inter- and intra-language asymmetries in the behavior of singular count nouns in argument and predicate position.
- Even in languages which generally prohibit or severely restrict bare nominals in argument position (e.g. French), bare singulars regularly appear in predicative positions.
- On the other hand, Present Day English (PDE), which freely allows bare plurals and mass nouns in argument position, generally disallows bare singular predicates.
- In this paper we show that the loss of bare singular predicates largely took place during the Middle English (ME) period.

### Analysis

- We suggest that the loss of these kinds of predicates in PDE is due to the emergence of the indefinite determiner (cf. Crisma, to appear) and the loss of gender distinctions in English (cf. Curzan 2003).
- As a result, English lost the ability to separate pure syntactic agreement from semantic number, forcing the appearance of the indefinite article in singular predicates (Munn and Schmitt 2005; Schmitt and Kester, to appear).

## Bare Singulars in ME

(1) Seoððen nes he neauer meiden þe unhalre.  
'Afterwards though, was he not a virgin nevertheless complete.'  
(CMANCRIW-1,II.128.1667) [c. 1200]

(2) God wimman scæ wæs  
'She was a good woman'  
(CMPETERB,59.602) [c. 1150]

(3) alse me king undertstant þanne he to his home cumeð  
'as one receives a king when he comes to his home'  
(CMTRINIT,115.1552) [c. 1225]

(4) þe frerys telde þe ʒemen þat woman sche was  
'the friars told the yeomen that she was a woman'  
(CMKEMPE,129.3007) [c. 1450]

## Methods

Two sets of data were coded: (1) all lexical count and mass noun phrases, and (2) noun phrases with heads from a selected list of count nouns. All data came from the PPCME2, PPCME, and PCEEC.

Nouns on the selected list were chosen to fill one of four semantic classes: (i) unique role, (ii) non-unique role, (iii) animate non-role or (iv) inanimate concrete noun.

- Roles were determined unique/non-unique based on a norming study of native English speakers who judged whether an article was necessary.
- Unambiguous spellings of the words were lemmatized, and noun classes coded using CorpusSearch.

For details of lemmatization and coding, see Hanson 2014.

## Results

Because the bare option is in competition with the indefinite article in most cases (for unique roles, it is the definite article), for some graphs the data was restricted to noun phrases with the indefinite article (definite for unique roles) and bare nouns only.

### Token Counts

- A total of 457835 tokens were obtained for all nouns (Table 1), and 14987 tokens for the selected list (Table 2).

Table 1: Data Counts, All Nouns

Period	Predicate	V-Arg	P-Obj
1150-1250	1488	10139	12282
1250-1350	967	5749	8487
1350-1420	3523	25046	38239
1420-1500	2443	18625	28218
1500-1569	3800	33344	50248
1570-1639	6550	54620	72727
1640-1720	5118	31536	44684

Table 2: Data Counts, Selected Nouns

Period	Predicate	V-Arg	P-Obj
1150-1250	155	1206	500
1250-1350	52	458	327
1350-1420	656	2312	1640
1420-1500	507	2079	1176
1500-1569	312	2226	1607
1570-1639	506	2388	2265
1640-1720	353	1233	976

### All Nouns

Results are presented in terms of probability that a singular noun will be bare in the corpus.

Figure 1: Probability Bare for All Lexical Noun Phrases by Position (Logistic Model)

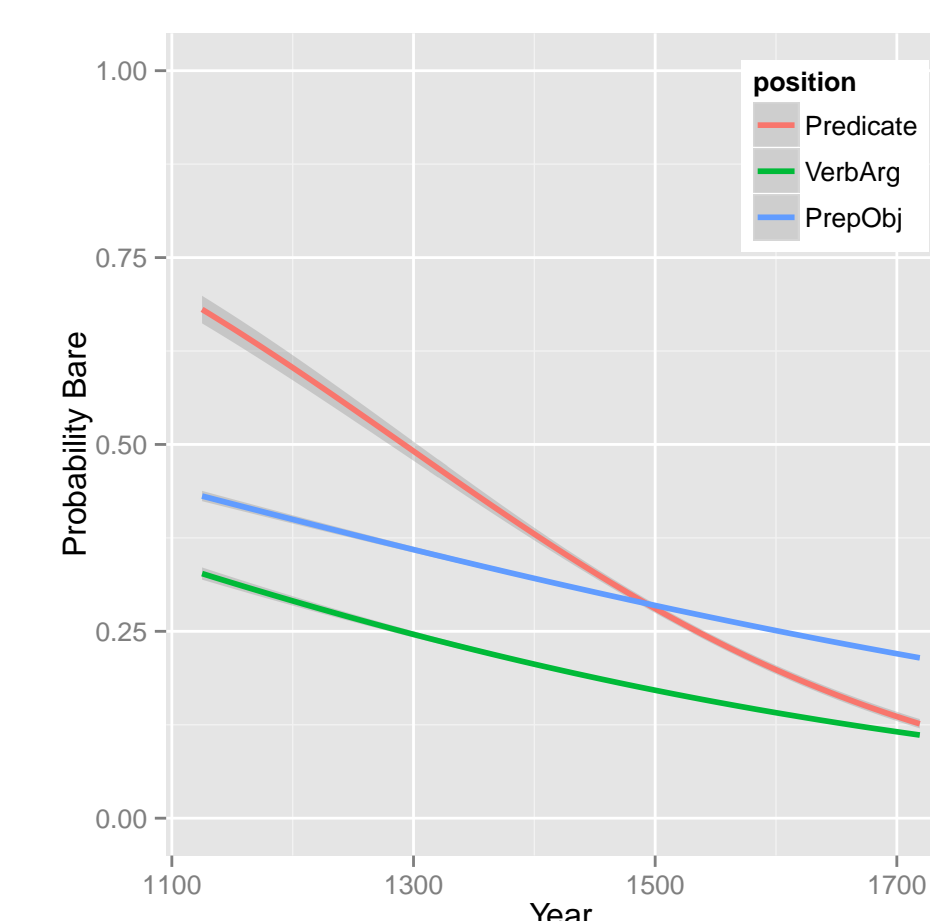
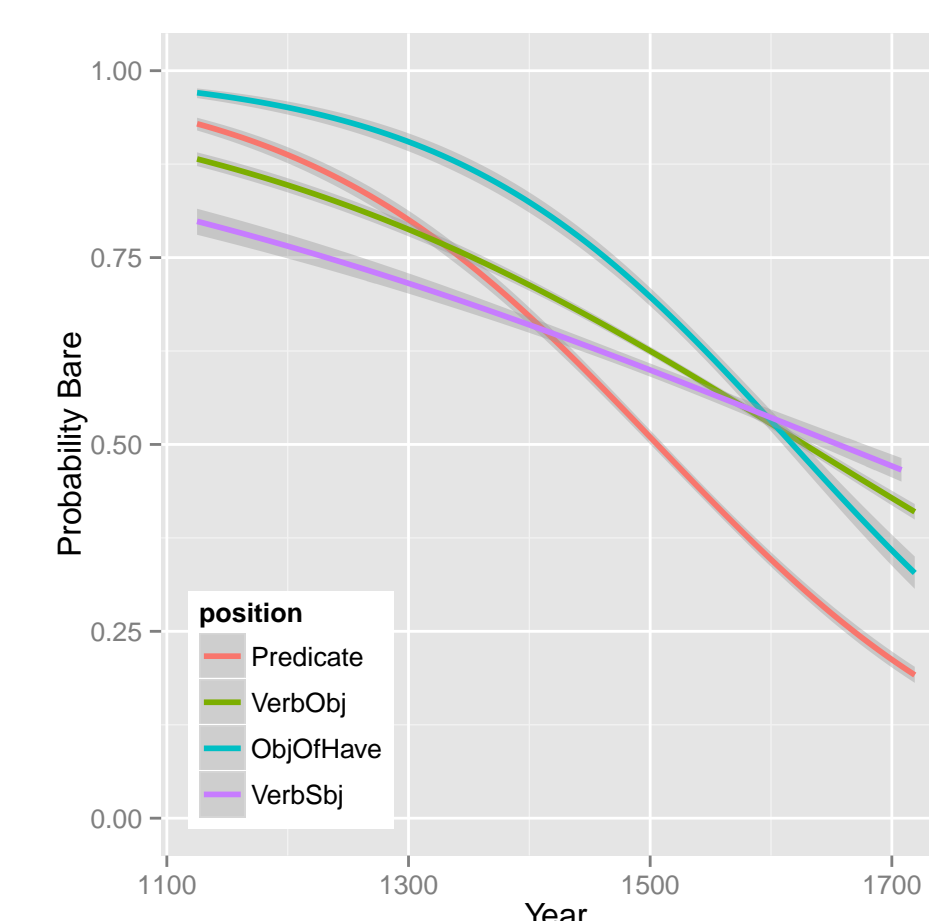


Figure 2: Probability Bare for All Lexical Noun Phrases by Position, Indefinites Only (Logistic Model)



Noun phrases from both data sets were coded for syntactic properties (position; presence of definite/indefinite articles, other determiners, quantifiers, possessors, and numerals; adjective modification), and date of the source text.

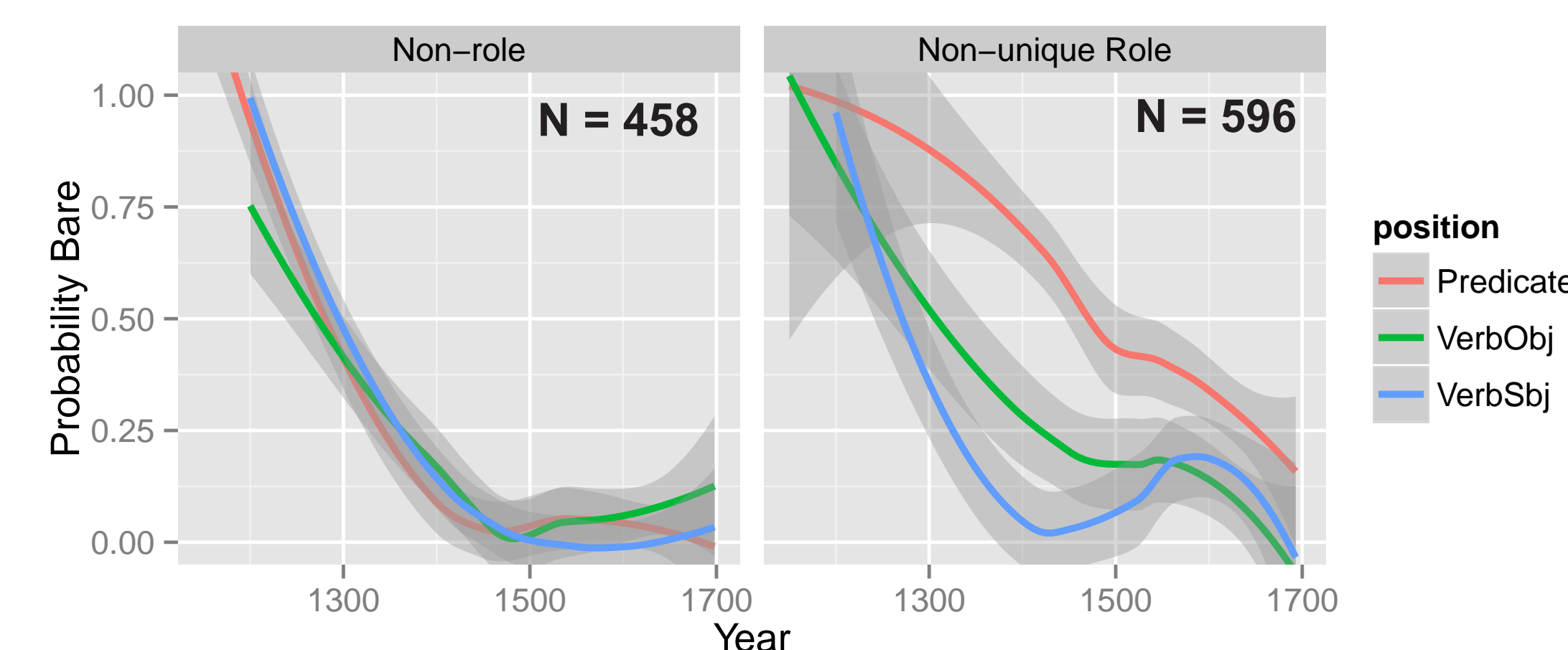
- Noun phrases were coded as bare if they did not contain a (i) determiner, (ii) quantifier, (iii) possessor, or (iv) numeral.
- Noun phrases containing compounds and conjunctions were filtered out.

Graphs of the proportion of noun phrases that appear bare over time by class and position were produced with LOESS and binomial logistic regressions using R and ggplot2.

- Bare nouns are predominant in predicate position but still possible in argument position in early ME.
- Restricting the data to indefinites (defined as containing an indefinite article only) and bare nouns, it is clear that the bare option was overwhelmingly preferred in all positions c. 1150, and was gradually lost over the ME and EME periods (Figure 2).

## Selected Count Nouns

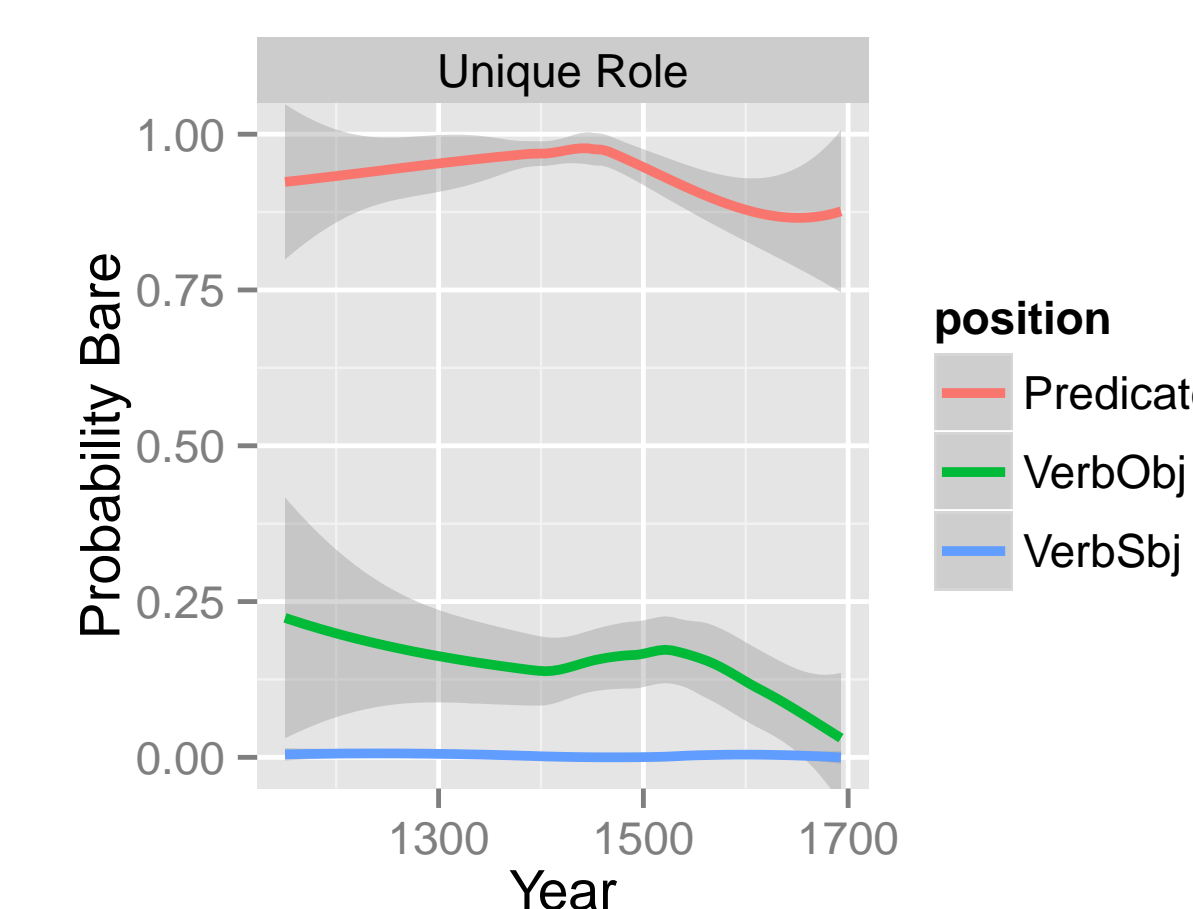
Figure 3: Probability Bare for Selected Roles and Non-roles by Class and Position, Indefinites Only (LOESS Models)



### Role Nouns

- Comparing role nouns to non-role nouns we find a much faster drop off with non-roles in predicate position.
- Unique roles behave differently, as expected since they are still possible in predicate position in PDE.

Figure 4: Probability Bare for Unique Roles by Class and Position, Definites Only (LOESS Model, N = 2855)



## Discussion

### The Indefinite Article

- Although the emergence of an indefinite article in English (cf. Crisma, to appear) is implicated in the loss of bare singular predicates, it cannot be the explanation for it since other languages have indefinite articles and still allow bare predicates.
- We hypothesize that the complete loss of gender in English (Curzan 2003; Markus 1988) is also implicated in the loss of bare singular predicates. Gender can be taken as the instantiation of Bobaljik's 1995 Free Agr parameter which Munn and Schmitt 2005 use to account for the bare singular predicates in the Romance languages.

### The Role of Gender in the Loss

- To test the hypothesis that gender might be relevant to the loss of bare singular predicates we examined the proportion of gender distinct definite determiners to uninflected definite determiners.

Table 3: Loss of gender distinctions as reflected by proportion of inflected definite determiners

Period	Uninflected	Inflected	(%)
1150-1249	2072	1519	73.3
1250-1349	1806	320	17.7
1350-1419	6741	193	2.9
1420-1500	6405	368	5.6

- The data shows a steep drop-off after 1250, but not a complete loss until 1350 which corresponds well with the drop-off of non-role predicates.

## Acknowledgements

- This work was partly funded by the MSU College of Arts and Letters Undergraduate Research Initiative (CAL-URI) and the MSU Office of Undergraduate Research.
- Special thanks to: Beatrice Santorini for help with writing CorpusSearch queries, Aaron Ecay and Ashley Hesson for help with statistics and modeling, and Aaron Ecay again for providing us with a query file for text dates.

## Selected References

- Crisma, Paola. to appear. The "indefinite article" from cardinal to operator to expletive. In Chiara Gianollo, Agnes Jäger & Doris Penka (eds.), Language change at the syntax-semantics interface, Mouton de Gruyter.
- Curzan, Anne. 2003. Gender shifts in the history of English. Cambridge: Cambridge University Press.
- Hanson, Kenneth. 2014. Methods for tracking lexical classes in parsed historical corpora. Poster presented at the Michigan State Undergraduate Linguistics Conference (MSULC), East Lansing, MI.
- Markus, Manfred. 1988. Reasons for the loss of gender in English. In Di-eter Kastovsky & Gero Bauer (eds.), Luick revisited: papers read at the Luick-Symposium at Schloss Liechtenstein 15-18.9.1985, 241-258. Tübingen: Gunter Narr.
- Munn, Alan & Cristina Schmitt. 2005. Number and indefinites. Lingua 115. 821-855.
- Schmitt, Cristina & Ellen-Petra Kester. to appear. Predicate nominals in Papiamentu: a comparison with Brazilian Portuguese and other languages. In Ana Aguilar, Bert LeBruyn & Joost Zwarts (eds.), Advances in weak referentiality, John Benjamins.