

Typological Studies in Language (TSL)

A companion series to the journal *Studies in Language*. Volumes in this series are functionally and typologically oriented, covering specific topics in language by collecting together data from a wide variety of languages and language typologies.

General Editor

Michael Noonan
University of Wisconsin-Milwaukee

Assistant Editors

Spike Gildea
University of Oregon

Suzanne Kemmer
Rice University

Editorial Board

Wallace Chafe
Santa Barbara

Matthew S. Dryer
Buffalo

Paul J. Hopper
Pittsburgh

Ronald W. Langacker
San Diego

Doris L. Payne
Oregon

Sandra A. Thompson
Santa Barbara

Bernard Comrie
Leipzig / Santa Barbara

John Haiman
St Paul

Andrej A. Kibrik
Moscow

Charles N. Li
Santa Barbara

Frans Plank
Konstanz

Dan I. Slobin
Berkeley

R.M.W. Dixon
Melbourne

Jerrold M. Sadock
Chicago

Edith Moravcsik
Milwaukee

Andrew Pawley
Canberra

Bernd Heine
Köln

Volume 82

Formulaic Language. Volume 1. Distribution and historical change.
Edited by Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali
and Kathleen M. Wheatley

Formulaic Language

VOLUME 1

Distribution and historical change

Edited by

Roberta Corrigan

Edith A. Moravcsik

Hamid Ouali

Kathleen M. Wheatley

University of Wisconsin-Milwaukee

John Benjamins Publishing Company
Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

This book is dedicated to the memory of our dear friend and colleague Michael Noonan (September 14 1947 – February 23 2009). The conference on formulaic language that these papers come from and the publication of this book would not have happened without Mickey's enthusiastic leadership, encouragement, and loving attention.

Library of Congress Cataloging-in-Publication Data

Formulaic language : volume 1 : distribution and historical change/ edited by Roberta

Corrigan, Edith A. Moravcsik, Hamid Ouali and Kathleen M. Wheatley.

p. cm. (Typological Studies in Language, ISSN 0167-7373 ; v. 82)

Includes bibliographical references and index.

1. Linguistic analysis (Linguistics) 2. Linguistic models. I. Corrigan, Roberta.

P126.F67 2009

410--dc22

2008042109

ISBN 978 90 272 2995 3 (HB; alk. paper) – ISBN 978 90 272 2997 7 (SET; alk. paper)

ISBN 978 90 272 9017 5 (EB)

© 2009 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

The role of prefabs in grammaticization

How the particular and the general interact in language change

Joan Bybee & Rena Torres Cacoullos
University of New Mexico

1. Introduction 187
2. Cognitive consequences of skewed frequency distributions in constructions 190
3. The role of prefabs in grammaticization: English *can* 193
 - 3.1 Prefabs with 'say' and 'tell' 194
 - 3.2 Cognitive verbs 197
 - 3.3 Frequent items as the centers of expanding classes 197
4. Spanish progressive and other imperfective gerund periphrases 198
 - 4.1 The grammaticization of the progressive in Spanish 198
 - 4.2 Grammaticization indices for Spanish progressives: locatives and unithood 200
 - 4.3 Prefabs and *estar* + *V-ndo* grammaticization 204
 - 4.3.1 Prefabs lead as units 205
 - 4.3.2 Prefabs contribute to productivity via associated semantic classes 207
 - 4.3.3 Prefabs and productivity: Evidence from motion-verb gerund periphrases 209
5. Conclusions 212

Abstract

Studies of grammaticization often reveal skewed distributions of lexical items in grammaticizing constructions, suggesting the presence of prefabs using these constructions. We examine here the role of prefabs in the grammaticization of *can* in English and the progressive *estar* 'be (located)' + *V-ndo* (Gerund) in Spanish. The data suggest that prefabs play a role in advancing formal and semantic change. We argue that (1) prefabs are ahead of the general construction in unit-hood status in early stages and thus demote the independent lexical status of the emerging grams, and (2) in their association with semantic classes of which they are the most frequent member, prefabs promote the productivity of the general construction. The evidence shows that prefabs maintain associations with the related general construction.

1. Introduction

The fact that natural discourse relies heavily on repeated, conventionalized multi-word strings has implications for processing theories, production theories and grammatical theory. In this paper we explore the interaction of specific conventionalized multi-word strings, which we will call “prefabs” (following Erman and Warren 2000), and the more general constructions that make up the grammar of a language. In particular we focus on the way prefabs participate in the process of grammaticization by studying prefabs that have developed along with the Progressive construction with *estar* in Spanish and the auxiliary construction with *can* in English.

Our theoretical perspective is that of usage-based construction grammar in which cognitive representations are affected by the speaker’s experience with language (Goldberg 2006; Bybee 2006). Tokens of experience are represented in memory as exemplars of varying strengths. The representations of constructions consist of categories that group together all the exemplars of a given construction, based on semantic and formal similarity.

The model of lexical associations proposed for morphologically complex words in Bybee (1985, 1988, 2001) can be extended to multiword units and constructions. In this model, associations made among related forms are gradient and depend upon the degree of semantic and phonological similarity and the token frequency of the specific items (as we explain below). One of the main determinants of memory storage is frequency in experience; thus specific instances of constructions may occur as units in memory storage, even if their meaning and form is predictable from the more general construction. An expression such as *that drives me crazy* may occur as a unit of storage and may be accessed in one step. However, when stored units are themselves complex they can still be related in representation to the smaller units that comprise them (*that*, *drives*, *me*, and *crazy*) as well as to the general construction the stored expressions instantiate.

In this view, there is no discrete division between fixed expressions and productive formations, rather, these two types of linguistic expressions form the two poles of a continuum. Evidence of the continuum between the processing of fixed expressions and productive constructions includes the fact that even highly fixed expressions sometimes undergo expansion, as when a radio news reporter was heard to say *all chaos broke loose*. One might have thought that *all hell broke loose* was entirely fixed, but he was able to make a substitution inside this expression. Some expressions allow for considerable expansion, as when the adjectives that follow *drives someone + ADJ* are studied in a corpus. It is found that *mad*, *crazy*, *insane*, *wild*, *nuts*, *up the wall* and several others are possible in this construction (Boas 2003).

Moreover, constructions that are often thought of as rather general and schematic nonetheless often have lexical restrictions, as the ditransitive construction occurs only with a certain set of verbs, such as those denoting ‘giving’: *give*, *pass*, *had*, *sell*, *trade*, *lend*, *serve*, *feed* and other classes as well (Goldberg 1995: 126).

Given this continuum, we can identify the dimensions that determine the variation from one pole to the other. In this paper we present evidence for the following three dimensions.

- i. Productivity: If the expression is schematic, that is, if it has slots that can be filled by a class of items, then it will also vary on a scale of productivity depending upon the number of types that can occupy its open position and the semantic generality of the class.
- ii. Transparency of meaning: fixed expressions can have fully compositional meaning, as in expressions such as *open the door* or *pass the salt*. Less than transparent meaning occurs in idioms that have a metaphoric interpretation (e.g., *pull strings*) or in frequent expressions that have developed some pragmatic or semantic nuances or changes that distance them from the more general constructions with which they are related (e.g., *I don’t know* as a discourse marker [Scheibman 2000]).
- iii. Analyzability: expressions may also differ in the extent to which the units composing the expression are associated with the etymologically same units in other constructions. Diagnostics for analyzability include the ability to add modifiers or other elements that separate the units of the expression or to appear in different constructions, as for example, when the elements are recomposed into a passive.¹

Many researchers propose two modes of processing to underlie the Open Choice Principle and the Idiom Principle (as Sinclair 1991 put it; see Erman and Warren 2000; Van Lancker [this volume]; Jackendoff 2002), despite the gradient between monomorphemic units on the one hand and conventionalized, multiword sequences on the other that we have just described and for which we present further evidence below. This gradience suggests that two distinct types of processing are not involved. In contrast, we propose that the access of stored units in production and perception is the same process whether the units are simple or complex; the observed gradience is not a property of the type of processing but

1. See Langacker 1987 for a discussion of analyzability (pp. 292–298; 457–460) and compositionality or transparency of meaning (452–457).

rather of the length, complexity and degree of fixedness of the stored units. Thus accessing the stored linguistic representations is essentially the same whether the unit is a monomorphemic word, such as *wall*, a phrase such as *the wall* or a partially schematic construction such as *X drives me Y*, where the X position can be filled with almost any sort of NP and the Y position contains an adjective or prepositional phrase from the class related to *mad, crazy, up the wall*.

Besides degrees of complexity in storage, another source of complexity in utterances arises from the fact that the schematic slots in constructions can themselves be filled with either simple or complex material. Dąbrowska and Lieven (2005) use the term “superimposition” to describe the process by which an accessed unit is used to fill a position in a partially schematic accessed unit or construction. To use Dąbrowska and Lieven’s example, *shall I PROCESS?* (where *PROCESS* stands for the set of verbs or verbal complexes that may occur in that position) and *open that* can be superimposed to derive the expression *shall I open that?* All the properties of the two units – their phonetic form, meaning and pragmatics as derived from previous experience are carried along in the process of superimposition. Note that units involved in superimposition may in themselves be the result of superimposition, as in the example *open that* which was derived by superimposing *that* and *open OBJECT*. Thus the process of superimposition is one of the sources of syntactic complexity in utterances; the other source is the complexity that is inherent to the stored unit.

Given this general framework coupled with an exemplar model of linguistic representation, usage data suggests that certain exemplars of constructions have differential representation depending upon their frequency of use (Bybee 2003, 2006). One of our interests in this paper is to examine how specific exemplars of constructions affect the overall meaning and use of the construction. We cast this question in a diachronic context and examine the way conventionalized instances of constructions or prefabs interact with the more general construction as grammaticization proceeds. Rather than viewing prefabs as something distinct from and perhaps peripheral to grammar in the traditional sense, we argue that prefabs constitute important loci of grammatical development in the diachronic domain. By implication, such conventionalized expressions have important interactions with more general constructions in the synchronic domain.

2. Cognitive consequences of skewed frequency distributions in constructions

Corpus-based studies of constructions reveal an uneven topology for the distribution of lexical items in constructions. In many cases, one or a small number of

lexical items occur frequently in the construction and other lexical items occur once or twice in the construction. Thus Goldberg, Casenhiser & Sethuraman (2004) find that in mothers’ speech to children aged 20 to 28 months certain verbs occurred frequently in certain constructions: for instance, *go* accounted for 39% of verbs in *SUBJ VERB OBLIQUE* constructions; *put* accounted for 38% of all *SUBJ VERB OBJECT OBLIQUE* constructions and *give* 20% of all *SUBJ VERB OBJ₂* constructions.

Bybee & Eddington (2006) studied Spanish change-of-state verbs and the adjectives that accompanied them and found that certain pairings were of very high frequency, e.g., *quedarse solo* ‘to end up alone’; *quedarse quieto* ‘to become still’; *quedarse sorprendido* ‘to be surprised’; *ponerse nervioso* ‘to get nervous’. These expressions are prefabs in that they represent the normal, conventionalized way of expressing certain commonly-referred to changes of state. It was also found that these expressions formed the centers of exemplar categories, as the corpus also contained many single examples that were related semantically to these more frequent expressions. Thus the prefabs play a central role in determining the range of use of the constructions. See Wilson (this volume) for details about the diachronic development of these constructions from prefabs.

Goldberg and colleagues (Goldberg, Casenhiser & Sethuraman 2004, 2005; Casenhiser & Goldberg 2005) argue that the skewed distribution in constructions aids in acquisition because the frequent expressions or prefabs play a crucial role in helping the child grasp the meaning of the constructions. They designed an experiment to test the contribution of type and token frequency in which both children and adults were taught a nonce argument structure construction in English. The construction had a nonce verb (with a suffix in some of the conditions) and the verb appeared at the end of the clause. The meaning of the construction was taught through a video presentation that accompanied the linguistic stimuli. In one condition nonce verbs appeared in the stimuli with the same token frequency, while in the other condition the same number of verbs was presented, but one had a higher token frequency than all the others. In the latter condition, learning was more successful. The hypothesis about the facilitation of learning is that the repetition of a particular verb in a particular construction helps to establish the correlation between the meaning of the construction and its formal expression. Goldberg (2006) goes on to demonstrate that in category learning in general a centered, or low variance, category is easier to learn. The condition with one instance of higher token frequency is just such a category.

Lieven and colleagues (Lieven et al. 1997; Lieven et al. 2003; Dąbrowska & Lieven 2005; Lieven et al. this volume) demonstrate that early children’s utterances are strongly based on utterances the children have experienced before, in

their own speech or in the speech of adults. Dąbrowska & Lieven (2005) argue that children start their acquisition of grammar with multiword sequences that are rather fixed and repeated verbatim and gradually learn to substitute lexical items into the slots in the construction represented by the sequence. Thus the analysis of the repeated utterances and the build-up of more abstract and schematic constructions emerges gradually out of experienced and repeated tokens. However, even after the more abstract constructions are established (say, in adults), many utterances may still be produced by accessing large, pre-assembled and lexically-specific sequences from memory.

These studies, then, all show a significant interaction of prefabs with more combinatorial tokens of constructions. This is possible because prefabs have not necessarily lost their internal structure, nor have their component parts necessarily lost their identities. Nunberg et al. (1994) argue that many phrases taken as idioms actually retain their compositionality in the sense that their parts “carry identifiable parts of their idiomatic meanings” (496). In addition, such “idiomatically combining expressions” retain their morphosyntactic analyzability. Thus it is argued that even in expressions with unpredictable meaning, such as *pull strings*, the two words each still contribute to the idiomatic meaning in the sense that one can identify for any given case what or who were the “strings” and what was done to “pull” them. So if such idioms have discrete parts that are associated with other VERB – OBJECT constructions as well as with other instances of the lexical items involved, then other sorts of prefabs can certainly have these properties as well. That is, despite holistic processing and chunk-like storage, prefabs can still be related (to varying degrees) to the words and constructions of which they are constituted. It follows then, that in language change, prefabs might have an impact on the nature and rate of change in constructions.

It is known from studies of discourse variation and grammaticization that increasing token frequency of an expression leads to increasing opacity of internal structure and increasing autonomy from the more general construction, which enables the resulting single processing unit to gain new discourse-pragmatic functions (Bybee 2003: 618; cf. Thompson & Mulac 1991; Company 2006; Torres Cacoullous 2006). Nevertheless, we argue that prefabs can maintain associations of gradient strength with the more general construction unless and until increases in frequency and concomitant semantic/pragmatic change reach high levels.

In a study of current variation reflecting ongoing grammaticization, Torres Cacoullous & Walker (2009) showed that the patterns affecting the general construction also affect fixed formulas: even though prefabs develop their own discourse-pragmatic characteristics, they retain traces of the constraints on their associated construction. These researchers used multivariate analysis to discover

a number of language-internal factors conditioning the variation between *that* presence and absence in naturalistic speech data. *I think, I guess* and a handful of other frequent 1st person singular and Present tense collocations (*I remember, I find, I'm sure, I wish, I hope*) have become conventionalized as discourse formulas that function more as epistemic or evidential adverbial phrases than as main-clause propositions (e.g., Thompson & Mulac 1991; Diessel & Tomasello 2001; Thompson 2002). Torres Cacoullous and Walker found that even though the rate of *that* with prefabs *I think, I guess* is low, the linguistic conditioning parallels instances of the more general construction with more robust variation: the two strongest constraints, intervening material and type of subject, are both operative, and with the same direction of effect (the presence of intervening material and full NP subjects favor *that* presence). They argue that not only do grammaticizing constructions retain lexical meaning (Bybee & Pagliuca 1987; Hopper 1991), but prefabs retain grammatical properties, manifested in the parallelism of constraints on variation. Thus, the units of formulaic language maintain associations with productive constructions, contra the view that would isolate the former in a lexicon separate from the grammar.

Other studies of grammaticization have also revealed skewed distributions of lexical items in grammaticizing constructions. It is often noted that grammaticization gets its start in constructions with particular classes of items. For instance, Carey (1994) finds that the Old English resultative construction that becomes the Perfect was used most frequently with verbs of mental state and reporting verbs and its meaning first conventionalized in expressions with these verbs. Thus to study both the meaning changes in grammaticization and the way grammaticizing constructions expand and generalize, it is instructive to examine the use of such constructions in prefabs.²

If prefabs are processed more holistically than more compositional word combinations, the meaning of the individual units making up the expression will be less transparent. We regard the effect of holistic processing to be cumulative; the more often a sequence is accessed as a whole unit, the stronger the path to that type of access will become (Hay 2001). We will argue that the cumulative effect of this more holistic processing contributes to the pragmatic and semantic changes that occur in grammaticization. Our consideration of diachronic data on the development of the English auxiliary *can* from the Old English verb *cunnan* ‘to know’ and

2. Another study of grammaticization that shows how specific instances of constructions interact with more general ones is Traugott (in press). Traugott argues that certain partitive modifiers, such as *a kind/sort/bit/lot of* break off (so to speak) from the Partitive construction and realign themselves with the extant Degree Modifier construction.

the development of the Spanish auxiliary constructions that express the progressive have led us to the following more specific hypotheses:

First, prefabs are more advanced than the general construction in unit-hood status. As a result, the independent lexical status of the emerging gram is weakened with the effect that the gram within the prefab may be bleached of its meaning, thus contributing to the general bleaching of the meaning of the gram.

Second, in their association with semantic classes of which they are the most frequent member, prefabs promote the productivity of the general construction.

Third, a more minor tendency is that an emerging gram can be locked in a prefab, the whole of which retains an older meaning.

3. The role of prefabs in grammaticization: English *can*

Bybee (2003) traces the development of the modal auxiliary *can* from OE *cunnan* 'to know' through the end of the ME period. In OE, *cunnan* had limited use with infinitive complements; it occurred primarily with the following three classes of infinitives:

- i. communication verbs, such as 'say' or 'teach', where *cunnan* meant to have the knowledge to say or teach truthfully;
- ii. cognitive verbs such as 'understand', 'comprehend' or 'perceive'. As argued in Bybee (2003), these infinitives are harmonic with the 'know' meaning of *cunnan*, reinforcing it and sometimes adding more specific meaning;
- iii. verbs indicating skills, again reinforcing the perhaps weakening meaning of 'know', as in 'I know the harp' by adding 'to play'.

In the Middle English (ME) texts composed by Chaucer, *can* (or *kan*) had a greatly expanded range of usage, but it continued with the same verbs and verb classes found in OE. Bybee (2003) argues that the new verbs used with *can* are related to the earlier classes of OE. In addition, in Chaucer's texts, Bybee notes some prefabs that can be identified by their relative frequency of occurrence and that the frequency of use of these tokens (such as *I can say you namoore*) may contribute to the bleaching of the meaning of *can*.

The current study investigates the latter proposal in more detail, considering the meaning of *can* in these prefabs compared to its meaning in other combinations. We find that with reporting verbs, the prefabs seem to lead to a meaning change from 'having knowledge to say' to 'being able to say' while for the cognitive verbs, where the combination of modal with main verb is harmonic, the older

usage is retained into ME and perhaps even into present day English. In this case, the older distribution is maintained, but *can* adds very little meaning.

3.1 Prefabs with 'say' and 'tell'

In the 300 tokens of *can* examined from the works of Geoffrey Chaucer, verbs of communication accounted for 102 tokens and 31 types.³ The verbs with the highest token frequency were *tellen*, which occurred 30 times and *seye/sayn* which occurred 29 times. In general, verbs of saying and telling occur frequently in the texts because they are often used as rhetorical devices for managing the topics of the text. This is certainly true in the *Canterbury Tales*; in addition, in these tales there is often talk of who has the ability or knowledge to tell a tale and this also elevates the number of such verbs.

The following prefab with *seye* was identified on the basis of its occurrence three times in 300 tokens:

- (1) I kan sey yow namoore (B. ML. 175; B. NP. 4159; G. CY. 651)

This prefab is used as a rhetorical device to end a chunk of discourse before entering another topic or scene. In this prefab, *can* indicates a notion as general as root possibility in interlocutors' interpretation of 'I can say no more because I want to get on with my narrative'. Some variations of this prefab also occur, as in (2) which omits *yow* and puts the main verb at the end:

- (2) I kan no more seye (TC. 1. 1051)

Another variation uses a different negative element:

- (3) I kan sey yow no ferre (A. Kn. 2060)

Another possible variation on this prefab occurs with a different verb:

- (4) I kan no moore expound in this manner (B. Pri. 1725)

A different prefab shows an alternation between *sey* and *tell*. This prefab is also used as a rhetorical device to indicate the end of a portion of narrative or description. Here, however, the sense of ability is more apparent because of the adverb *bettle* which clearly points to 'ability to describe' rather than 'knowledge to say'. Note the older word order with *seye* in (5) and the word order variation with *telle* in (6). The adverb *feithfully* in (7) meant 'with faith or confidence' reinforcing the ability meaning of *kan*.

3. The tokens were the first 300 listed in Tatlock and Kennedy (1927).

- (5) I kan no better sayn (B. ML. 42; B. ML. 874; E. Mch. 1874; I. Pars. 54)
 (6) I kan telle it no better (B. ML. 881)
 (7) I kan no better telle, faithfully (D. Fr. 1433)

Outside of these prefabs, a large majority of the uses of *can seye* still express the notion of 'knowledge to say', and only a few indicate ability, as indicated in Table 1.

Table 1. Other (non-prefab) uses of *kan seye* in Chaucer's English

Knowledge to say	16
Ability	3
Both	2

The two examples where both interpretations apply have the sense of 'can tell a tale', which we interpret as involving both knowledge and ability.

In comparison, the prefab uses of *can seye* do not involve knowledge to say, but are discourse devices, in the one case with a meaning of root possibility and the other a clear meaning of ability.

The situation with *can telle* is quite similar. *Telle* occurs in the prefab shown above and also in a *more than I can + V* construction exemplified by the following. One token with *telle* occurred and the others involve different communication verbs.

- (8) A thousand foold wel moore than I kan telle (B. ML. 1120)
 (9) And mo than I kan make of menciouun (A. Kn. 1935)
 (10) And deyntees mo than I kan yow devyse (B. ML. 419)
 'And dainties more than I can describe to you'

Again, this construction appears to be used as a rhetorical device for emphasizing great quantity, but the interpretation of *can* in these examples strongly suggests ability rather than knowledge.

Outside these prefabs, *can telle* is still used preferentially to express knowledge to tell, as indicated in Table 2.

Table 2. Other (non-prefab) uses of *can telle* in Chaucer's English

Knowledge to tell	15
Ability	4
Both	3

As with *seye*, the uses that allow both interpretations have as the object of *telle* a story or tale. Two of the examples that express ability are identified by the accompanying adverb and occur in a specific construction:

- (11) and telle yow as pleyonly as I can (A. Kn. 2481)
 (12) I wol yow telle, as wel as ever I kan (A. Co. 4342)

These examples demonstrate the construction:

- (13) *as* ADVERB *as* SUBJ *can*

This construction is also used with other verbs in the corpus, as shown in the following:

- (14) Bot I wol passe as lightly as I kan' (B. NP. 4129)
 (15) As shortly as I can it trete (PF. 34)
 (16) As well as that my wit can me suffyse (PF. 460)
 (17) To serve you as hertly as I can (TC. 5. 941)

In these examples, the sense of *can* is clearly ability. The use of *telle* in this construction may be one of the means by which ability comes to be an interpretation of *can* with *telle*. Thus the expansion of specific constructions can be one means of spreading a new sense to a range of verb classes.

The conclusion of this section is that in the class of reporting verbs, the prefabricated or formulaic uses led the meaning change from knowledge to ability.

3.2 Cognitive verbs

Another major verb class that is used with *cunnan* in Old English contains cognitive verbs, such as *understandan*, *ongietan* 'understand', *tocnawan* 'to distinguish, discern' *geþenkan* 'to comprehend', and so on (Goossens 1992; Bybee 2003). As argued in Bybee (2003), these verbs are used with *can* in a way that is harmonic: the main verb echoes the meaning of *cunnan*, adding meaning that is more specific and shoring up the meaning of *cunnan* which seems to be becoming too weak to express 'knowing' on its own. These same verbs continue to occur with *can* up to the present time. Because of the harmonic nature of these expressions, *can* contributes very little to the meaning. Thus *can understand* or *can remember* are not that different in meaning from *understand* or *remember*. Indeed in most languages, no modal would be added to clauses with these verbs. Because *can* in these phrases is nearly meaningless, these expressions have likely contributed to the bleaching of *can* throughout the history of its development.

This class expands in ME as the lexicon is enhanced by borrowings from Old French. The new verbs entering the language in the 14th century come to be used with *can*. Examples found in our small corpus are: *imagine*, *conclude*, *construe*, *judge*, *remember* and *espy* (in the sense of 'discover').

3.3 Frequent items as the centers of expanding classes

We hypothesize that in the examples from the works of Chaucer the high frequency verbs are serving as the centers of the expanding classes of verbs used with *can*. This is especially clear with the two classes of verbs just discussed – the reporting verbs and the cognitive verbs. Both classes expanded greatly with the influx of lexical borrowings from Old French.

As noted above, in the Chaucer texts used, 102 of the 300 tokens were verbs of communication. There were 31 types; two of these – *say* and *tell* – accounted for 59 tokens. The evidence that these more frequent tokens serve as the central members of the category and attract other verbs with similar semantics is that some of the less frequent verbs or phrases are used in the same constructions or prefabs as *say* or *tell*. For instance, examples (9) and (10) above show *make of menciou* and *devyce* ‘describe’ in a construction also used with the more frequent verbs. Of the 31 types found, 19 are verbs borrowed from Old French, suggesting that their appearance in this construction could easily have been on analogy with the other native verbs of communication that were used with *can*.

Similarly, the class of cognitive verbs found with *can* in the Chaucer texts included 18 types. The most frequent members are native English verbs – *see*, which was used in a cognitive sense nine times, *deem* and *understand* each used six times. A borrowing, *espy* ‘discover’, was used five times. Of the other verbs and expressions in this class, ten were borrowed from Old French. Since we have argued that the origins of *can* with cognitive verbs is an harmonic construction, it follows that the new verbs and expressions were used with *can* on analogy with the established, and more frequent, verbs in this construction.

4. Spanish progressive and other imperfective gerund periphrases

The development of a set of progressive constructions from Old Spanish to Modern Spanish provides us with the opportunity to study the structural as well as semantic properties of grammaticizing constructions and their conventionalized instantiations.

4.1 The grammaticization of the progressive in Spanish

In Old Spanish (12th – 15th centuries) texts we find occurrences of a general gerund construction, in which finite forms of spatial (locative, postural, or motion) verbs combine with another verb in gerund (*-ndo*) form to mean ‘be/go VERB-ing’, as shown in (18):

- (18) Gerund construction: [Verb_{locative-postural-motion} + gerund (*-ndo*)] = ‘be/go VERB-ing’

The verbs occurring in the finite verb slot are:

- | | | |
|------|-------------------------------------|--------------------------|
| (19) | <u>Location-Postural</u> | <u>Movement</u> |
| | <i>estar</i> ‘be (located)’ | <i>andar</i> ‘go around’ |
| | <i>quedar</i> ‘remain, stand still’ | <i>ir</i> ‘go’ |
| | <i>yacer</i> ‘lie’ | <i>salir</i> ‘go out’ |
| | | <i>venir</i> ‘come’ |

The finite form is an independent lexical item with full spatial meaning, as illustrated in the 13th c. examples in (20–22). Lexical status is indicated by a co-occurring locative, which may (as in [20], ‘in that pond’) or may not (as in [21], ‘along the road’) intervene between the finite form and the gerund. Lexical status is also evident in the combination of a motion verb with another motion verb in a harmonic use, where the gerund describes the manner of motion (as in [21], ‘go (by) walking’). Finally, in (22), the repetition of *andar* ‘go around’ and *buscar* ‘look for’ separately shows that *andando buscando* is a combination of two independent lexical items.

- (20) Et alli ESTAUa el puerco en aquella llaguna BOLCANDO se (XIII, GE.II)
‘And there WAS the pig in that pond TURNING itself’
- (21) YUASSE ANDANDO por la carrera que ua al pozo (XIII, GE.I)
‘He WENT WALKING along the road that goes to the well’
- (22) Et ANDANDO BUSCANDO los. encontrasse con un omne quel pregunto como andaua o que buscaua. (XIII, GE.I)
‘And GOING AROUND LOOKING FOR them he met a man who asked him how he was going or what he was looking for’

Particular instances of this general gerund construction grammaticize, yielding a set of aspectual constructions (cf. Bybee 2006). Thus, *estar* ‘be located’, *ir* ‘go’, and *andar* ‘go around’ + V-*ndo* evolve from lexical spatial expressions into grammatical aspectual morphemes in these constructions. In present-day varieties of Spanish, these gerund periphrases cover a range of meanings in the domain of imperfective aspect (e.g., Camus Bergareche 2004). In particular, the construction *estar* + V-*ndo* as shown in (23) is on its way to becoming an obligatory expression of progressive aspect in the Present tense (Torres Cacoullós 2000, Chapter 5; García Fernández et al. 2006: 140).

- (23) [*Estar* + V-*ndo*] = progressive

Throughout the evolution of these gerund periphrases, there is retention of spatial meaning from the source construction (Bybee & Pagliuca 1987; Hopper 1991) and spatial and aspectual meanings coexist synchronically, often in the same token. For example in (24), from a corpus of New Mexican Spanish, *está cuidando*

televisión means both 'he is there, in front of the TV' (locative – lexical) and 'he is in the midst of an activity at reference time, i.e., watching TV' (progressive – grammatical) (Torres Cacoullós 2000: 9).

- (24) - *¿Aquí está?*
 - *Sí ESTÁ CUIDANDO televisión.*
 - *Oh.*
 - *Ahi en en en el cuarto allá del otro lado. Está dormido en la silla.* (NMBil/Vig)
 '- Is he here?
 - Yes he IS WATCHING television.
 - Oh.
 - There in in in the room over there on the other side. He's asleep in the chair'

At the same time, aspectual meaning is present from the earliest texts. In the next set of 13th c. examples, locative or physical motion meaning is less discernable than in (20–22), rather the meaning is more aspectual, with *estar* + *V-ndo* indicating a situation in progress (25), *ir* + *V-ndo* a gradually developing process (26), and *andar* + *V-ndo* figurative motion together with continuous meaning (27).

- (25) *cato por una finiestra & uiol estar con ella [...] como esta marido FABLANDO con su muger* (XIII, GE.I)
 'he looked through a window and saw him (be) there with her [...] as is a husband SPEAKING with his wife'
- (26) *porque non poblara el y [...] & YUAN ya las yentes SEYENDO muchas.* (XIII, GE.I)
 'so that he wouldn't settle there [...] and the people already WERE GROWING'
 (literally: went the people being many)
- (27) *el que [...] quiere andar los caminos peligrosos ANDA BUSCANDO su muerte* (XIII, Calila)
 'he who [...] wants to walk dangerous roads IS LOOKING for his death'

4.2. Grammaticization indices for Spanish progressives: Locatives and unithood

Grammaticization of the finite locative-motion verb in gerund periphrases proceeds via semantic reduction, which in this case involves the loss of spatial meaning (Torres Cacoullós 2000: 71–113). Yet we cannot establish that grammaticization is occurring by comparing isolated examples from earlier and later periods, since throughout the evolution of gerund periphrases there is retention of spatial meaning from the source construction, even in present-day examples (such as [24]). Nor would quantitative comparisons across periods of the proportion of tokens with aspectual as opposed to spatial meaning be a replicable measure, since tokens may be compatible with both meanings (again as in [24]) and analysts' interpretations may

well differ. Instead, we can show the advance of grammaticization by uncovering changes in distribution patterns.

Tokens of the gerund construction were exhaustively extracted from 13 texts, representing four periods: late 13th c. (three texts, approximately 900,000 words), late 15th c. (five texts, approx. 500,000 words), early 17th (one text, approx. 400,000 words), late 19th (four texts, approx. 350,000 words) (see Corpus, before References, and Table 6 for token counts).⁴ We first present distribution patterns for *estar* 'be located' and then some results for *ir* 'go' and *andar* 'go around' + *V-ndo*.

We hypothesize that bleaching of spatial meaning will be shown in a decrease of co-occurring locatives, in the aggregate.⁵ Table 3 shows the percentage of *estar* + *V-ndo* tokens with a co-occurring locative, in the four chronological sets. The rate of co-occurring locatives diminishes, from an average of 38% (91/238) in the 13th and 15th c. (Old Spanish) data combined, to 24% (51/217) in the 17th c. and 16% (35/217) in the 19th c. data. We take this result as a measure of loss of spatial meaning and thus advancing grammaticization.

Table 3. Co-occurring locatives in Progressive *estar* + *V-ndo*

XIII	XV	XVII	XIX
36% (37/104)	40% (54/134)	24% (51/217)	16% (35/217)
XIII-XV 38% (91/238) vs. XVII-XIX 17% (86/434) Chi-Square 15.10291903; p = 0.0001.			

A second measure of the grammaticization of Progressive *estar* + *V-ndo* and the motion-verb (*ir*, *andar*) + *V-ndo* periphrases is the degree of unithood. Bybee (2003: 603) proposes that frequent collocations become automated as single processing units, gaining autonomy in two ways. Analyzability is lost when the erstwhile individual constituents of the frequent collocation weaken their association with other instances of the same constituents and with other instances of the same construction. We examine three indices of unithood: adjacency, association, and fusion (Torres Cacoullós 2000, Chapter 2).

4. The texts are chronicles (13th and 15th c.) and novels; the 15th c. corpus includes two plays (the *Celestina* and the early 16th c. *Lozana*).

5. Locative co-occurrence need not always indicate a lesser degree of grammaticization; the locative may promote the aspectual meaning of the auxiliary when it refers to the main verb in harmonic uses (for example, *ir* plus another motion verb) (cf. Hopper & Traugott 1993: 83) or when it is incompatible with the auxiliary's original spatial meaning (for example an allative locative with *estar*).

1. Adjacency: the locative-motion finite verb and the gerund may be adjacent or they may be separated by intervening material. In the 13th c. data, nearly two thirds of *estar* + *V-ndo* tokens have an intervening locative or temporal adverbial, subject or object, or a combination of elements, as in (28) and (29).⁶

- (28) ESTÁ Melibea muy affligida HABLANDO con Lucrecia sobre la tardança de Calisto (XV, Celestina, XIV, 282)
 '[Stage instructions] IS Melibea, deeply distressed, TALKING to Lucrecia about the tardiness of Calisto' (cf. Singleton, 197)
- (29) Pero hombre, ¿estamos locos? ... ¿qué ESTÁ usted HABLANDO? (XIX, Perfecta, 284)
 'But man, are we crazy? ... What ARE YOU TALKING about?'

2. Association: multiple gerunds may co-occur, as in (30), or the finite verb may be more tightly associated with a single gerund as in (31), where *ir* is repeated for each gerund.

- (30) le YVAN MENGUANDO los bastimentos e CRECIENDO las necesidades (XV, CRC LIV, 178)
 'supplies WERE [lit: went] SHRINKING and needs GROWING'
- (31) la vida vulgar VA PENETRANDO y se VA INFILTRANDO en mi naturaleza. (XIX, Pepita, 55)
 'ordinary life IS (gradually) [lit: goes] PENETRATING and IS [lit: goes] INFILTRATING my nature'

3. Fusion: object pronouns may appear as enclitics on the gerund or proclitics on the finite verb. This latter configuration, called "clitic-climbing" (e.g., Myhill 1988), is a manifestation of greater fusion between the emergent auxiliary and the gerund: in (32), proclitic *los* indicates that *fueron conservando* is a unit, just like single-word *conservan*.⁷

- (32) otros, que tuvieron principios grandes, y LOS FUERON + CONSERVANDO y los conservan y mantienen en el ser que comenzaron; (XVII, Quijote II, VI)
 'others had noble beginnings, and PRESERVED [lit: went preserving] them, and still preserve and maintain them just as they were' (Grossman, 494) [lit: went preserving, i.e., continued (went on) preserving them]

6. In counting *estar* + *V-ndo* tokens, we included cases of intervening adjectives (N=35) (but not *estarse quedo* + *V-ndo* in the Quijote, N=7); though it could be argued that *estar* + Adjective + *V-ndo* is a different construction, it does not exclude progressive meaning and thus is associated with the more general *estar* + *V-ndo* construction.

7. Excluded from the count were cases of structurally ambiguous reflexive marking, which may have contributed to the increase of clitic climbing over time (Torres Cacoullós 2000: 50–51).

Table 4 shows a diachronic increase in adjacency, association, and fusion for *estar* + *V-ndo*. The proportion of occurrences without intervening material increases significantly between all the data sets (from 36% in the 13th, to 50% in the 15th, 67% in the 17th, and 78% in the 19th century). The proportion of occurrences with a single as opposed to multiple gerunds increases from 80% in the 13th to 92% in the 19th c. data. And the rate of "clitic climbing" shows an increase between the combined 13th and 15th c. data, at 57%, and the combined 17th and 19th c. data, at 76% (we attribute the later drop in rate to the development of stylistic meaning in clitic climbing in the 19th c. (Torres Cacoullós 1999)).

Table 4. Grammaticization (unithood) measures for *estar* + *V-ndo*: Adjacency (lack of intervening material), Association (absence of multiple gerunds), Fusion ("clitic climbing")

	XIII	XV	XVII	XIX
Adjacency	36% (37/104)	50% (67/134)	67% (145/217)	78% (169/217)
Association	80% (83/104)	86% (115/134)	88% (192/217)	92% (199/217)
Fusion	63% (15/24)	50% (11/22)	82% (61/74)	70% (54/77)

Adjacency: XIII vs. XV Chi-Square 4.950998521; p = 0.0261; XV vs. XVII Chi-Square 9.799123895; p = 0.0017; XVII vs. XIX Chi-Square 6.634394904; p = 0.0100. Association: XIII vs. XIX: Chi-Square 9.323668501; p = 0.0023. Fusion: Combined XIII-XV vs. XVII-XIX: 57% (26/46) vs. 76% (115/151) Chi-Square 6.682716664; p = 0.0097

Based on these three unithood indices, we constructed a cumulative "grammaticization index", weighted to take account of adjacency more than association and fusion, as follows:

Adjacency: two points for no intervening material, one for an intervening subject, object, temporal or manner expression, zero for an intervening adjective, locative or more than one of the above.

Association: one point for a single as opposed to multiple gerunds.

Fusion: one point for a proclitic as opposed to enclitic. Since clitic climbing does not apply to all tokens, the index is calculated as a fraction.

Table 5 shows a diachronic increase in the value of this index for *estar* + *V-ndo*.

Table 5. Cumulative grammaticization (unithood) index for *estar* + *V-ndo*

	XIII	XV	XVII	XIX
	.60 (62.2/104)	.74 (99.33/134)	.79 (172.5/217)	.83 (180.75/217)

*Between parentheses is the point total for all tokens divided by the number of tokens

A final measure of the advancing grammaticization of *estar* + *V-ndo* is relative frequency. Table 6 shows the changing relative frequency of the locative-motion verbs in gerund periphrases. From having half the relative frequency of

ir in the 13th c. data, *estar* goes on to become the most frequent in the cohort of emerging auxiliaries; *ir* + *V-ndo* remains viable, but its frequency relative to *estar* decreases; and *andar* + *V-ndo* ceases to be productive (in these Peninsular Spanish data).

Table 6. Relative frequency of gerund (*V-ndo*) periphrases

	XIII (N=477)	XV (N=301)	XVII (N=505)	XIX (N=557)	
<i>estar</i> 'be'	26%*	45%	43%	39%	Rises
<i>ir</i> 'go'	50%	27%	37%	35%	Decreases
<i>andar</i> 'go around'	21%	16%	12%	3%	Declines
<i>venir</i> 'come'	3%	9%	4%	3%	Always minor
<i>seguir</i> 'follow'	1	1	0	15%	Appears late
<i>quedar</i> 'remain'	0	3%	4%	2%	Always minor
<i>continuar</i> 'continue'	0	0	0	3%	Appears late

The relative frequency of *estar* is greater in the XV c. than in the XIII c. (Chi-square 29.99123288; $p = 0.0000$); differences in the relative frequency of *estar* between the XV c., XVII c., XIX c. are not significant.

*13th c. *estar* count includes 18 tokens of *seer* + *V-ndo*.

In summary, *estar* + *V-ndo* shows bleaching of locative meaning (Table 3), an increasing unithood index (Tables 4, 5), and increasing relative frequency (Table 6). In the next section we examine the role prefabs have played in the grammaticization of *estar* + *V-ndo*.

4.3 Prefabs and *estar* + *V-ndo* grammaticization

In identifying prefabs, we consider relative frequency rather than token frequency, both with respect to the "auxiliary" and the gerund (cf. Torres Cacoullos 2000: 57–59, 2006; Hay 2001). We operationally define prefabs as "auxiliary"-plus-gerund combinations making up 2% or more of the corresponding "auxiliary" data and 50% or more of the corresponding gerund data. For example, *estar hablando* 'be talking' makes up 5% (32/672) of *estar* data and 71% (32/45) of *hablando* data. Combining the data of all time periods, we identified the prefabs appearing in Table 7 (listed alphabetically, by "auxiliary").⁸

8. *Estar hablando* total includes four 13th c. tokens of *seer hablando*. High frequency *diciendo* 'saying, telling' (N=50), which makes up 3% (22/672) of the *estar* and 2% (15/700) of the *ir* data, is not overwhelmingly associated with either auxiliary (44% (22/50) *estar*, 30% (15/50) *ir*).

Table 7. Prefabs (as percentage of aux and of gerund; all time periods combined)

		% "auxiliary"	% gerund	
ESTAR	<i>aguardando</i> 'waiting'	2% (14/672)	93% (14/15)	
	<i>diciendo</i> 'saying'	3% (22/672)	44% (22/50)	
	<i>durmiendo</i> 'sleeping'	2% (14/672)	93% (14/15)	
	<i>escuchando</i> 'listening'	3% (23/672)	96% (23/24)	
	<i>esperando</i> 'waiting'	7% (48/672)	89% (48/54)	
	<i>hablando</i> 'talking'	5% (32/672)	71% (32/45)	
	<i>mirando</i> 'looking'	7% (49/672)	84% (49/58)	
	<i>oyendo</i> 'hearing'	2% (15/672)	94% (15/16)	
	<i>pensando</i> 'thinking'	2% (13/672)	62% (13/21)	
	IR	<i>creciendo</i> 'growing'	3% (24/700)	86% (24/28)
<i>diciendo</i> 'saying'		2% (15/700)	30% (15/50)	
<i>entrando</i> 'entering'		3% (22/700)	100% (22/22)	
<i>haciéndose</i> 'becoming'		3% (24/700)	100% (24/24)	
<i>huyendo</i> 'fleeing'		3% (22/700)	67% (22/33)	
<i>yendo</i> 'going'		2% (14/700)	88% (14/16)	
<i>llegando</i> 'approaching, arriving'		3% (20/700)	95% (20/21)	
<i> viniendo</i> 'coming'		2% (12/700)	92% (12/13)	
ANDAR		<i>buscando</i> 'looking for'	25% (57/229)	84% (57/68)
VENIR		<i>huyendo</i> 'fleeing'	13% (10/77)	
SEGUIR	<i>andando</i> 'walking'	5% (4/85)		
	<i>creciendo</i> 'growing'	4% (3/85)		
	<i>siendo</i> 'being'	5% (4/85)		
QUEDAR	<i>esperando</i> 'waiting'	12% (5/43)		

We will make a case that (1) prefabs are in the advance of the general construction in unithood status in early stages and thus demote the independent lexical status of the emerging auxiliary, and (2) in their association with semantic classes of which they are the most frequent member, prefabs promote the productivity of the general construction.

4.3.1 Prefabs lead as units

The first column of Table 8 shows *estar* + *V-ndo* prefabs by time period. Two prefabs in particular, *estar hablando* 'be talking' and *estar esperando* 'be waiting', are evident throughout the time periods examined and continue in present-day data. *Estar hablando* is the single most frequent *estar* + *V-ndo* collocation (165/2270) in conversational Peninsular Spanish data (COREC, Marcos Marín 1992) and *estar esperando* (38/2270) is still among the top ten collocations.⁹

9. *Haciendo* 'doing' is more frequent than *hablando* in the COREC data (N = 216), but it combines with (often non-referential) objects to form different predicates, thus we don't view it as a single collocation like *estar hablando*.

Table 8. *Estar + V-ndo* prefabs, by time period: Comparison of grammaticization indices

		Unithood index		Locative	
		prefab	general <i>estar + V-ndo</i> (Table 5)	prefab	general (Table 3)
XIII	<i>hablando</i>	.67 (8/12)	.60 (62.2/104)	17%	36%
	<i>esperando</i>	.82 (4.9/6)		33%	
XV	<i>hablando</i>	.89 (8/9)	.74 (99.33/134)	11%	40%
	<i>esperando</i>	.74 (13.3/18)		56%	
XVII	<i>hablando</i>	.72 (4.33/6)	.79 (172.5/217)	33%	24%
	<i>diciendo</i>	.90 (9/10)		0	
	<i>mirando</i>	.79 (29.83/38)		5%	
	<i>escuchando</i>	.79 (14.25/18)		6%	
	<i>esperando</i>	.78 (16.33/21)		38%	
	Avg	.79 (73.74/93)			
XIX	<i>hablando</i>	.93 (4.67/5)	.83 (180.75/217)	20%	16%
	<i>diciendo</i>	1.00 (4/4)		25%	
	<i>mirando</i>	.73 (4.42/6)		0	
	<i>oyendo</i>	.76 (6.83/9)		22%	
	<i>pensando</i>	.96 (8.67/9)		0	
	<i>esperando</i>	.75 (2.25/3)		67%	
	Avg	.86 (30.84/36)			

As with English *can*, gerund construction prefabs may begin as harmonious expressions, where the original lexical meaning of the emerging auxiliary is compatible with the main verb. That is, as Torres Cacoullós (2000: 175) has argued, frequent collocations such as *estar hablando* ‘be talking’, *ir creciendo* ‘be (go) growing’ and *andar buscando* ‘be (go around) looking for’ (Table 7) “follow from the original uses of the source constructions”. Such harmonious prefabs may appear conservative in manifesting retention of meaning from the source construction, for example, a locative meaning component in *estar esperando* ‘be waiting’, as in (33): over one-third of present-day oral Peninsular Spanish (COREC) tokens (34%, 13/38) have a co-occurring locative, whereas the rate of co-occurring locatives with *estar hablando* is 5% (9/165) in the same corpus. Nevertheless, retention of original meaning in the unit, originally a harmonic combination, does not detract from grammaticization. On the contrary, since the locative meaning is contributed by *esperar* ‘wait’, the meaning contribution of *estar* is minimized.

- (33) y él nos ESTABA ESPERANDO en San Sebastián (COREC, CCON035B)
‘and he WAS WAITING for us in San Sebastián’

Other prefabs may conventionalize as fixed discourse formulas. For example, *estoy hablando de* ‘I’m talking about’ or *estamos hablando de* ‘we’re talking about’,

as in (34), (see also [29]) may play more of an interactional role akin to discourse markers or connectives rather than actually referring to a situation in progress. Some scholars call such developments “pragmaticalization” (e.g., Erman & Kotsinas 1993; cf. Aijmer 1997: 3). As with prefabs manifesting meaning retention, prefabs with formulaic discourse uses detract from the independence and meaning contribution of the erstwhile lexical item (locative or motion verb).

- (34) ESTAMOS HABLANDO de la madre no del matrimonio. (COREC, PEDU010A)
‘WE’RE TALKING about the mother not the couple’

Both these prefabs show an early lead in their unithood index. The columns in Table 8 compares the unithood indices and rate of co-occurring locatives for the prefabs and the general construction (all tokens of *estar + V-ndo*). *Estar hablando* leads the grammaticization of *estar + V-ndo* in the earliest (Old Spanish) stage, with a unithood index of .67 compared to .60 for the general construction, in the 13th c., and .89 compared to .74, in the 15th c. data. The rate of cooccurring locatives is also lower with *estar hablando*, at 17% and 11%, compared to 36% and 40%, in the 13th and 15th c. data, respectively. *Estar esperando* also shows a higher than average unithood index, in the 13th c. data, though not a lower locative rate. Over time, as the productivity of the general construction increases, *estar hablando* makes up a smaller portion of the data, from 12% (12/104) of all *estar + V-ndo* tokens in the 13th c. to 2% (5/217) in the 19th c., and appears to follow general patterns.

Thus, in early stages, prefabs score higher than the general *estar + V-ndo* construction on the unithood measures shown above (Section 4.2). This empirical result provides evidence that frequent collocations become automated as single processing units (Bybee 2003). As we argued earlier, prefabs contribute to grammaticization because they are accessed holistically, which means that the erstwhile independent lexical item contributes less meaning, which promotes the semantic bleaching of the emerging auxiliary in this construction. Thus, it is the unithood of prefabs, meaning retention or formulaic discourse uses notwithstanding, that is conducive to grammaticization.

Now, given the relative autonomy of high frequency collocations (Bybee 2003), how do these prefabs contribute to the productivity of a general grammatical construction? Our argument is that prefabs maintain associations with the more general construction. In the next section we will show that prefabs contribute to productivity via the semantic classes centered around them.

4.3.2 Prefabs contribute to productivity via associated semantic classes

Estar + V-ndo prefabs are *estar hablando* and *estar esperando* in the Old Spanish data, as we have seen; these plus *estar diciendo* ‘be saying, telling’, *estar mirando* ‘be watching’, *estar escuchando* ‘be listening’ in the 17th c. data; and all of the above plus *estar pensando* ‘be thinking’ and substituting *oyendo* ‘hearing’ for *escuchando* ‘listening’ in

the 19th c. data. From Table 8 we can deduce the proportion of the data made up by the prefabs by adding their tokens (the second number between parentheses in the first column) and taking this sum over the total number of tokens per time period (the second number between parentheses in the second column). This proportion seems to remain steady over time at 17% (18/104) in the 13th c. data, 20% (27/134) in the 15th, 43% (93/217) in the 17th (38 tokens of *estar mirando* 'staring or gazing' in the *Quijote* contribute to this inflated figure), and 17% (36/217) in the 19th c. data. Nevertheless, considering we have listed six instead of two prefabs in the 19th c., it is fair to conclude that while there is continuity of particular prefabs over time, these make up a declining proportion of the general construction data. This is as expected, since grammaticization involves generalization to more and more types.

The prefabs participate in classes with other semantically related verbs, ranging from the large class of verbs of speech (e.g., *alabar* 'praise', *demandar* 'request', *explicar* 'explain', *gritar* 'shout', *murmurar* 'murmur', *razonar* 'argue', *rogar* 'beg, pray') to the small class of verbs of 'waiting'. Intuitively apparent semantic classes for the prefabs identified in Table 8 are shown in (35). Besides noting the verbs of speech and 'waiting' verbs, we coded all tokens for affiliation with verbs of perception, bodily activity (e.g., *bañarse* 'bathe', *doler* 'ache', *llorar* 'weep', *respirar* 'breathe', *sangrar* 'bleed', *temblar* 'tremble'), and cognition-emotion (e.g., *figurar* 'imagine, think', *morirse de miedo* 'be scared to death', *penar* 'suffer', *rumiar* 'ruminate', *sentir* 'feel', *temer* 'fear'). Table 9 shows the distribution of *estar + V-ndo* tokens in semantic classes, by time period. The distribution and concentration of tokens in the semantic classes we defined appears steady (we will return shortly to the decline of the 'waiting' class).

(35) *Estar*: Prefabs (Table 8) and semantic classes

<i>hablando, diciendo</i>	SPEECH class size:	big
<i>pensando</i>	COGNITION (also emotion)	big
<i>durmiendo</i>	BODY ACTIVITY	medium sized
<i>escuchando, mirando, oyendo</i>	PERCEPTION	small
<i>esperando, aguardando</i>	WAITING	very small

Table 9. Semantic classes: *Estar + V-ndo*

	XIII (N=104)	XV (N=134)	XVII (N=217)	XIX (N=217)	
Speech	16% (17)	18% (24)	13% (28)	15% (32)	steady
Cognition	11% (11)	7% (9)	4% (9)	12% (27)	steady
Body activity	13% (13)	4% (5)	8% (17)	11% (23)	steady
Perception	12% (12)	7% (10)	28% (60)*	10% (22)	steady
Waiting	11% (11)	15% (20)	13% (29)	3% (6)	decline
Other	38% (40)	49% (66)	34% (74)	49% (107)	steady

*In *Quijote*, *estar escuchando* N=18, *estar mirando* N=38

** Difference proportion "Other" XIII-XV combined 45% (106/132) vs. XVII-XIX combined 42% (181/253) is not significant.

Though contrary perhaps to our expectations we do not see an increase in the "Other" category, that is, an expansion outside the original semantic classes over time, generalization of *estar + V-ndo* is shown in a count of type/token ratios, where "types" are the different verbs appearing in the open slot in the construction. Table 10 shows type/token ratios, calculated for each data set based on a random sample of 100 tokens (since an increased sample size is likely to show a lower type/token ratio, as lexical types are repeated). The ratio increases from 48–49 in the 13th and 15th c. data, to 55 in the 17th and 69 in the 19th c. The increase in type/token ratio over time indicates the increased productivity of the construction but the lack of increase in the "other" category in Table 9 indicates that much of the generalization is taking place within the established verb classes.

Table 10. Type/token ratio: *Estar + V-ndo* (randomized sample 100)

XIII	XV	XVII	XIX
49/100	48/100	55/100	69/100

Even though *estar hablando* and *estar esperando* are both high frequency prefabs, an important difference is precisely that the former is part of the large class of verbs of speech appearing in the *estar + V-ndo* configuration, while the class of verbs of 'waiting' is tiny, including only *aguardar* and *atender* besides *esperar* ([35]). Since *estar hablando* is associated with a high type frequency semantic class, the contribution of this prefab to the development of a general *estar + V-ndo* construction should be greater than that of *estar esperando*. Besides the striking decline in the relative frequency of 'waiting' verbs (shown in Table 9), two pieces of evidence show the weaker contribution of *estar esperando* to the grammaticization of the general construction. First, recall that *estar esperando* has had a higher than average rate of co-occurring locatives from the 15th c. data onwards (Table 8). Second, though *esperando* is still among the top ten or so gerunds combining with *estar* (in the present-day COREC data), its exclusive association has eroded. While in the 13th and 15th c. data, 100% (24/24) of *esperando* tokens co-occurred with *estar* as opposed to another "auxiliary", beginning with the 17th c. data, *quedar* 'remain' combines with this gerund, so that *quedar esperando* is somewhat of a prefab (by our operational definition) in its own right, making up 12% (5/43) of all *quedar + V-ndo* tokens (Table 7). In contrast, no other "auxiliary" competes with *estar*'s association with *hablando*. So as predicted, the contribution of *estar esperando* and its low type frequency class to the grammaticization of a general *estar + V-ndo* construction is less consistent than that of *estar hablando*.

The conclusion of this section is that prefabs may participate in classes with other semantically related verbs and that these classes may be higher or lower type frequency categories. We reason that participation in high type frequency categories, as in the case of *estar hablando*, contributes to a more general schema and thus greater productivity (Bybee & Eddington 2006; cf. Torres Cacoullous 2000: 13, 130). In contrast, if the prefab cannot be associated with a many-membered class, as is the case with *estar esperando*, it will not contribute as consistently to the productivity of the grammaticizing construction.

4.3.3 Prefabs and productivity: Evidence from motion-verb gerund periphrases

Further support for the hypothesis that prefab exemplars of a grammaticizing construction must be associated with semantically related instances in order to contribute to the productivity of the construction is provided by *ir + V-ndo* and *andar + V-ndo* distributions.

Ir + V-ndo has developed a meaning of 'gradually developing' or prospective imperfective aspect (cf., e.g., Dietrich 1983; Olbertz 1998; Squartini 1998). The data suggest that this more general construction emerges from more particular *ir + V-ndo* constructions, including a harmonic motion construction and a change-of-state construction (Torres Cacoullous 2000: 151). One set of *ir + V-ndo* prefabs in the early data is harmonic motion expressions with *yendo* 'going', *llegando* 'arriving, nearing', *viniendo* 'coming'; another prefab set is process verb expressions *ir creciendo* '(gradually) grow' and *ir haciéndose* '(gradually) become' (Table 7, above). The two corresponding semantic classes, motion verbs and process (change-of-state) verbs, which include many other members, have been the mainstay of the construction, making up between one-third and one-half of all the *ir + V-ndo* data in all time periods, as shown in Table 11. While the proportion of motion verbs has declined, as expected if the construction has grammaticized from a harmonic motion verb expression, process verbs appear to remain stable. A measure of the association of *ir + V-ndo* with processes is cooccurrence with reflexive (*se*)-marked lexical types, a number of which refer to changes of state, for example, *mudarse* 'change', *tornarse* 'become', and which pair up with *ir* as opposed to *estar* (though *estar + V-ndo* has generalized even to this context).¹⁰

10. The ratio of *ir + V_{REFLEXIVE} -ndo* to *estar V_{REFLEXIVE} -ndo* tokens shows a decline: 13th c. 48: 7 > 15th c. 10: 5 > 17th c. 23: 13 > 19th c. 38: 23.

Table 11. Semantic classes: *Ir + V-ndo*

	XIII (N=238)	XV (N=80)	XVII (N=188)	XIX (N=194)	
Motion	37% (88)	23% (18)	29% (55)	23% (44)	decline
Process	18% (43)	11% (9)	6% (12)	20% (38)	steady
All other verbs	45% (107)	66% (53)	64% (121)	58% (112)	

Though it starts out with more than double the relative frequency in 13th c. data, *ir + V-ndo* is not as productive as *estar + V-ndo*. Over time it is overtaken by *estar + V-ndo* in relative frequency (Table 6) and the pace of grammaticization has been slower for *ir + V-ndo*, as indicated in Table 12: we find no significant decrease in co-occurring locatives and two of the three unithood indices, association (single vs. multiple gerunds) and fusion (clitic climbing) fail to show an increase (adjacency, that is, lack of intervening material, does increase, from 58% (137/238) in the 13th c. data to 89% (172/194) in the 19th c. data (Chi-Square 50.749219454; $p = 0.0000$)). Furthermore, some of *ir + V-ndo*'s uses have been taken over by newcomer (in the 19th c.) *seguir* 'follow, continue' + *V-ndo*, at least in some varieties (Tables 6 and 7).

Table 12. Grammaticization indices *ir + V-ndo*: Co-occurring locatives and unithood measures

	XIII N = 238	XV N = 80	XVII N = 188	XIX N = 194
Locatives	25%	24%	19%	21%
Adjacency	58%	61%	86%	89%
Association	85%	84%	90%	90%
Fusion	95%	88%	95%	66%

How do we explain the restricted productivity of *ir + V-ndo* compared to *estar + V-ndo* despite an early lead in relative frequency (Table 6)? Contributing to grammaticization is the persistence of early prefabs (such as *ir creciendo* 'be [go] growing') and their association with high type frequency semantic classes participating in the construction (such as the process verb class). At the same time, however, from the beginning the construction has been heavily concentrated in a small number – only two – classes due to its more specific meaning, in contrast to the more general *estar + V-ndo*, which has been more evenly distributed across different semantic classes (Table 9).

The single-most remarkably robust prefab is *andar buscando* 'be [go] looking', which makes up an average of 20% of the tokens of the *andar* construction.

Andar buscando is so frequent that it has been said to be “really a set phrase” (Spaulding 1926: 259) or a case of “lexical specialization” (Squartini 1998: 261). Just like *estar hablando* and *ir creciendo*, *andar buscando* continues as a well-established routine in present-day varieties (with 9% (8/89) of all *andar + V-ndo* tokens in a corpus of popular Mexican Spanish (Torres Cacoullós 2000: 168)). But unlike *estar hablando* and *ir creciendo*, this prefab is not associated with a large class of semantically related items (though other lexical types in the Old Spanish data take on a ‘looking for’ meaning, for example *andar demandando* ‘enquiring’ and *andar catando manera* ‘looking for a way’ or *andar guisando cómo* ‘arranging how’ (Torres Cacoullós 2000: 164–165)). As we would predict, *andar + V-ndo* shows a sharp decline, dropping from 21% in the 13th c. to a relative frequency of 3% in the 19th c. data. Social factors are clearly important, since *andar + V-ndo* is much more frequent in other varieties, especially Mexican Spanish, where it has developed social associations (Torres Cacoullós 2001). Nevertheless, the restriction of *andar + V-ndo* compared to *ir + V-ndo* and especially *estar + V-ndo* is consonant with the notable strength – and isolation – of its prefab.

In summary, *ir + V-ndo* remains a viable aspectual expression, though largely concentrated in two semantic classes, while *andar + V-ndo* is geographically and socially restricted. Both the viability of *ir + V-ndo* and its slower grammaticization as well as the restriction of *andar + V-ndo* would be predicted by the view of prefabs and their associated semantic classes that we are advancing: early prefabs persist but contribute to productivity (generalization) of a grammaticizing construction only if they are associated with relatively large semantic classes of lexical types participating in the construction.

5. Conclusions

Our study, then, contributes to the understanding of the relation between the specific and the general in the development of constructions over time. We hope to have shown that prefabs are important to the understanding of the fabric of grammaticization. At any given point in time, prefabs will be responsible for increasing the frequency of grammaticizing constructions as well as for serving as the loci for extensions of the construction. Their lack of compositionality, their frequency and conventionalization play an important role in providing meaning for the construction as a whole while at the same time affecting the meaning of the constituent parts, usually by loss of earlier, lexical meanings. These interactions demonstrate that prefabs and their related constructions remain associated and interact in language change.

Specifically, we have demonstrated that prefabricated instances of constructions lead in the semantic reduction of the meaning of the construction as well as in manifesting structural indices of unithood. As predicted from their relative frequency of use, prefabs grammaticize earlier or at a faster rate than the general construction.

We have also presented evidence that prefabricated instances of constructions serve as the centers of subclasses of the grammaticizing construction, attracting more lexical types into the construction and thereby contributing to the productivity of the construction. This process is apparent in the Middle English verbs of communicating and cognition with *can*, as well as in the verbs of communicating with the *estar*- and in the process verbs with the *ir*- progressive construction in Spanish.

To a lesser extent, we find prefabs retaining the older meaning or distribution of the construction. As we mentioned, *estar esperando* ‘to be waiting’ still often co-occurs with locative expressions, suggesting that this exemplar implies ‘waiting somewhere’. We argue that it is not so much that *estar* has retained its locative meaning here as that the whole prefab has a locative implication that derives as much from *esperar* as from *estar*. Similarly in English, the expression *I kan nought sayn* from Middle English and its modern descendent, *I can’t say*, in some uses gives a knowledge interpretation: ‘I cannot say because I don’t know’. Again, the parts of the construction are harmonic in that saying itself implies knowledge to say. Thus it is the whole prefab that retains the earlier meaning, not just the auxiliary.

In other cases, an older distribution is maintained by a prefab, while the older meaning has eroded. Thus *andar buscando* ‘to be looking for’ is purely aspectual, but the use of *andar* with *buscar* reflects an older compatibility of the two lexical items. In the English examples, we have the continued use of *can* and *can’t* with main verbs such as *understand*, *remember*, *imagine*, *guess*, *believe*, where the modal contributes very little if any meaning. The use of *can* with these cognitive verbs is retained from the very earliest period when *cunnan* meaning ‘know’ was harmonic with these more specific verbs.

Our study has both diachronic and synchronic implications. To come back to the dimensions along the continuum between prefabs and more general constructions that we presented in the introduction, the data we have examined shows how essential it is that we consider prefabs to be highly integrated with the more general constructions.

Productivity: Even within a general construction, such as *can + VERB* or *estar + VERB -ndo*, there can be expressions with varying degrees of productivity: *can + cognitive verb* occurs with many different types, as does *estar + speaking verb*, while *estar* with *esperando* is quite isolated.

Transparency of meaning: *can't say* retains the knowledge interpretation while *can't understand* has no real semantic role derivable from *know* for *can*, yet these are clearly instances of the same construction. Also, *estar esperando* retains some locative nuance, while other instances of *estar* + gerund have lost all such meaning.

Analyzability: As the Spanish data show, the degree of analyzability can also vary, with more frequent collocations showing less analyzability as demonstrated by less frequent occurrence of modifiers and multiple gerunds and the more frequent occurrence of proclitics before the whole expression.

Thus it appears that grammaticization of a construction is not a uniform process with all instances or subclasses of the construction marching through the changes in lockstep. Rather, certain instances of the construction lead the charge, attracting other similar expressions, while low frequency uses may drag along at the rear. Some high-frequency instances may become fossilized early on, maintaining older meanings, while others rush ahead to become bleached and generalized. Our more general point, then, is that prefabs are not marginal or peripheral to grammar at all, but rather highly integrated with the more general structures of the language. Thus language use with its varying lexical specificity and uneven contours of token and type frequency is highly involved in the creation and maintenance of grammatical constructions.

Corpus [word counts-tokens]

- Calila (1250) = Anonymous. 1987. *Calila e Dimna*, ed J.M. Cacho Blecua and M.J. Lacarra. Madrid: Castalia. [86,000–30]
- GEI (1260–1280) = Alfonso X. 1930. *General estoria. Primera parte*, ed. Antonio G. Solalinde. Madrid: Centro de Estudios Históricos, 1930. [572,000–250]
- GEII (1260–1280) = Alfonso X. 1957. *General Estoria. Segunda parte*, 2 vols., ed. A. Solalinde, Ll. Kasten and V.R.B. Oelshläger. Madrid: CSIC (Consejo Superior de Investigaciones Científicas). [263,500–197]
- Grimalte (~1486) = de Flores, Juan. 1971. *Grimalte y Gradissa*, ed. Pamela Waley. London: Tamesis. [25,500–10]
- Cárcel (1492) = de San Pedro, Diego. 1972. *Cárcel de amor*, ed. Keith Whinnom. Madrid: Castalia. [25,500–5]
- CRC (1482–1490) = Hernando del Pulgar, *Crónica de los Reyes Católicos*, 2 vols., ed. Juan de Mata Carriazo, Madrid: Espasa Calpe, 1943. [322,000–155]
- Celestina (1499) = de Rojas, Fernando. 1987. *La Celestina*, ed. D.S. Severin, Madrid: Cátedra. [67,000–81]; Translated by Mack Hendricks Singleton, Madison: The University of Wisconsin Press, 1968.
- Lozana (1528) = Delicado, Francisco. 1984. *La lozana andaluza*, ed. Bruno M. Damiani. Madrid: Castalia. www.cervantesvirtual.com. [64,000–50]

- Quijote (1605–1616) = Miguel de Cervantes, *Don Quijote de la Mancha*. www.cervantesvirtual.com. [384,000–505]; Translated by Edith Grossman, *Don Quixote*, Harper Perennial, 2005.
- Pepita (1870) = Valera, Juan. *Pepita Jiménez*. www.cervantesvirtual.com. [56,500–63]
- Regenta (1870–1880) = Alas “Clarín”, Leopoldo. *La Regenta*, vol I. www.cervantesvirtual.com. [141,000–258]
- Perfecta (1876) = Pérez Galdós, Benito. *Doña Perfecta*. www.cervantesvirtual.com. [65,000–107]
- Pazos (1886) = Pardo Bazán, Emilia. *Los pazos de Ulloa*. www.cervantesvirtual.com. [83,500–129]

References

- Aijmer, Karin. 1997. *I think* – an English modal particle. In *Modality in Germanic languages: Historical and comparative perspectives*, T. Swan & O. Jansen Westvik (Eds), 1–47. Berlin: Mouton de Gruyter.
- Boas, Hans C. 2003. *A constructional approach to resultatives*. Stanford CA: CSLI.
- Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Bybee, Joan. 1988. Morphology as lexical organization. In *Theoretical morphology: Approaches in modern linguistics*, M. Hammond & M. Noonan (Eds), 119–141. New York NY: Academic Press.
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: CUP.
- Bybee, Joan. 2003. Mechanisms of change in grammaticization: The role of frequency. In *Handbook of historical linguistics*, B. Joseph & R. Janda (Eds), 602–623. Oxford: Blackwell.
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4): 529–551.
- Bybee, Joan & David Eddington. 2006. A usage-based approach to Spanish verbs of ‘becoming’. *Language* 82(2): 323–355.
- Bybee, Joan & William Pagliuca. 1987. The evolution of future meaning. In *Papers from the 7th International Conference on Historical Linguistics*, A.G. Ramat, O. Carruba & G. Bernini (Eds), 109–122. Amsterdam: John Benjamins.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago IL: The University of Chicago Press.
- Camus Bergareche, Bruno. 2004. Perífrasis verbales y expresión del aspecto en español. In *El pretérito imperfecto*, L. García Fernández & B. Camus Bergareche (Eds), 511–572. Madrid: Gredos.
- Carey, Kathleen. 1994. The grammaticalization of the Perfect in Old English: An account based on pragmatics and metaphor. In *Perspectives on grammaticalization*, W. Pagliuca (Ed.), 103–117. Amsterdam: John Benjamins.
- Casenhiser, Devin & Goldberg, Adele E. 2005. Fast mapping of a phrasal form and meaning. *Developmental Science* 8(6): 500–508.
- Company Company, Concepción. 2006. Zero in syntax, ten in pragmatics, or subjectification as syntactic cancellation. In *Subjectification: Various paths to subjectivity*, A. Athanasidou, C. Canakis & B. Cornillie (Eds), 375–398. Berlin: Mouton de Gruyter.

- Dąbrowska, Ewa & Elena Lieven. 2005. Towards a lexically-specific grammar of children's question constructions. *Cognitive Linguistics* 16(3): 437–474.
- Diessel, Holger & Michael Tomasello. 2001. The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics* 12(2): 97–141.
- Dietrich, Wolf. 1983. *El aspecto verbal perifrástico en las lenguas románicas*. (Translated by Marcos Martínez Hernández.) Madrid: Gredos.
- Erman, Britt & Ulla-Britt Kotsinas. 1993. Pragmaticalization: The case of *ba'* and *you know*. *Studier i modern sprakvetenskap* 10: 76–92.
- Erman, Britt & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20: 29–62.
- García Fernández, Luis (dir.), Ángeles Carrasco Gutiérrez, Bruno Camus Bergareche, María Martínez-Atienza & María Ángeles García García-Serrano. 2006. *Diccionario de perifrasis verbales*. Madrid: Gredos.
- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago IL: The University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: OUP.
- Goldberg, Adele E., Devin Casenhier & Nitya Sethuraman. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 15: 289–316.
- Goldberg, Adele E., Devin Casenhier & Nitya Sethuraman. 2005. The role of prediction in construction learning. *Journal of Child Language* 32(2): 407–426.
- Goossens, Louis. 1992. *Cunnan, conne(n), can*: The development of a radial category. In *Diachrony within synchrony: Language history and cognition*, G. Kellerman & M.D. Morrissey (Eds), 377–394. Frankfurt: Peter Lang.
- Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39: 1041–1070.
- Hopper, Paul J. 1991. On some principles of grammaticalization. In *Approaches to grammaticalization*, Vol.1, E. Closs Traugott & B. Heine (Eds), 17–35. Amsterdam: John Benjamins.
- Hopper, Paul J. & Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: CUP.
- Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: OUP.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar*, Vol.1: *Theoretical perspectives*. Stanford CA: Stanford University Press.
- Lieven, Elena, Julian Pine & Gillian Baldwin. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language* 24(1): 187–219.
- Lieven, Elena, Heike Behrens, Jennifer Speares & Michael Tomasello. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30(2): 333–370.
- Marcos Marín, Francisco (dir.). 1992. COREC: Corpus de Referencia de la Lengua Española Contemporánea: Corpus Oral Peninsular. (www.lllf.uam.es/~fmarcos/informes/corpus/corpusix.html)
- Myhill, John. 1988. The grammaticalization of auxiliaries: Spanish clitic climbing. *Berkeley Linguistics Society* 14: 352–63.
- Nunberg, Geoffrey, Ivan A. Sag & Thomas Wasow. 1994. Idioms. *Language* 70: 491–538.
- Olbertz, Hella. 1998. *Verbal periphrases in a functional grammar of Spanish*. Berlin: Mouton de Gruyter.
- Scheibman, Joanne. 2000. *I dunno but ... a usage-based account of the phonological reduction of don't in conversation*. *Journal of Pragmatics* 32: 105–24.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: OUP.
- Spaulding, Robert K. 1926. History and syntax of the progressive constructions in Spanish. *University of California Publications in Modern Philology* 13: 229–284.
- Squartini, Mario. 1998. *Verbal periphrases in Romance: Aspect, actionality, and grammaticalization*. Berlin: Mouton de Gruyter.
- Tatlock, John S.P. & Kennedy, Arthur G. 1927. *A concordance to the complete works of Geoffrey Chaucer*. Washington DC: Carnegie Institute.
- Thompson, Sandra A. 2002. 'Object complements' and conversation. *Studies in Language* 26(1): 125–164.
- Thompson, Sandra A. & Anthony Mulac. 1991. A quantitative perspective on the grammaticalization of epistemic parentheticals in English. In *Approaches to grammaticalization, Vol. II*, E.C. Traugott & B. Heine (Eds), 313–329. Amsterdam: John Benjamins.
- Torres Cacoulios, Rena. 1999. Construction frequency and reductive change: Diachronic and register variation in Spanish clitic climbing. *Language Variation and Change* 11: 143–170.
- Torres Cacoulios, Rena. 2000. *Grammaticalization, synchronic variation, and language contact. A study of Spanish progressive -ndo constructions*. Amsterdam: John Benjamins.
- Torres Cacoulios, Rena. 2001. From lexical to grammatical to social meaning. *Language in Society* 30: 443–478.
- Torres Cacoulios, Rena. 2006. Relative frequency in the grammaticalization of collocations: Nominal to concessive *a pesar de*. In *Selected proceedings of the 8th Hispanic Linguistics Symposium*, T.L. Face & C.A. Klee (Eds), 37–49. Somerville MA: Cascadilla Proceedings Project. (<http://www.lingref.com/cpp/hls/8/index.html>)
- Torres Cacoulios, Rena & James A. Walker. 2009. On the persistence of grammar in discourse formulas: A variationist study of *that*. *Linguistics* 47.1
- Traugott, Elizabeth Closs. In press. Grammaticalization, constructions and the incremental development of language: Suggestions from the development of degree modifiers in English. In *Language evolution: Cognitive and cultural factors*, R. Eckardt & G. Jaeger (Eds), Berlin: Mouton De Gruyter.