INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700  800/521-0600
RICE UNIVERSITY

Cyclical Grammaticalization
and the Cognitive Link Between Pronoun and Copula

by

Aya Katz

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE

James E. Copeland, Chair
Professor of Linguistics and German

Stephen A. Tyler
Professor of Anthropology and Linguistics

Spike Gidea
Assistant Professor of Linguistics

E. Douglas Mitchell
Adjunct Professor of Linguistics

Lilly C. H. Chen
Lecturer on Linguistics

Houston, Texas

May, 1996
ABSTRACT

Cyclical Grammaticalization
And the Cognitive Link Between Pronoun and Copula

by

Aya Katz

The process of grammaticalization is the transition from a less 'grammatical' state to one that is more so. But the results of the process, if we follow the history of any given linguistic unit as it undergoes grammaticalization over and over again, may involve achieving a state similar to one that was in effect at an earlier stage of its history.

The phenomenon explored in this dissertation is the grammaticalization of pronouns into copulas and copulas into pronouns.

A crosslinguistic examination of copulas in ten languages, Chinese, English, Finnish, French, Hebrew, Hungarian, Korean, Russian, Turkish, and Vietnamese, reveals the following functional tendencies:

(1) one morpheme is used for possession and existence,
    and
(2) another is used for identity and class membership.

It is from these equative copulas in (2) that third person pronouns are developed, and it is from third person subject pronouns that we get newly formed copulas.
The following instances of grammaticalization of pronoun and copula are demonstrated here:

(1) Chinese -- pronoun to copula
(2) Hebrew (Biblical to Modern) -- pronoun to copula
(3) Finnish -- pronoun to copula
(4) Turkish -- copula to pronoun
(5) Hebrew (pre-proto-Seriitic to Biblical) -- copula to pronoun

The Hebrew example, combining (2) and (5) above, provides us with a full cycle, from copula to pronoun to copula. At each step, the unit becomes more bleached and conventionalized. Starting from a form of a stative verb with marking for conjugation, gender, number and person, it becomes a demonstrative pronoun, marked only for person, gender and number, and then is transformed once more to a copula, marked only for number and gender, but not person. Pronouns and copulas are equally abstract, but at each stage of the change from one to the other we have seen an example of generalization and conventionalization.

While grammaticalization progresses along a unidirectional cline from concrete to abstract, the history of a particular linguistic unit may reveal that it has travelled the same semantic path more than once.
ACKNOWLEDGEMENTS

I would like to thank the members of my thesis committee for their invaluable advice and guidance. My principal advisor, Jim Copeland, introduced me to the concept of grammaticalization and spent countless hours in detailed discussions of unidirectionality. His contributions are reflected in every portion of this work, as well as in my linguistic training as a whole.

Spike Gikdea has been a great help. His enthusiastic interest in my topic has spurred me on, and his criticism has helped to sharpen my focus and improve my prose. His observations on Panare have likewise enriched my outlook on pronouns and copulas.

Douglas Mitchell has supported my outlandish claims from the start and has read my drafts with great alacrity. In addition to introducing me to Gothic and Sanskrit, he has been very supportive of my investigations into Indo-European, even when they proved inconclusive. And he has been most tolerant of my stylistic choices and my sometimes less than scientific vocabulary.

Stephen Tyler’s contributions are unutterably salicnt to my project. In addition to his tutelage in cognitive linguistics and Indian culture, I owe him a great debt in the matter of Cabeza de Vaca. His support never flags.

Lilly Chen provided me with the Chinese example, and without her help Chapter Three could never have been written. In addition, she has shared with me her work on verbal expansion in Chinese with its fresh outlook on the motivation behind the grammaticalization process. She has shown me that language contact is not necessary to bring about linguistic change, and that synthetic reanalysis need not come about as a reaction to loss. It may be that the human desire to add more and more verbiage is a contributing factor to grammatical attrition, since the practical limitations on the size of an utterance are a crosslinguistically stable constant. She has taken the
trouble to point out once and for all that infinitely long sentences have never occurred in natural language nor ever will.

I would like to express my deep appreciation to Elizabeth Traugott for taking the time to discuss unidirectionality and cyclicity with me at the 1996 SECOL conference, and for her subsequent correspondence and the articles to which she referred me. It was heartening to hear from a pioneer in the field of grammaticalization and a researcher who continues to expand the horizons of our discipline that unidirectionality has not been borne out as an essential attribute of the grammaticalization process.

Svi Rin’s expertise on Semitic languages has been of great help, and his prompt responses to my queries have been invaluable to this project.

I thank Shin Ja Hwang, Insun Yang, MiHyun Baek and Jeong Hwa Lee for their help with Korean grammar and Linh-Chan Brown for her assistance with Vietnamese. I would also like to express my gratitude to Manuel Datuin, my consultant for Pangasinan, and Efe Ustundag, my Turkish consultant.

I acknowledge a heartfelt debt to the late Albina Nikolaevna Hill, who instructed me in conversational Russian and also provided a rich grounding in sociolinguistic variation in Russian usage.

Finally, my colleague Tim Pulju proofread the entire manuscript, and in addition to suggesting novel measures for the retrieval of spilt milk, corrected numerous typographic, grammatical and stylistic errors.

The contributions of the above named scholars have been instrumental in bringing forth the best that this work has to offer. Any errors, infelicities, fallacies or dubious claims remaining are my own.
# TABLE OF CONTENTS

Abstract ................................................................. ii
Acknowledgements ....................................................... iv
Contents ................................................................. vi
The Circle of Grams ................................................... x
List of Abbreviations .................................................. xi

Chapter 1. Introduction with Reference to General Theories of
Grammaticalization, Typology and Language Change

§1.1 Definition ......................................................... 1
§1.2 General Assumptions about Grammaticalization ............... 17
  §1.2.1 Unidirectionality and Irreversibility ..................... 17
  §1.2.2 Cyclicality and Counter-Examples to Unidirectionality ... 22
  §1.2.3 Isolating Starting Point and Degrees of Grammaticalization 25
§1.3 Conclusion ....................................................... 26

Chapter 2. The Cognitive Function of Copula and Pronoun

§2.1. Introduction .................................................... 29
§2.2 Introduction to the Copula .................................... 31
  §2.2.1 The Relationship Between Being and Having ............. 35
  §2.2.2 Semantic Sources for the Copula: From the Concrete ..... 44
  §2.2.3 Semantic Extensions of the Copula: ... to the Abstract .. 45
  §2.2.4 Conclusions on Copular Morphology ..................... 46
  §2.2.5 The Cognitive Framework of the Copula ................. 47
§2.3 Pronouns and Demonstratives: The Pronominal System ....... 51
§2.4 Distinguishing Pronouns from Copulas ....................... 60
§2.5 Conclusion ..................................................... 61

Chapter 3. Classical Demonstrative to Modern Copula: The story of
Chinese shi.

§3.1 Introduction .................................................... 63
  §3.1.1 The Guises of shi ......................................... 64
§3.2 Classical Chinese to Modern Mandarin: General Trends ...... 65
Chapter 4. Classical Pronoun to Modern Copula: Hebrew נֹּֽעַ [hú].
§4.1 Introduction ........................................... 85
§4.2 Functional Marking
  §4.2.1 Synchronic Co-Existence of Copular and Pronominal Uses . . 85
  §4.2.2 Agreement for Gender and Number -- But Not Person .... 89
  §4.2.3 Agreement for Definiteness .......................... 91
§4.3 Review of the Literature ................................ 93
§4.4 Biblical Sources for the Modern Day Construction .......... 97
§4.5 Typological Environment ................................. 99
§4.6 Possible Source Constructions ............................ 100
§4.7 Conclusions ............................................ 102

Chapter 5. The cycles of Finno-Ugric on
§5.1 Preview .................................................. 103
§5.2 A Sidetrip into Russian ................................ 103
§5.3 The Uses of Finnish on ................................. 105
§5.4 The Hungarian Reflex of *on ............................ 109
§5.5 Mechanisms of Reanalysis ............................... 115
§5.6 Conclusions: Typological Dating and Grammaticalization .... 117

Chapter 6. The Turkish Third Person Pronoun
§6.1 Preview .................................................. 118
§6.2 The Turkish Third Person Singular Pronoun ............... 118
  §6.2.1 Historical Evidence for the Use of the Pronoun as a Copula . 119
§6.3 The Spurious Nasal in the Paradigm ....................... 120
§6.4 Zero Morphemes ......................................... 122
Chapter 7 Biblical Hebrew Third Person Pronouns and Verb 'to be' . 137

§7.1 The Hebrew Third Person Pronouns and the Verb 'to be' are Related . 139

§7.1.1 Internal Reconstruction . 139

§7.1.1.1 In Hebrew, Roots Consist of Consonants . 139

§7.1.1.2 The Root of the Verb 'to be' Has Two Alternates: \\h-j\\ and \\h-w\\ . 143

§7.1.1.3 The Root of the Third Person Pronoun Has Two Alternates: \\h-j\\ and \\h-w\\ . 146

§7.1.1.4 The Orthography of the Feminine and Masculine Pronouns is Identical in Some Parts of the Bible . 147

§7.1.1.5 Some Alternate Orthographic Renditions of the Verb 'to be'. 147

§7.1.1.6 Biblical Hebrew Pure Identity Statements with Pronoun as Linker . 148

7.1.1.7 In Both Form and Meaning the Verb 'to be' and the Third Person Singular Pronouns of Hebrew are Related . 150

§7.1.2 Functional Development . 151

§7.1.2.1 Unmarked Time/Aspect Was Present/Continuing . 158

§7.1.2.2 The Default Verb Was the Copula . 161

§7.1.2.3 From Suffixed Verbal Morphology, Third Person Was Once the Default Person . 162

§7.1.2.4 The Need for a Third Person Developed in Unmarked Aspect . 163

§7.2 Comparative Reconstruction . 165

§7.3 Conclusion . 169

Chapter 8. Typological and Theoretical Implications of the Findings

§8.1 Introduction . 171
§8.2 Typological Survey ........................................ 171
§8.2.1 Copula to Pronoun: Turkish and Biblical Hebrew .... 173
§8.2.2 Pronoun to Copula: Chinese, Finnish and Modern Hebrew . 177
§8.2.3 Typological Conclusions ............................... 181
§8.3 Theoretical Conclusions ..................................... 181
§8.3.1 Unidirectionality and Hierarchies ...................... 182
§8.3.2 Unidirectionality and Cyclicity ......................... 185
§8.4 Conclusion ..................................................... 185

Chapter 9. Conclusions ........................................... 187

Appendix--Index of Cited Authors ............................. 193

Bibliography ....................................................... 194
The Circle of Grams

Based on The Circle of Life by Elton John and Tim Rice

From the first time they ever are uttered,
Before they’ve been bleached and reduced,
There’s more they can mean than we mean them to mean,
More usage than ever gets used.

Some say human tongues are evolving,
And some claim they grow worse and worse,
But most will concede that speakers fill needs
By converting words while they converse.

Chorus

It’s the circle of grams,
It’s the way of language,
  From the nonce to rote,
  From the true to trite
Till we turn old tropes
  From clichés to grammar ...
In the circle, the circle of grams.

Old phrases have lost all their lustre,
And we wonder just where it all went,
But the things that we blurt when we’re searching for words
Are newly coined crowns till they’re spent.

Some are reduced into nothing,
And others get very abstract,
And some are subverted, until they revert
To what they were once, way, way back.

Chorus

It’s the circle of grams,
It’s the way of language,
  From the nonce to rote,
  From the true to trite
Till we turn old tropes
  From clichés to grammar ...
In the circle, the circle of grams.
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>abs</td>
<td>absolute</td>
</tr>
<tr>
<td>acc</td>
<td>accusative</td>
</tr>
<tr>
<td>ag</td>
<td>agent</td>
</tr>
<tr>
<td>AGFOC</td>
<td>agent focus</td>
</tr>
<tr>
<td>CONJ</td>
<td>conjunction</td>
</tr>
<tr>
<td>COP</td>
<td>copula</td>
</tr>
<tr>
<td>dat</td>
<td>dative</td>
</tr>
<tr>
<td>DEM</td>
<td>demonstrative</td>
</tr>
<tr>
<td>EMPH</td>
<td>emphatic</td>
</tr>
<tr>
<td>fem</td>
<td>feminine</td>
</tr>
<tr>
<td>FNK</td>
<td>Funknet, electronic discussion group on functional linguistics.</td>
</tr>
<tr>
<td>imprf</td>
<td>imperfective</td>
</tr>
<tr>
<td>inf</td>
<td>infinitive</td>
</tr>
<tr>
<td>interrog</td>
<td>interrogative</td>
</tr>
<tr>
<td>masc</td>
<td>masculine</td>
</tr>
<tr>
<td>NEG</td>
<td>negation</td>
</tr>
<tr>
<td>pat</td>
<td>patient</td>
</tr>
<tr>
<td>pl</td>
<td>plural</td>
</tr>
<tr>
<td>prf</td>
<td>perfect</td>
</tr>
<tr>
<td>PRN</td>
<td>pronoun</td>
</tr>
<tr>
<td>PRT</td>
<td>particle</td>
</tr>
<tr>
<td>PRT-Q</td>
<td>interrogative particle</td>
</tr>
<tr>
<td>prtcpl</td>
<td>participle</td>
</tr>
<tr>
<td>sg</td>
<td>singular</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction with Reference to General Theories of Grammaticalization, Typology and Language Change.

1.1 Definition

Grammaticalization bridges the gap between synchrony and diachrony. It describes grammar in the process of forming, language in the throes of change, speakers in the act of making and remaking their language, not consciously, but with all the relentless vigor of a natural force.

Instead of relegating descriptions of morphosyntax to static snapshots, frozen in time, and studying historical change only in reference to a series of fossilized remains, we can now follow the pragmatic choices of speakers at the inception of a new grammatical form, examine how typological factors play a part in language drift, and tie discourse to syntax and syntax to morphology. We trace the paths of independent words as they are recruited into newly formed grammatical paradigms, become cliticized and bleached, and perhaps even disappear altogether.

Grammaticalization allows us to study specific cases of particular constructions in individual languages and to examine cognitive, functional, and formal aspects, all at the same time.

But what exactly is grammaticalization? The literature gives us broad guidelines, but no delimiting statement of the application of this term. The following is from Traugott and Heine (1991:2-3):

Alongside the lexical item:morpheme tradition which derives from Meillet, there has been a more recent tradition associated with Talmy Given, Charles Li, Sandra Thompson and others that focuses on the "packaging" of discourse and evolution of syntactic and morphological structure through fixing of discourse strategies. For example, Givon (1979:209) characterizes the process as one of cyclic waves involving

Discourse $\rightarrow$ Syntax $\rightarrow$ Morphology $\rightarrow$ Morphophonemics $\rightarrow$ Zero.
The preceding would serve as a handy description, but not necessarily as a formula for determining whether something does or does not qualify as grammaticalization. Traugott and Heine (1991:3-4) go on to describe the relationship between grammaticalization and language change as follows:

...grammaticalization is a kind of language change, subject to certain consequences and mechanisms of change... and characterized by certain consequences such as changes in grammar. These are increased syntacticization in its early stages, and increased loss of morphosyntactic independence in its later stages, ultimately leading to zero... Like other changes, grammaticalization spreads gradually across linguistic contexts on the one hand ... and across social contexts on the other...

But it does not mean that all change is grammaticalization. To take an obvious example, the semantic change from Middle English *bede* 'prayer' to its modern meaning of a small spherical object often used in necklaces ('bead'), is an example of change by metonymy, but not of grammaticalization.

The above statement strikes us as true and clear, yet it begs a definition. Examples are useful explanatory devices, but in order to test the bounds of our notions of grammaticalization, we cannot rely on description and example alone.

Mechanisms of language change listed in Traugott and Heine (1991:7) include:

(1) metaphorical transfer

(2) metonymic transfer

(3) reanalysis

and

(4) analogy.

But again, these mechanisms are common to forms of language change that do not constitute grammaticalization.

One of the strongest claims for grammaticalization is that it is unidirectional. "...more specifically of a pathway that channels change through a limited number of structures that are minimally different from one another." (Traugott and Heine
Unidirectionality is not limited to grammaticalization. "...for example, the well known phonological tendency for [sy] to become [s] but not vice versa is in itself not an instance of grammaticalization. However, if it results from morpheme boundary loss, then it is an instance of a stage in grammaticalization." One would tend to assume that unidirectionality, where documented, is a property of grammaticalization, but not part of its definition. However, Traugott and Heine go on to say: "The question is precisely what kinds of unidirectionality are necessary or at least typical of grammaticalization..." The exact meaning to be ascribed to the word "necessary" is unclear: a necessary precondition to the occurrence of grammaticalization as a phenomenon in the real world, or a necessary condition for us to recognize a change as being an instance of grammaticalization?

Counterexamples to unidirectionality are also mentioned in passing. "Virtually nothing is exceptionless, and there are of course instances of change in languages that are counterexamples of tendencies that can be characterized as 'less>more grammatical', 'main clause>subordinate clause', etc." (Traugott and Heine 1991:6)

However, Traugott and Heine also suggest that perhaps these instances that go against unidirectionality are not examples of grammaticalization after all.

"It is likely that all these examples are strictly speaking actually not cases of grammaticalization (although once they have occurred they may be subject to the generalization, reduction, loss, and other typical changes of grammaticalization. Rather, the examples Campbell and Greenberg cite can be regarded as instances of reanalysis." (Traugott and Heine 1991:7) [Emph. mine, AK].

This is all rather unclear, because we have already established that reanalysis is one of the mechanisms of grammaticalization. To characterize a change as reanalysis is not an automatic negation of its status as an instance of grammaticalization. Yet there is apparently some such implicature here.
Others have have different formulations. A recurring concept underlying many grammaticalization studies is the lexical/grammatical opposition. In this view of language, morphemes are classified by means of a binary system: each is deemed to be either ‘lexical’ or ‘grammatical’, but not both. The process of grammaticalization, from this point of view, is one in which lexical items are recruited for ‘grammatical’ use, thereby losing their lexical status.

"As is usually the case with words rich in implications, there are a number of different conceptions of grammaticalization. Yet there are central, prototypical instances of grammaticalization which most linguists would recognize ... For example, it is usually accepted that some kind of distinction can be made in all languages between ‘content’ words, also called ‘lexical items’ and ‘function words’, also called ‘grammatical’ words. ... Frequently it can be shown that function words have their origins in content words.” (Hopper and Traugott 1993:4)

The possible areas of confusion that come into play with such a dichotomy between lexical and grammatical categorization are as follows:

(1) words are often multimorphemic;

(2) in the process of deriving a word, lexical material and grammatical material are mixed;

(3) the admixture of lexical and grammatical material often crosses syllabic boundaries, so that in the derived form, the phonological word cannot easily be divided into separate grammatical and lexical parts;

(4) often synchronically the same lexeme has more or less grammatical function depending on its place in the phrase or sentence;
(5) in many languages, grammatical category on the lexemic level is very fluid, and a word can be used as a noun, a verb or a preposition without any change in its form. (Chinese, English, Pangasinan and Hebrew, to name just a few, have many words that transcend categorial boundaries without morphological change.)

All of the above points can be illustrated by the following Hebrew sentence:

"התכונתי את המדריכים לא לדורף על ההורר אל אשר עבורי זרר והע.

hidræti ‘et hamadrixim lo lidrox ‘al haderex ka-aser ′avarnu derex haja’ar.

instruct-1sg DO the-instructors neg to-step on the-road when passed-1pl through the forest.

‘I instructed the instructors not to step on the road when we passed through the forest.’

Of the eleven words in the above sentence, five are from the same root. I will list the related words below, numbering them consecutively as WORD (1) through WORD (5), but ignoring unrelated words that intervene in the sequence of the sentence.

WORD

(1) hidræti- ‘I instructed’. Literally : ‘I caused to step’

verb: hiph’il (causative), perfect, first person singular
root consonants: D R K (realized as d r x)
derivational elements hi- (depends on both derivation and inflection:perfect)
inflectional elements θ-a-ti (depend on both derivation and conjugation and person, and number)
WORD
(2) madrixim 'instructors'. Literally: 'causers to step'
-- noun: form present participle, hiph'il, masc. pl.
root consonants: D R K (realized as d r x)
derivational element ma- (depends on both derivation and inflection: participle)
inflectional elements: -ο-ι-ιμ (depends on both derivation and inflection: number and gender)

WORD
(3) lidrox -- 'to step'
nominal verb: infinitive, pa'al (simple derivation)
root consonants: D R K (realized as d r x)
prepositional prefix: li- (vowel realized this way by euphonic rule)
inflectional elements: θ-ο

WORD
(4) haderex -- 'the way (road)'. Literally: 'the stepping thing'
noun: segholate † derivation, definite
root consonants: D R K (realized as d r x)
definite article prefix: ha-
derivational elements: -ε-ε-
infectional elements: null, singular.

† Segholates are noun templates, which in the absolute, manifest the pattern Ce-CeC and penultimate stress. A segol is the short [e]. Historically, segholates derive from monosyllabic noun templates with a consonant cluster at the end (e.g. dark). Hebrew does not tolerate the cluster, hence the [e-e] pattern in the absolute.
WORD

(5) *derex* -- ‘through’. Literally: ‘By way of’ (construct)

preposition: from noun, segwolate derivation

root consonants: D R K (realized as d r x)

derivational elements: -e-e-

inflectional elements: null, singular.

The above analysis of the constituent parts of the following words illustrates the five points against the strict lexical/grammatical dichotomy as follows:

(1) **words are often multimorphemic**: All five of the words from the DRK root that I cited above are multimorphemic.

(2) **in the process of deriving a word, lexical material and grammatical material are mixed**: Those who adhere to the lexical/grammatical dichotomy would then be forced to try to decide which of the constituent morphemes are more grammatical and which are more lexical. Chances are that the root consonants would be judged to be more lexical, while the vowels and derivational consonants would be seen as more grammatical. My reply to this argument will be taken up below, in the discussion of open and closed classes. For the time being, however, let us simply note the futility of assigning a binary lexical or grammatical categorization to words whose constituent parts include both functors and contentives.

(3) **the admixture of lexical and grammatical material often crosses syllabic boundaries, so that in the derived form, the word cannot easily be divided into separate grammatical and lexical parts**: In the case of all five words in the above sentence, lexical and grammatical elements cross syllabic
boundaries. In fact, all the vowels come from derivational and inflectional morphology, so that no purely ‘lexical’ syllable can be found. In addition, choice of forms is dependent on both derivational and inflectional information, so that an element such as ‘ma-’ in *madrixim* cannot be assigned to either a derivational or an inflectional category alone. The choice of the form is dependent both on the *hiph’il* derivation and participial inflection.

(4) often synchronically the same lexeme has more or less grammatical function depending on its place in the phrase or sentence: The word *derex* appears as a noun in WORD (4) and as a preposition in WORD (5). This fact also illustrates the next point.

(5) in many languages, category is very fluid, and a word can be used as a noun, a verb or a preposition without any change in its form. Not only *derex* is subject to a categorization based on its syntactic position in the sentence, but the same is true of the word *madrixim*, which is seen to be nominal because it is preceded by the direct object marker, but might just as easily have been used as a present tense verb in a different sentence.

For the reasons illustrated above, I would strongly caution against any conception of grammaticalization that is based on the distinction between function words and content words. Perhaps we all have some sort of intuitive feel for what we mean by those terms, but our intuitions are very much influenced by the languages to which we have been exposed.

Most of us might agree without thinking that a preposition is more of a function word and less of a content word than a noun. We might also be tempted to proclaim that a preposition that is identical in form to a noun in a given language must have
been derived from the noun, and not the other way around. As a factual matter, I neither agree nor disagree with such a statement from the outset. But it is up to grammaticalization studies to prove such a historical progression: we cannot beg the question by defining it away.

If we think long and hard, we might find that we have difficulty deciding why a preposition must, perforce, be seen as more of a function word than a noun or a verb. The root DRK in Hebrew can mean ‘to tread on something, step, a way, direction, instruction ...’ It is not marked for part of speech. Its meaning is relational. So is the meaning of a verb, whose semantic variables include subject and object, a noun, whose significance in a sentence is determined by its relation to other sentence elements, or a preposition that ties together two participants in a spatial relation.

One way to get around the fuzzy nature of the functor/contentive terminology (Hockett 1958:264-265) is to resort to open and closed classes. Open classes are the sets of words that are most numerous in the language and to which one can presumably add new elements with the greatest ease. Closed classes are those that have few elements, and to which new elements can be added with great difficulty, if at all. In addition, the elements that are in the open classes appear less frequently, because they can be replaced by other elements from the same class. Elements of closed classes appear statistically much more frequently. Thus, in the following English sentence, the article the and the preposition over can be seen to belong to the closed class, while the word cow and jump and moon are considered to belong to an open class.

(2) The cow jumped over the moon.
The frequency of the as a member of the closed class in English is illustrated in the example itself, since it appears twice. Presumably, there are more occurrences of over than there are of jump or cow or moon. Likewise, the -ed affixed to jump has a very high frequency of occurrence and very few substitutes. But if we follow statistical guidelines alone, we might find that jump occurs much more frequently than either cow or moon, so that we would once more end up with a continuum of classes, rather than a binary opposition between closed and open classes.

Cummings (1995 FNK) had the following to say on the subject: “People usually talk about ‘open’ and ‘closed’ classes here, on the understanding that closed classes are ‘more grammatical’ because they are listed in the grammar instead of in the lexicon. But I think many linguists today reject the hard-and-fast distinction between ‘lexicon’ and ‘grammar’ that we used to operate with, and it’s not far from there to rejecting the distinction between closed and open classes. On the one hand, there are lots of reasons to want to let items like prepositions, conjunctions and even complementizers to be listed in the lexicon -- since they are clearly not simply ‘structure markers’, but rather they have the kinds of semantics we associate with lexical items -- and on the other, why should we consider ‘verb’ to be an open class in English, for instance, when virtually every high-frequency verb has its own unique set of distributional properties? The insights about the highly idiosyncratic grammatical distribution of all kinds of items brought to us courtesy of construction grammar should also shed some doubt on the open/closed opposition. So: if you use the term grammaticization, what’s your theory of what is and isn’t ‘in the grammar?’

An examination of the general structure of Hebrew militates against using open
and closed classes as a measure of lexical or grammatical status, at least for crosslinguistic purposes. Hebrew has a strong root system, and most words can be identified as belonging to a particular root. Most roots consist of three consonants. Thus, using an idealization based on ordered groups of three, with an inventory of twenty-two consonants, theoretically, 10,648 roots are possible. In reality, most of those possibilities do not actually occur. But in theory, this presents us with a rather large, but closed, potential set. Thus what would traditionally be considered the lexical material of the language is organized in a closed set.

The derivational system for verbs is rather small, consisting of seven primary derivations, but nouns and adjectives can be formed from a much larger set of templates.

In Hebrew, the boundaries between derivation and inflection are blurry, since one can always derive a new noun or adjective by taking a verbal root and conjugating it into a nominal pattern. Patterns for derivation are equally accessible to speakers as patterns for conjugation and inflection.

If we lump derivation and inflection together, then in Hebrew lexical (root: consonantal) and grammatical (derivational/inflectional: mostly vowels) forms are in about one to one correspondence in most utterances (even unto the syllabic level.) Common roots and common grammatical formations are likewise about equal in frequency. Thus open and closed classes and frequency of occurrence are of little help in determining lexical or grammatical status in Hebrew.

The only words that might qualify for pure grammatical status under such a definition are the affixed prepositions whose roots have been so worn down that they consist of a single consonant. But by and large, most of the grammatical structure of
Hebrew is to be found in the seamless interweaving of triconsonantal roots with vocalic patterns. The set of consonants and the set of vowels are equally closed, even if the former are more numerous than the latter. Thus a proposed measure of this sort would apply to Hebrew only in those places where its grammar is uncharacteristically isolating or agglutinative.

If our aim is to find a crosslinguistic formulation of grammaticalization, we should use the terms ‘lexical’ and ‘grammatical’ guardedly and not as absolute binary oppositions. Grammaticalization includes a larger scope than the conversion of lexical material into grammatical use. “Grammaticalization phenomena are essentially gradient and variable. They proceed by minimal steps ...” (Traugott 1996:3). Grammatical as opposed to lexical status is merely a matter of degree. The terms are useful when applied within a particular language, but their application varies considerably form language to language.

There are other issues that must be addressed in a clear delineation of grammaticalization. One of the questions that arises is precisely what is it that becomes grammaticalized? Is it always a single word or affix?

... “in most works of grammaticalization, there appears to be an assumption to the effect that this process affects only single lexical items. This is in line with the by now classical definition volunteered by Kuryłowicz (1965:52), according to which grammaticalization ‘consists in the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status.’ In such studies, the important fact tends to be ignored that many instances of the process involve at the same time more than one linguistic item. In fact, quite a number of conceptual processes leading to the development of grammatical more complex conceptual entities. For example, in the grammaticalization of perfect aspects in a number of European languages, at least two markers were involved: an
auxiliary 'have' or 'be' and a marker of nonfiniteness, which typically was a passive participle morpheme, and progressive aspect constructions in many languages world-wide even involve three distinct morphological elements: an auxiliary verb, a nominalization marker, and a locative element. Such a complex contraction can also be observed, for example, in the English Future marker be going to." (Heine 1993:30).

Another issue of concern to some linguists is the regularity by which newly grammaticalized material displaces older forms in a paradigm.

"Is it appropriate to use grammaticalization to refer to the kind of change which leads to suppletive paradigms for the surviving word?" (Gildea 1995 FNK)

My own response is that the suppletive or non-suppletive nature of the item recruited into a paradigm does not take away from its relative grammatical status, and that to the extent that an item has become more conventionalized and abstract, so it can be held to have taken a step along the grammaticalization continuum. That I am not alone in this view can be seen by the following response from Elizabeth Traugott:

"Go-went involves decategorialization of the parts of the verb that became suppletive, but there is no generalization; therefore I think that go-went suppletion is lexicalization not GR. On the other hand, dem > copula involves change in function (and in the example given) generalization. The same holds for the third case, reanalysis of V > Pro." (Traugott 1995 FNK)

The consensus seems to be that if a lexeme has been generalized, abstracted or bleached, and conventionalized to the point of serving as part of a language-wide grammatical paradigm, then it has undergone grammaticalization. On the other hand, if it has merely been recruited as an allolex of a single lexeme, then
grammaticalization has not taken place. The key issue is the cognitive process involved, not the morphological results.

“It seems to me that the ‘correct’ answer to this is whether the type of human cognitive processing involved in all ... changes is the same, or different. If the type of change cognitively is the same, then call them all by the same term. If the type of change is somehow substantively different, call them by different terms. Perhaps I'm naive here, but it does seem to me that all ... involve (re)routinization or (re)automization of language behavior.” (Payne 1995 FNK).

If we accept that grammaticalization is an ongoing process, we can conclude that that items are not grammaticalized or ungrammaticalized, or lexical as opposed to grammatical, but that it is a question of degree of grammaticalization.

Bybee has formalized this approach in her research and has given quantitative measures of degree of grammaticalization. She has proposed three measures of formal grammaticalization: shortness, dependence and fusion.

“For measuring reduction, or as we shall call it, shortness, we simply count the number of consonants and vowels present in the gram ...” (Bybee 1994:108).

“We distinguish between the general loss of autonomy, which we call dependence, and the actual fusion of the gram with the verb, which we call fusion.” (Bybee 1994:110)

“Dependence ... includes all those indices of loss of autonomy that are not specific to a growing fusion with the verb. They are:

“(1) the number of allomorphs of the gram, not counting those that are purely phonetically conditioned ...

“(2) Conditioning for allomorphs ...
"(3) Suprasegmental reduction ...[e.g. can it carry stress?] (Bybee 1994:110)

"Fusion with the verb includes all the indices that give evidence for a degree of fusion of the gram with the main verb stem. They are:

"(1) Written bound ...
"(2) Open class intervening
"(3) Phonological process conditioned by the stem
"(4) Lexical conditioning ...
"(5) Conditions stem change ..." (Bybee 1994:112)

Bybee then seeks to match up differing levels of semantic bleaching with various degrees of formal grammaticalization as per the above quantifiers.

The approach is laudable in its attempt at objective measurement. However, it presents the following pitfall: more reduced and dependent forms are assumed to be ipso facto more grammaticalized than those which are longer and more independent. This is because Bybee assumes that grammaticalization is a unidirectional process that leads from independent forms to affixes.

According to Bybee’s terminology, a more dependent form is a more grammaticalized form. I prefer an approach whereby that assumption is open to falsification, and is not a part of the definition of grammaticalization.

Is grammaticalization a process or a result?

If we opt for the result definition, there is really not much difference between saying that a linguistic unit is ‘grammatical’ (as opposed to ‘lexical’) and saying that it is ‘grammaticalized’.

If we opt for the process definition, we may find that after undergoing a considerable amount of grammaticalization, a unit becomes less grammatical (and more lexical) than when we started observing it.
To avoid this lack of clarity, I would propose the following working heuristic, which accords with both the spirit and the letter of the majority of available studies on grammaticalization. By this heuristic, grammaticalization is a process (from point A to B in time) that leads to an effect: more conventionalized patterns, not a particular method by which that effect is achieved, and not a synchronic result of any stage of that process.

**DEFINING HEURISTIC:** Grammaticalization is a process whereby independent linguistic units are recruited into grammatical paradigms or into systems that form as a result of this recruitment. Such independent linguistic units may be lexemes, phrases or even larger units, they may be derived or opaque as to derivation. In the process of recruitment, these units lose their individual identity and merge into a newly formed grammatical pattern. The grammatical patterns into which such units are recruited may be on the morphological, lexical, syntactic or discourse particle level. The process may involve more than one of these levels simultaneously. Degree of Grammaticalization refers to the number of specific identifiable changes that a unit has undergone under the grammaticalization process, and not to the extent to which it is reduced or dependent at any given point of time.

Naturally, grammaticalization is a continuum, and in plotting its progress, we cannot expect for there to be discrete and numerable instances of its occurrence in reality. The number of identifiable changes is a very rough measure. Nevertheless, what we are interested in doing is finding out how far along the continuum of change the item has travelled. We are accustomed to slicing up the continuum into discrete units for our own purposes. So an example of a single change could be something like: from *a head* (noun) to *ahead* (adverb). Every change that can be listed as a single event can be further sliced into as many subevents as the researcher prefers, depending on degree of precision desired. To compare the degree of grammaticalization of two items, we must use the same degree of precision in our slicing. The important point is that if a unit first becomes dependent and later breaks free, we will not consider it to be less grammaticalized because it is less dependent: we will consider it more grammaticalized, because it has undergone more numerable changes than its affixed form, not fewer. By limiting the definition of
grammaticalization to a process leading to temporary results, we have left open the question of methods of change or immutable laws concerning direction and manner of change. We have also left open the issue of whether after several such processes the ultimate results may not nullify intermediate developments.

§1.2 General Assumptions about Grammaticalization

Keeping the above working definition of grammaticalization in mind, we can now turn to a list of properties generally attributed to the phenomenon, with a view to confirming, disproving, or delimiting the bounds of each.

It is generally agreed that grammaticalization:

(1) is unidirectional.

(2) has an isolating starting point.

(3) involves increased synthesis, bleaching and reduction.

(4) ultimately may result in complete loss.

(5) can stop at any point, but a form can never return to an earlier stage, once it has reached a certain degree of grammaticalization.

(6) is quantifiable into degrees.

(7) leads to layering, where older forms persist alongside newly grammaticalized versions.

This study accepts most of the above assumptions, not as axiomatic, but rather because they are tendencies which have been established by countless examples. We will examine (1) and (2) and (5) and (6) more closely, not necessarily with a view to disproving them altogether, but with a critical eye to their universal application.

§1.2.1 Unidirectionality and Irreversibility

In one sense, all language change -- and in fact all change of whatever sort -- is unidirectional. This results from the unidirectionality of time. Whatever has happened cannot be made to unhappen.
In addition to this rather ontological fact, there are the less dramatic effects of entropy: everything tends to disorder. There is no point crying over spilt milk for two reasons: (1) you cannot reverse time and cause the whole incident never to have happened and (2) while it is possible to take positive action to clean up the mess, spilt milk is difficult to retrieve: you cannot refill the glass from the spillage.

This is generally why changes (both linguistic and physical) tend to leave detectable traces. If it were not for this fact, historical linguistics (or geology or forensic pathology) could not yield as many useful results.

However, under appropriate circumstances things can come full circle. For instance, the glass can be refilled with new milk. Or, to use a more appropriate metaphor, rain water evaporates, collects into clouds and condenses and rains down again.

Cycles run through nature and language. Nothing is ever restored to its original state, (hence we can prove that a change occurred), but sometimes things achieve a state very similar in form and function to that which prevailed at a previous point in time. It is to this type of ‘restoration’ of the \textit{status quo ante} that we will refer as a cyclical change. Thus, while there is absolute unidirectionality in the process of any change, the results of the unidirectional change can be cyclical.

My view of grammaticalization is entirely consonant with the unidirectionality of change: it leaves open the issue of cyclicity of result.

In claiming unidirectionality as one of the cardinal attributes of grammaticalization, we seek to obtain the same status for this phenomenon as is enjoyed by other well established forms of language change: chief among them, sound change. But if we examine the matter closely, we will find that sound change, too, is subject to cyclical waves and that a sound or phonerne that has been lost can, and sometimes does, reappear in the system.
According to Hockett (1965:203): "Whenever sound change does bring about a specific SOUND SHIFT, the structural ramifications are so numerous, so scattered, so subtle, and communicatively so utterly trivial that a return to a state of affairs before the sound shift has a vanishingly small probability." Yet, "sound change may threaten a contrast that is of some communicative importance. When this happens, people say 'I beg your pardon?' more often, and speakers either repeat with more precise enunciation or find paraphrases." (Hockett 1965:202). How do we know that a sound change is merely threatening to happen and is averted by more precise enunciation, as opposed to having occurred and then been reversed?

The issue here is not of a hypothetical change upon which we base an argument: the issue is this, what constitutes historical evidence of change?

Bybee (1994:10) grapples with the difficulty of identifying the inception of grammaticalization as follows: "One problem in identifying the properties of lexical items that are candidates for grammaticization is the problem of determining at exactly what point we can say that grammaticization has begun ... A case to consider in this regard is the use of body part terms in grammatical constructions signaling spatial relations ... [N]ote that it is not 'face' with its specific body part meaning that enters into grammatical constructions. Rather 'face' generalizes perhaps by metaphorical extension to mean 'front' ... and only after it has taken on the sense of a general spatial relation does it enter into the grammaticization path by which it can become a spatial adposition."

If we regard the process of grammaticalization as one that comes about through increasing generalization, as Bybee has suggested (1994:6) -- "Since at all levels we see increasing generality ... leading to wider ... use, the evolution of grammatical material is best viewed as a single continuum along which the same processes are
operative" -- then it follows that the generalization of 'face' to mean spatial location is already a step along the path of grammaticalization. When we consider that the more literal application of 'face' coexists with the spatial use, even greater difficulties in tracking the path present themselves.

Returning to the unidirectionality of phonological change, discussed above, I would like to invoke the example of the French *h aspiré*, a segment whose phonological status is still being hotly debated. The real problem may be that while the phone [h] has for the most part disappeared, the segment has not entirely been eradicated as a contrastive unit in the phonology of the language. When something is both "there" and "not there", we have an opportunity for a change to seemingly reverse itself.

Popular French lost its Latin 'h' at the end of first century A.D., while the upper classes retained it in their speech for some time thereafter. The Germanic 'h' began to fade away in the thirteenth century A.D. In 1582, H. Etienne criticizes the ignorance (or negligence) of the Parisians, who among their many other faults, fail to pronounce the *h aspiré*: "Multi perinde pronuntiant ac si scriptum esset *un ocqueton*, et *un'aute maison*, sic *un'onte*, *un'aquenee*, *un'arpe*... eodemque caetera modo proferunt. Et ..*il m'ait pro il me haiit." (Fouche 1961.580). [Many pronounce it as if it were written *un ocqueton*, and *un'aute maison*, thus *un'onte*, *un'aquenee*, *un'arpe*... And in the same way they carry out other <words>. And *il m'ait for il me hait."].

While the process of weakening and eventual loss was well underway, as late as 1777 grammarians still described the Germanic *h* as being pronounced with an audible blowing out." (Fouche 1961.581).

Eventually, all that remained of the Germanic *h* was its disjunctive function, in that it prevented liaison and elision from taking place as they would with a word beginning in a vowel or an 'unaspirated' Latin *h*. 
And yet, even today, occasionally, in emphatic speech, where unusual stress is placed on the first syllable, an actual aspiration is pronounced for an *h* aspiré. E.g. *Il est hideux!* -- ‘He/it’s hideous!’ or *C’est un honte!*. -- ‘It’s a shame!’ (Grevisse 1980,44).

If the aspiration has been lost, and as some linguists argue, phonemic status has also been lost, how is it that speakers still know where to pronounce the *h*?

Socioeconomic factors and literacy play a part in this phenomenon. The less literate, rural communities are the ones who have retained the older more aspirated pronunciation, while the urban working class has gone so far as to ignore the *h* aspiré for purposes of elision and liaison. The educated middle class finds itself somewhere between these two extremes. Elision and liaison are maintained, and when an occasion calling for great emphasis arises, the actual aspiration is produced.

A similar situation can be found with modern Hebrew *ayin* and *aleph*, the first having been a pharyngeal in Classical Hebrew, the second, a glottal stop. In fact, both of these phonemes are currently being realized as beginning a new vowel-initial syllable. The loss is complete for working class speakers from European descent, while Middle Eastern ethnic speakers still pronounce the pharyngeal. The educated middle class speaker straddles the fence. An *ayin* is not pronounced as a matter of course, but is readily produced in situations of ambiguous context. For instance, the word for ‘happiness’ starts with an *aleph* and the word for ‘wealth’ starts with an *ayin*. in all other respects they are the same. For a middle class European native speaker of Hebrew, they are homonyms. But if the context is ambiguous and the hearer looks confused, the middle class speaker will repeat the word for wealth with the pharyngeal inserted. For most middle class speakers who are not from a Middle Eastern background, this phonetic clarification is performed naturally and automatically and is not an affectation.
Have the $h$ aspiré and ayin been lost and restored? Probably not. A simpler explanation would be that they enjoy phonemic status, which occasionally sees its way to phonetic expression (Bennett 1988.2). But suppose that as a result of regular overemphasis they were to become pronounced as an aspiration and a pharyngeal, respectively? That might be viewed as an example of a sound change that came full circle. In any event, if that did occur, the process of change would be unidirectional, while the result would be cyclical.

According to Labov (1977:229), something of that nature actually happened in English. A complete merger seems to have taken place between ea and long a, but later the ea and long a words, under the influence of another sound change, split into the same two groups once more.

Even so, it is difficult to deny that most sound change is irreversible. In fact, most changes in language are not reversed. But cyclical change does occasionally happen. Its rarity makes it all the more intriguing. We would do well to ask ourselves, what special conditions allow a change to come full circle?

That is what I propose to do in this dissertation with regard to pronouns that become copulas and copulas that become pronouns.

### 1.2.2 Cyclicality and Counter-Examples to Unidirectionality

Delancey (1985) describes a three-stage cycle for the grammaticalization of deictic motion verbs, ‘come’ and ‘go’, in Tibeto-Burmese. The cycle begins with a general motion verb, unspecified for directionality, which is placed in collocation
with deictic particles. In stage two of the cycle, the particles are morphologized on
the verb. In the third stage we have recruitment of new deictic elements into the
obligatory syntax, despite the older deictics whose relics form part of lexical
material in the verb. In other words, the process has come full circle.

(3) Harsa bharad wan-a ‘Harsha went to India’
           Harsa India go-perf

(4) Harsa nepal wal-a ‘Harsha came to Nepal’
           Harsa Nepal come-perf

In (3) and (4) the ‘come-go’ words are etymologically bimorphic, coming from
a general motion verb and a deictic, but they are no longer seen that way by
speakers. (Delancey 1985.368)

(5) jhego baw-ya wan-a ‘the bird went flying, flew away’
            bird fly go-perf

(6) jhego baw-ya wal-a ‘the bird came flying, flew hither’
            bird fly come-perf

In (5) and (6) the generalized motion verbs are being used as deictics in
combination with more specific motion verbs.

Delancey gives a cognitive explanation for this unidirectional cycle. His is the
lexis-synthesis circle, in which a lexeme once synthesized, never gets reinterpreted
as an independent unit again. Instead, other lexemes are recruited to fill its place.
Nevertheless, this cyclicity of process, might be helpful to understand other instances
where there is actual cyclicity of result: "...we have a competition between iconicity
as applying to two different conceptualizations of the same phenomenon." (Delancey
1985:384). That is, we can think of ‘coming’ as movement first and foremost and
deal with direction of motion separately, or we can think of the entire event as a
concept. People do both. "Thus the inability of the TB languages to settle on a stable
mechanism is a consequence of the insufficiency of the resources of linguistic
structure for dealing with the full complexity of cognition."

The idea of a conceptual phasing in and out is useful. It is one we will explore
further, after our discussion of the cognitive domain of pronouns and copula.

Bybee presents an actual counterexample to unidirectionality of result. It is
particularly a propos here, since it deals with the lexicalization of a pronominal
affix. "Only one example of affixed material that has become free has come to our
attention, and in this Irish case there is a strong paradigmatic pressure for the
reanalysis of a person/number suffix as a free pronoun." (Bybee 1994:14)

The first person plural suffix ‘mid/muid’ is the person/number suffix that has not
been lost in most Irish dialects. "It occurred in the paradigm in the same position as
the free subject pronouns. Now the suffix ... can occur as an independent pronoun,
replacing the earlier first plural pronoun 'sinn'. " (Bybee 1994:14)

Bybee dismisses this as an isolated example, but she concedes that "under very
special circumstances, an affix can become free again." (Bybee 1994:14).

This dissertation will discuss other examples of cyclical change and will explore
in some detail the "very special circumstances" that make them possible.
§1.2.3 Isolating Starting Point and Degrees of Grammaticalization

Heine et al. (1991.6-7) quote Von Humboldt approvingly with regard to his proposed four stage evolution of language:

Stage I (which he calls “the lowest stage”: idioma, phrases and clauses;
Stage II: fixed word order and words vacillating between “matter and form meaning”;
Stage III: “analogs of forms,” which are “pure expressions of relations”;
Stage IV (“the highest stage”): “true forms, inflection, and purely grammatical words”
(Humboldt 1825:66)

The difficulty with this approach to language development is that it begs an isolating starting point for the whole grammar, and then assumes increased synthesis without any deviation. Many works on grammaticalization tend to treat every language studied as if, at its inception, the language had been isolating. Given that languages develop from other languages, as English developed through various stages from Proto-Germanic, it is unreasonable to postulate a holistic starting point without any specific information to that effect about the language in question.

A fall back position would be to reply that while the language as a whole may not have been isolating at any given point in time, when we study the history of a particular morpheme, we can properly assume that it began as an lexical item. This, too, is potentially misleading. Words usually derive from other words, as younger
languages do from their predecessors. To know the degree of grammaticalization of a particular morpheme, we would have to reconstruct its development from point A to point B, and even then, we would not be able to prove what happened to this morpheme prior to point A.

According to Bybee, unidirectionality of grammaticalization has direct consequences in typological dating of morphemes. "...we might conclude that grams in isolating languages are less grammaticized -- in both meaning and form -- than grams in agglutinative languages, which are in turn less grammaticized than grams in fusional or inflectional languages." (Bybee 1991.44-45)

Though Bybee's assertion may be intuitively appealing, the possibility of recycling of the same gram, from an independent to a cliticized to an independent form once again, precludes the ability to determine the degree of grammaticalization by a synchronic examination of the state of a language. The assumption that grams in isolating languages are ipso facto less grammaticalized than those in inflectional languages implies that the isolating language has undergone fewer cycles of change. I believe such an assumption is unfounded.

§1.3 Conclusion

Of what possible significance is the source of a free lexical item to its recruitment into a grammatical paradigm?

The source of any lexical item, or of any morpheme for that matter, free or dependent, bears little relevance for speakers of the language to the extent that its
origin is synchronically opaque. When we decide to study the history of any morpheme from a given point in time, we do not need to know its history prior to that point. However, neither can we categorically assume that because a gram presents as independent, it has undergone less grammaticalization than one that presents as affixed. Also, for purposes of prediction, the longer the chain of grammaticalization that we observe, the more we can say about the possible long-term results of grammaticalization of similar grams under similar circumstances. For those of us interested in long range comparison, observation of lengthy cycles of grammaticalization can teach us a great deal about the sorts of permutations that can occur over long periods of time, when the same principle of change is applied over and over again.

The issue of whether languages always become more synthesizing with the passage of time, or can in fact tend toward isolation after a period of relative morphological complexity, is an important one for purposes of exploring language genesis and development. Models of language genesis that assume all words of evolving or ‘primitive’ languages must have been noun-like and that grammar arose from grammaticalization of concrete nouns into verbs and adpositions†, and later on into inflection, are simplistic, at best. They are not far from Von Humboldt’s classification of existing language according to level of ‘sophistication’. Such models underlie what is worst in current grammaticalization theory. These are some of the reasons why the sources of free lexical items are of interest to those who wish to explore language development.

† I believe it is more common for nouns to develop from verbs. But see Heine, Claudi and Hunnefeyer (1991(2):160) where a developmental hierarchy is suggested of ‘human noun, non-human noun, verb, adverb, adposition, adjective, adverb’. Observation of bodypart terms and their extensions seems to foster such a view: e.g. head (verb) from head (noun).
Grammaticalization is not a static property, nor a specific result. It is a process whereby independent linguistic units are recruited into new grammatical systems, whether on the phonological, morphological, syntactic or pragmatic level. The results of grammaticalization are temporary and subject to the re-application of the same process over and over again. Language change, like most forms of change, tends to be unidirectional and irreversible, by and large. This holds true for grammaticalization. However, cycles and cyclical change do occur. This dissertation will examine such cyclical change and the cognitive and functional conditioning that allows it to take place.
Chapter 2. The Cognitive Function of Copula and Pronoun

§2.1. Introduction

This chapter discusses briefly the cognitive and functional connection between the copula and the pronoun. The method used is crosslinguistic comparison of morphological and syntactic expressions in surveyed languages. I have made an effort to include in my survey languages whose genetic affiliation and synchronic typologies are very different, in order to avoid basing my conclusions on extremely localized and language specific phenomena. No claim is made to absolute universality, and any generalizations are based on the data and are open to falsification in their application to languages not surveyed.

The goals of this chapter are as follows:

(1) To establish broad functional parameters for the copula and the pronoun,

(2) To determine how morphological and syntactic patterns may reflect cognitive metaphors that underlie the origin of such expressions,

(3) To determine what the copula and the pronoun have in common cognitively, and

(4) To establish a heuristic for determining when a morpheme is being used as a copula rather than a pronoun, despite their common elements.

In the following sections we will establish that the copula in the surveyed languages is expressed by two separate patterns at most, even though the statements conveyed can be grouped into four pseudo-logical categories of identity, class membership, existence and possession.
In eight of the ten languages surveyed, (Chinese, French, Hebrew, Hungarian, Korean, Russian, Turkish and Vietnamese), one morpheme was used for both expressions of existence and possession, while another was found in statements of identity and class membership. In one language, (Finnish), the same morpheme was used for all four categories. In yet another, (English), the morpheme for existence, identity and class membership was the same, while that for possession was separate and distinct.

We will see that it is the copular morpheme for identity and class membership that patterns with the third person pronoun. Third person pronominal marking is, in turn, separate and distinct from first and second, as evidenced by morphological characteristics crosslinguistically.

The pronominal grammatical category has deixis at its cognitive core, and while first and second person refer to situationally salient participants, third person requires a greater level of abstraction. When referring to a third person participant for the first time in discourse, speakers are in fact also presupposing the existence of such a participant. Thus third person anaphora is cognitively related to predication of existence.

Since the flicker between a nominal interpretation and a verbal interpretation of a morpheme denoting ‘a being’ or the ‘act of being’ is what motivates reanalysis of copulas as pronouns and vice versa, the question arises: How do we know whether a particular morpheme is a pronoun or a copula at any given point in time and in any specific utterance? The heuristic I will use throughout this dissertation will be whether morphosyntactically the morpheme in question is patterning more closely with the nominal or verbal category of the language as it is synchronically constituted. If it behaves like a nominal, it is a pronoun. If it behaves like a verb, it is a copula. If it is hard to distinguish, then we are in flicker mode, the mechanism that allows pronouns and copulas to develop one from the other.
§2.2 Introduction to the Copula

The copula is distinguished in the ten languages surveyed below by a number of recurring characteristics. Basic copular sentences, for identity and class membership, are considered to be equative, with nominals taking on roughly equal status. In most languages where case is morphologically marked, the nominals “equated” in such a copular phrase are both in the nominative. The order of the nominals can be reversed without changing the purport of the proposition. The copula, even when expressed overtly, and even when it fills the syntactic position of a verb, is very unverb-like in its function.

If a prototypical verb expresses a specific event relationship between participants A and B, the copula is the most empty of all verbs semantically, since it merely states the existence of a specified pseudo-logical relationship between A and B, and there are no event countours. It is not simply that the copula is intransitive, and thus automatically further removed from the verbal prototype. An intransitive verb of motion, of action, or one expressing a mental state is far more concrete and lexical in its purport than one conveying a relationship of existence, identity or class membership. Not that existence, identity and class membership are, strictly speaking, accurate descriptions of the semantic relations that hold in such sentences. We will see that they are in fact metaphors by which the grammatical coding of copular expressions is organized. The forms are based on these paradigms; the semantics is richer. Recurring logical metaphors on which copular morpho-syntax is based are listed below:

(1)(a) ‘A exists’ [in many languages we must specify where]
      (b) ‘A is identical to B’
      (c) ‘A is a member of the class B’
      (d) ‘A is a member of the class of things belonging to B’.

[in (d) B would be in an oblique case, since this a variation on (a):‘ A exist to B’]
The exact configuration of semantic relationships conveyed by a single copular expression varies from language to language. This chapter seeks to identify typological traits of languages that correlate with copular function.

For instance, of the languages surveyed, those with strong, productive morphologically expressed case systems, whether agglutinative or inflectional, and flexible word order, tend to omit any overt expression of the copula in the present tense in sentences expressing identity (e.g.: Biblical Hebrew, Hungarian, Korean and Turkish). When the two participants, A and B, are both in the nominative, the relationship between them is one of metaphorical identity, and the expression of the copula as a morphologically full verb is deemed superfluous. On the other hand, in these languages, where existence is being expressed, an overt copula is used, because direct statements about existence involve only one participant, and absent contextual priming, are not self explanatory.

In addition to correlations between copular use in typologically similar languages, we will also take account of parallels in the use of copulas that transcend typological similarities. We will see that languages as different from one another as French and Chinese will sometimes show congruent patterns in the use of the copula, patterns that they do not share with other more closely related languages, or with languages whose morphosyntactic typology we might consider to be more similar.

The classification of a present tense copula as a type of verb is in itself debatable. In a majority of the surveyed languages (Chinese, French, Hebrew, Hungarian, Korean, Russian, Turkish and Vietnamese) the relation of identity or class affiliation is expressed syntactically (e.g. ‘A B’) rather than morphologically (e.g. ‘A is B’). When the expression of these relations is morphological, as in existence and possession, we sometimes find that the copula does not convey all the distinctions of person, number and gender normally marked in a verbal paradigm (Turkish, Hebrew), or the syntactic position of the copula is different from that of verbs in the language, and the copula is considered a ‘particle’, not a ‘verb’ (Classical Chinese).
In addition to the semantic relations listed in (1), above, the copular morpheme is also simultaneously recruited to express the following in a number of languages:

(2) (a) Affirmation (Chinese, Turkish)
   (b) Emphasis or echo (Chinese, English)
   (c) Interrogative particle or expression (*Hebrew, by reconstruction)
   (d) Auxiliary verb (Indo-European)
   (e) Demonstrative (Chinese, Hebrew, Turkish)
   (f) Definitizer (*Hebrew, by reconstruction)

Expressions either asserting or asking about the truth value of a previous statement have a tendency to resort to a form of the copula. By the same token, that which is asserted to exist or to be a token of a type is sometimes deemed definite. It is a short step from a demonstrative to a definitizer. And the road from a demonstrative to a pronoun is likewise not too circuitous.

For the purposes of this dissertation, I will first classify a form as a copula on the basis of the semantics in (1), but once it is so classified, I will list functional overlaps in (2), below.

Figure 1 lists traits of the copula in ten sample languages. For this broad view, the designation ‘HEBREW’ may be seen as encompassing both modern and ancient Hebrew. Responses tabulated would work for either. The Classical Chinese forms, in contrast to Mandarin, are given in parenthesis.

Figure 1 presents a small sample of the variety of ways in which copular functions are organized. The examples are all in the third person present, maximizing the similarities. It is important to note, however, that in Hungarian an overt copula would be used for any person other than third. Also, where A and B are pronouns rather than nouns, certain ambiguities develop in Modern Hebrew and in Turkish and to a lesser extent in Russian, as to whether a form is a pronoun or copula.
## Expressions of Identity, Class Membership, Existence & Possession

<table>
<thead>
<tr>
<th>CHINESE</th>
<th>ENGLISH</th>
<th>FINNISH</th>
<th>FRENCH</th>
<th>HEBREW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A shi B</td>
<td>A is (the) B</td>
<td>A on B</td>
<td>A B (A hu B)</td>
</tr>
<tr>
<td></td>
<td>(A B ye)</td>
<td></td>
<td></td>
<td>A is B</td>
</tr>
<tr>
<td>2</td>
<td>A shi B</td>
<td>A is (a) B</td>
<td>A on B</td>
<td>A B (A hu B)</td>
</tr>
<tr>
<td></td>
<td>(A B ye)</td>
<td></td>
<td></td>
<td>A is B</td>
</tr>
<tr>
<td>3</td>
<td>[LOCATION] you A</td>
<td>There is A</td>
<td>on A</td>
<td>Il y'a A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>jés A</td>
</tr>
<tr>
<td>4</td>
<td>B yóu A</td>
<td>B has A</td>
<td>B ade on A</td>
<td>B a A</td>
</tr>
</tbody>
</table>

1 = IDENTITY  2 = CLASS MEMBERSHIP  3 = EXISTENCE  4 = POSSESSION

<table>
<thead>
<tr>
<th>HUNGARIAN</th>
<th>KOREAN</th>
<th>RUSSIAN</th>
<th>TURKISH</th>
<th>VIETNAMESE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>van A</td>
<td>A-nom. iss-dec.</td>
<td>jest A</td>
<td>A var</td>
</tr>
<tr>
<td>4</td>
<td>a(s) B van A-poss.</td>
<td>B-top. A-i iss-dec.</td>
<td>u B jest A</td>
<td>B-gen. A var</td>
</tr>
</tbody>
</table>

(Common shading within respective boxes of a language indicates use of a common morpheme. Classical Chinese forms are in parentheses)

**Figure 1**
The rows of Figure 1 correspond to the subdivisions of (1) below.

(1)
(a) 'A is identical to B' -- **ROW 1** (A is B)
(b) 'A is a member of the class B' -- **ROW 2** (A is a B)
(c) 'A exists' -- **ROW 3** (There is an A)
(d) 'A is a member of the class of things belonging to B'. -- **ROW 4** (B has A)

The shaded boxes of Figure 1 indicate use of the same morpheme (overt or null) for the expressions of different relations under (1) above within the same language. By the same token, boxes within the same language that are not shaded share the same morpheme, or lack of a morpheme, in the expression of such relations. The collocations chosen were those most frequently used within each language, although alternate forms exist. Interestingly, for the most frequently used collocations expressing the four relations of identity, class membership, existence and possession, none of the languages examined used more than two different modes of expression. That is why a binary distinction of shaded versus white was sufficient.

§2.2.1 The Relationship Between Being and Having

If we continue to look at the copula in terms of a quasi-logical framework, which we will do for the present section before discussing a more pragmatic and discourse oriented approach, we will find that certain patterns emerge that are consistent despite vast differences in language typology.

There seems to be a marked morphological correlation between rows 3 (existence) and 4 (possession) on the one hand, as opposed to row 1 (identity) and row 2 (class membership, attribute), which are also closely linked to each other. Thus, in Chinese, Hebrew, Korean, Russian, Turkish, and Vietnamese, the same morpheme that is used to signify existence (你, 有, 有, 是, 是, 可, 为, and so) appears also in the
expression for possession. On the other hand, any morpheme used for class membership or attribution statements is a different word altogether from that used for existence and possession.

The main exception here is English, where boxes 1, 2 and 3 share a morpheme, a form of the verb to be, whereas box 4 uses a form of ‘to have’. (Finnish is unusual in that all four boxes use on.) To have in English takes us entirely out of the equative paradigm; to have takes an object. We are subjecting the possessed to the possessor, so that the expression is not symmetrical. And yet in the expression of existence, English uses the equative pattern which displays complete symmetry.

Of course, English is one of many languages to use a specialized transitive verb to express possession. Several other Indo-European languages do so as well: German and French, to name two. Indo-Europeanists are quick to note that the verbs have and haben in German and English are entirely unrelated to the French avoir. The Latin capere, meaning to grasp, is the Italic reflex of its Germanic counterpart in haben.

PIE does not have a reconstructable specialized verb meaning ‘to have’. Both Latin habere and German haben are later developments. The semantic path, whereby ‘to hold’ means ‘to have’, is not too difficult to grasp. So Spanish tener meaning ‘to hold’ also means ‘to have’.

(3) Tengo muchos libros

‘I have many books.’

(4) Tengo un libro en la mano.

‘I’m holding a book in my hand.’

If we turn from the English, and similar languages with a specialized verb for possession, to an examination of Chinese in contrast to Hebrew, we find the following: that while Hebrew uses yesh for both existence and possession, and Chinese
uses יָוּע in its expression of the same two concepts, there is a decided difference between the semantic/logical meaning to be assigned to the respective morphemes in these two languages. Whereas in Hebrew, 'A exists' and 'A is to B' (= 'B has A'), in Chinese 'the world has A' (= 'A exists') and 'B has A'. So that we might decide to gloss יֶש as 'is' and יָוּע as 'has'.

(5) יֶש אֶלוהִים בְּיוֹסֵרָאֵל

is god in-Israel

'There is a god in Israel.'

(6) סִיַּגְי-סָנָג יָוּע רֶנ.

world has person.

'There is a man.'

All of the above distinctions in modes of expressing these relations are operative in the framework of the internal semantics of the languages discussed. They are not necessarily available to speakers through introspection. In fact, since these logical metaphors developed in the distant past of each language, and since these are extremely bleached, very grammaticalized elements of the language, they are in danger of being reinterpreted away. (So, for instance, colloquial English and standard French do in fact use objective case for the second element of an equative sentence to the first, although their Indo-European forerunner did not. Colloquial Hebrew treats the possessed as an object, and Chinese is tending toward the introduction of a locative element for the impersonal possessor in statements of existence.) I am not claiming absolute universality or immediate salience to speakers. The discussion of these logical metaphors is relevant to later chapters in this dissertation to the extent that it sheds light on the grammatical environment in which copulas are found at the time when they develop from more lexical material. I am less interested in what later happens to them when they cease to be used as equative. Thus, I am using the
older, less colloquial forms in the languages under consideration, even though I am aware of the equally strong tendency to move away from an equative paradigm long after a copula has grammaticalized from a less bleached element.

If we use a container metaphor to explain the Chinese example, we might think of 你 as implying that B contains A, in the expression B 你 A. A different Chinese verb, 招 ‘to be located at’, expresses the opposite relation, that ‘B is contained in A.’ But when containment is not at issue, Chinese will use 之.

Without resorting to a formal model of case, we can see that the relations described by the three verbs above are different, that the first two involve some sort of hierarchy and that the third does not. (Which is not to say that class membership doesn’t involve logical hierarchy, merely that the language does not reflect this hierarchy morphosyntactically, in that identity and class membership are not distinguished.)

Examination of the French may help to clarify the difference between the Hebrew model of existence assertions and the Chinese. While il y’a is admittedly an idiom whose component parts are bleached and fused, and not necessarily accessible to the speaker as separate elements, we will proceed to segment analytically, in order to understand the relational mechanism employed by French to express existence.

(7) Il y’a un homme.
he there has a man
‘There is a man.’
The pronoun at the beginning of the sentence is a vestige referentially -- but not grammatically. That is, no one imagines a person or entity that 'has a man'. But the verb *avoir* agrees with the pronoun, and the pronoun serves as the subject of the sentence. Even though case is not morphologically marked on French nouns, the grammatical object of the sentence is indisputably *un homme*. Grammatically, *avoir* is a transitive verb, and the thing that exists (here, a man) is the object of that verb.

Not so in Hebrew. The subject of *jeś* is always the thing that exists. This is evident not from verbal agreement, since *jeś* does not normally vary, but from the fact that the direct object marker is never placed before the noun whose existence is in question.

The *il* in French is functionally distinguishable from the *there* in English. The use of there, as opposed to some other word, is influenced by the need to provide a location for existence and is analogous to the *y* in French. That is its semantic function. While some of the semantic content of there is bleached, much of the resonating content remains. It fills the same function as the Chinese requirement of providing location in every statement of existence. So that on the semantic level there is not completely bleached. The syntactic reason for its obligatory inclusion in the phrase is that English does not tolerate verb-initial statements. In French the placement of *il* at the head of the phrase was similarly motivated. But here the analogy ends. In English, A is the subject of 'there is A'. In French, A is the object. *Il* is semantically a bleached, but syntactically an important element. The inclusion of *there* is syntactically motivated by word order constraints, though its semantics still resonates for speakers.

Two groupings emerge in the use of copular grams: possession is linked to existence, and languages vary in determining which of the two concepts is primary.
On the other hand, identity and class membership are almost universally equivalent in form. English is the only language in Figure 1 that has a completely distinct form for expressions of possession, while retaining the same form for identity and existence.

A separate issue entirely is one of hierarchical ranking or lack thereof. Equivalence implies no hierarchy in any of the languages included here. Possession does, and when possession is linked to existence, the question arises: which is primary, the possessor or the possessed?

Even though Chinese has no case marking and its syntactic subject and objects are much closer to topics and comments, its use of *yīwèi*, rather than *shì*, places it in the group of languages that consider the participant whose existence is being stated to be peripheral to some other participant. That is, the statements of existence and possession are viewed from the perspective of a putative possessor or container. When we say that 'the world has a man', our framework is in reference to the world, not the man. When a language uses this strategy, it distances statements of existence and possession from those of identity and class membership. In our table of ten languages the following members make this distinction: Chinese, English (only for possession, not existence), French, Korean and Vietnamese. The following languages do not make the distinction between possession/existence and identity/class membership: English (for existence only), Finnish, Hebrew, Hungarian, Russian and Turkish.

This leads us to an interesting discovery concerning the symmetrical relationship which appears to hold universally, where identity and class membership are concerned. Never is *A* hierarchically superordinate to *B*, or *B* to *A*, when the statement means essentially 'A is B' or 'A is a B.' This is perhaps surprising, especially when you consider that class membership is logically a hierarchical
system. In other words, ‘Spot is a dog’ does not imply ‘any dog is Spot’. So why is this hierarchy not reflected in the morphology of copulas?

I can demonstrate this grammatical fact about class membership even using colloquial English which does appear to put the second element of an identity statement in the objective case. The objective case here does not indicate hierarchical subordination, but rather topicality.

(8) (a) I am him
   (b) He is me.

Observe that what is actually coded by the change in case in (8) is the topic under consideration. The first nominal is the one topicalized and it appears in the nominative. While the logical purport of (8) (a) and (b) is the same, the outward forms are not.

Now watch what happens when we use the same morphology to express class membership.

(9) (a) He is a dog.
   (b) A dog is he.
   (c) * A dog is him

One might suppose that the mechanism of case would disambiguate class membership statements for the purposes of type-token hierarchy. It does not; neither does word order. Both the token and the type are in the nominative, and to express it otherwise is ungrammatical even in colloquial English.

In fact, (9) (c) could be grammatical, but only when the concept ‘dog’ is topicalized, as in response to the question ‘What’s a dog?’. The objective case codes lower topicality, not membership in a subclass. Case and word order in Modern Colloquial American English both serve to indicate topic, not role.
Word order does not distinguish type from token. A sentence such as (10)(a) does not mean that the first nominal (a mighty fortress) is a token of the type 'our God'. It means (albeit metaphorically) the opposite. Similarly, with the appropriate intonation, (10) (b) is an example of a statement of class membership where the type appears first. Notice the intonational emphasis on the first item.

(10)

(a) A mighty fortress is our God.
(b) A dog is Spot.
(c) A good boy am I.

While Modern American English word order is relatively fixed, historically English was much more flexible, and there may still be speakers in various parts of the world who are fluent in more archaic registers that allow such flexibility.

What did initial position mean for earlier speakers of English? One answer is topic. Likewise, modern speakers who use objective case for the second nominal in an identity are marking the difference between topic and comment, rather than type and token.

This is not meant to suggest that languages do not distinguish between statements of class membership and identity. They simply do not do so by means of separate copular morphemes, nor do they use case for this purpose.

As it turns out, definiteness is much more important in determining which is the type and which the token. Generally speaking, the token is definite, while the type is indefinite. (See §2.2.4) But in terms of case hierarchy, both sides of the copular equation are on an equal footing, indicating that historically the relation that gave rise to the construction was apparently that of identity.
Of primary interest here are those hierarchies which are evidenced by morphological residue that points to the origin of idiomatic copular expressions. Morphology changes more slowly than syntax, and thus it is more salient to a discussion of longterm historical change.

This dissertation is concerned with the origin of copular morphemes. As will be seen in later chapters, when a copula is derived from a pronoun, it is a subject form of the pronoun that is chosen, never one in oblique case. Often when a pronoun is suppletive for case, the subject pronoun is the one that has been derived from the copula.

It is therefore useful to note that the order of A and B is crosslinguistically interchangeable in attributive constructions, and that the source relation is symmetrical, even if the target meaning involves a logical hierarchy.

The conceptual system in question is not solely that of current speakers. Instead, we are interested diachronically in the metaphorical extensions that a language, seen as an entity, employs in order to express statements of identity, class membership, existence and possession. French speakers today may not be aware of the elements of Il y’a, but somebody, sometime, somewhere, must have said it for the first time. And it is the conceptual system of such a person that I consider here.

With this in mind, let us turn once more to the question, why is it that statements of identity and of class membership are coded identically in nearly all the surveyed languages.

The answer appears to be that metaphorically, and with respect to the use of the copula, all class memberships were once conceptualized as forms of identity. In terms of
the container metaphor, this makes a certain amount of intuitive sense. You can take the dog out of the world he inhabits, but you cannot take the canineness out of the dog. The inherent nature of class membership makes its association with identity acceptable.

The tendency, in all ten of our survey languages, to lump together identity and class membership is especially salient here because, as it turns out, when pronouns come to be reanalyzed as copulas, or copulas as pronouns, the reanalysis always takes place in the context of identity or class membership, rather than that of existence or possession.

§2.2.2 Semantic Sources for the Copula: From the Concrete...

When we talk about grammaticalization, we often have in mind the journey of a gram from a very concrete source, such as terms for a body part, to a more abstract, grammatical function, such as a preposition. While this dissertation will explore some alternate routes that grammaticalization can take, it will be helpful to note the sources of the copula that do fit within this traditional view of grammaticalization.

Copulas often come from verbs that describe position. Thus, in Spanish, the innovative verb estar has as its source the Latin stare, meaning ‘to stand’.

Tarahumara, a Southern Uto-Aztecan language spoken in Northern Mexico, normally specifies the position of a participant whose existence or location is being asserted. A participant cannot simply ‘be’: it either stands, lies, sits, squats, swims, flies or flows in its default stance. (Copeland 1995).

The Hebrew verb ‘to be’ הָיָה, haja/hawa appears to have been derived from הרָה, xaja/xawa, meaning ‘to live’. It is a short step from living to merely existing.
While such sources for a copula naturally arise in the process of deriving a new paradigm for the expression of equivalence or possession, in time, worn out copulas can become even more abstract, to the point that their function can be viewed as purely grammatical.

§2.2.3 Semantic Extensions of the Copula: ... to the Abstract.

The story of how copulas become auxiliary verbs is well known to anyone familiar with Indo-European languages.

A nominalized form of an active verb, a participle, for instance, is coupled with a finite form of a verb of existence or possession (to be or to have, être or avoir, sein or haben). In time this combination is interpreted as a new tense, with the copula bearing the grammatical agreement, while the participle carries the semantic load.

(11) He is judging.

In sentence (11) the ing of judging was originally a nominalizing element that later became reinterpreted as part of a verbal tense. This is significant to our inquiries into the copula in the following way: when a language doesn’t have an overt copula in the present tense or non-perfective aspect, one readily available element that may be recruited to serve this purpose is the pronoun.

(12) הָשָׁפֵט
he judge
‘he’s a judge’ or ‘he is judging’
§2.2.4 Conclusions on Copular Morphology

We have seen that the patterns of use of the copula vary little from one language to the next in our sample group, and that the variations that can be identified do not correlate closely with morphosyntactic typology or genetic grouping. We have learned that:

(a) Assertions of existence are often expressed by the same mechanism as expressions of possession, (nine out of the ten languages in Table 1, with the notable exception of English),

and

(b) Existence is patterned as a special case of possession, where B is primary over A, whether this primacy is expressed by morphological case or other relational means, (e.g. in Chinese, French, and Vietnamese.),

but

(c) Possession is patterned as a special case of existence, where A is primary over B (e.g. Finnish, Hebrew, Hungarian, Korean, Russian, Turkish.),

and

(d) Equivalence of two nominals is a logical metaphor that is readily extended to class membership (and disambiguated by definiteness marking).

Above all, we have seen that the copula is not so much an event-verb, as the barest relational expression between participants. Even when grammatically coded as a verb, the copula is the least verb-like of all verbs. It participates in the assertion of a stative situation, being essentially a linker, and in this way it is similar to other conjunctural elements.
§2.2.5 The Cognitive Framework of the Copula

The previous discussion was primarily morphological. Its significance for this dissertation lies in the fact that morphemes of identity and class membership are often derived from third person demonstratives (Chapters 3, 4, and 5). We will also see that third person demonstratives sometimes derive from a copular lexeme (Chapters 6 and 7).

While actual synchronic and universal features of the cognitive makeup of the copula are not particularly significant for purposes of tracking the origin of identity morphemes from more concrete sources, they become much more salient when we wish to understand how copulas can be reinterpreted as demonstratives.

My reasoning in this regard is that the conceptual and cognitive structure of the source construction is more significant in historical terms than that of the target. The target's salient features are sculpted by both its historical origins and its current linguistic environment. Later chapters will explore in detail the effect of typological change on already existing constructions. This chapter is more concerned with the sorts of constructions that are recruited to serve a particular communicative function.

Thus, while in §2.2.1 we were taking a historical approach, focusing on the morpheme by morpheme source interpretation of a copular expression and suppressing its synchronic function as a fused whole, in this section, we will take a holistic, cognitive view of the copula, as a source for idiomatic expressions of personal deixis.
Langacker (1991:II.65) starts out his discussion of the cognitive basis of the English *be* by noting that it is not merely relational. "*Be* is schematic for the class of imperfective processes: it profiles the continuation through time of a stable situation characterized only as a stative relation; it is a true verb, all of whose component states are construed as being identical, but apart from their being relational, it is maximally unspecific concerning their nature."

Langacker then describes predicate nominative constructions as being those in which the trajector and the landmark are identical (1991:II.65). Thereupon he turns to the distinction between generic and specific statements, as in (13) below. (Langacker 1991:II.69)

(13) (a) A wombat is a mammal.
    (b) The okapi is a mammal.

In (13) (a) "the profiled relationship is one of identity between an arbitrary instance of the *wombat* category, ... serving as trajector, and an arbitrary instance of the *mammal* category, ... serving as a landmark." (1991:II.70). By contrast, in 10 (b) the "typespace is construed as the domain of instantiation, with the consequence that type specifications count as instances of the *mammal* category. The identity relation holds between two such types: its trajector is the specific type *okapi*;... its landmark is an arbitrary instance in typespace of the *mammal* category, which inherits the name of the superordinate type and takes the article *a* because it is not unique."(1991:II.71)

Thus the problem of having the same morpheme serve as both an identity marker and a marker of class membership is resolved in Cognitive Grammar by allowing
nominals to sometimes stand for specific types, and at other times to represent entire conceptual categories.

This is an acceptable resolution of the issue on the cognitive level. It was not considered relevant to our earlier discussion of morphological sources for the copula, since our focus was on the common element representing the identity or class membership. (Possession and existence, where the same morpheme is used, are likewise additionally distinguishable by marking on the nominals.)

While the definiteness markers a and the morphologically disambiguate (13)(a) from (13)(b), they do nothing for (14) (a) and (b).

(14) (a) A good animal is a caged animal. [Identity of classes]

(b) A good dessert is a sorbet. [superclass:subclass]

We can resolve this problem by having recourse to two different schemas:

(15) (a) An X Y is a Z Y [Identity of classes]

(b) An X Y is a Q. [superclass:subclass]

Thus in (15), X and Z are adjectives, Y and Q are a nouns. If a speaker recognizes the pattern in (15) (a), he will interpret it as an identity. The pattern in (15) (b) triggers a logical hierarchy, with the second nominal phrase being subsumed in the class represented by the first nominal phrase.

Since this is a discussion of the cognitive scope of the copula, such devices are perfectly acceptable in order to describe what speakers do to make sense of their language as it is synchronically constituted. This resolution would be outside the scope of a crosslinguistic review of copular morphemes and their historical sources.

We turn now from the cognitive status of identity constructions to that of
constructions for asserting existence. The difference between languages where existence is expressed as a special case of possession (Chinese, French) and those where possession is expressed as a special case of existence (Hebrew, Hungarian, Russian, Turkish) can be explained in terms of a container metaphor. In the one case the possessor is the container and the possessed is the contained (Chinese, French), and in the other, the contained is the possessor and the container is the possessed (Hebrew, Hungarian, Russian, Turkish). See Figure 3, where the former group is represented by French, the latter by Hebrew and English shows its own unique pattern, outside both groups.

<table>
<thead>
<tr>
<th>EXISTENCE</th>
<th>FRENCH</th>
<th>HEBREW</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSSESSION</td>
<td>Il y’a A</td>
<td>ješ A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EXISTENCE</th>
<th>There is B</th>
<th>There is A</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSSESSION</td>
<td>B has A</td>
<td>A has B</td>
</tr>
</tbody>
</table>

**Figure 3**

In English, we have a mixed system, because in both the case of existence and possession, the participant that exists or possesses is the focus subject, while the possessed can never have that function (using the same basic morphology).

(16) (a) I had a dog.

(b) *A dog was had by me.
While all of the above exemplifies the distinction between a historical morphologically based discussion and one that is cognitive and synchronic, the area in which cognitive analysis of the copula is most necessary for this dissertation is its application to statements of existence. I return to Langacker’s first description the verb *be*: “...it profiles the **continuation through time of a stable situation** characterized only as a stative relation ...” (Langacker 1991:II.65).

Stability and continuation through time is the hallmark of an entity. When we divide the world into ‘actions’ and ‘entities’, ‘entities’ are those events that continue in a similar state over time in relative stability. Conceptual ‘entities’ are in turn generally coded as nouns.

By predicating existence, we are implying an entity. Since a copula of existence is “maximally unspecific concerning [the] nature” of the states which are being described, it is an excellent candidate for recruitment as a maximally bleached nominal: a demonstrative.

The significance of the above will become more clear in Chapters 6 and 7. For the time being, we will suspend our study of the copula, and move on to examine the pronoun.

§2.3 Pronouns and Demonstratives: The Pronominal System

To understand the change from pronoun to copula, we must examine the conceptual grounding of the pronominal category as a source domain.

The parameters generally available to pronouns are as described below. There are three basic persons: the speaker, the hearer and others. Some languages have a
fourth person, composed of either an amalgam of first and second (as in first person inclusive or a dual, referring to the speaker and hearer 'we two') or split in the third person (along the lines of proximal and distal, or present and absent.)

<table>
<thead>
<tr>
<th>1st</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCLUSIVE (1+2)</td>
</tr>
<tr>
<td>2nd</td>
</tr>
<tr>
<td>3rd</td>
</tr>
<tr>
<td>proximal/distal</td>
</tr>
</tbody>
</table>

**Figure 4**

Even in those languages that display a fourth person, the three persons are more basic and generally are more entrenched in the morphology of the language. There appears to be a cognitive connection between the three persons and the three levels of deixis present in a majority of the world's languages.

The morphology of person marking can involve independent pronouns and/or affixes on the verb or on nouns, usually marking a genitive relation. Pronominal morphemes can additionally be marked for number, gender, animacy, deixis (proximal/medial/distal), case and focus. Pronouns may also be subdivided as personal, interrogative or reflexive.

Pangasinan, an Austronesian language spoken in the Philippines, has two basic sets of pronominal agreement, for focused and unfocused participants. (Fieldwork with Manuel Datuin 1993-1994.)
<table>
<thead>
<tr>
<th></th>
<th>SINGULAR</th>
<th>DUAL</th>
<th>PLURAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ST</td>
<td>ak</td>
<td>ta</td>
<td>tayo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>kami</td>
</tr>
<tr>
<td>2ND</td>
<td>ka</td>
<td></td>
<td>kayo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>kayora</td>
</tr>
<tr>
<td>3RD</td>
<td>□</td>
<td></td>
<td>ira</td>
</tr>
</tbody>
</table>

**Figure 5**
PANGASINAN FOCUSED PRONOMINAL AGREEMENT

<table>
<thead>
<tr>
<th></th>
<th>SINGULAR</th>
<th>DUAL</th>
<th>PLURAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ST</td>
<td>ko (-k)</td>
<td>ta</td>
<td>tayo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mi</td>
</tr>
<tr>
<td>2ND</td>
<td>mo (-m)</td>
<td></td>
<td>yo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>yora</td>
</tr>
<tr>
<td>3RD</td>
<td>to</td>
<td></td>
<td>da/ra</td>
</tr>
</tbody>
</table>

**Figure 6**
PANGASINAN UNFOCUSED PRONOMINAL AGREEMENT
Verbal affixes determine which participant in a Pangasinan sentence is focused, and pronominal clitics agree with the verb in terms of focus.

(17) manga-kan ak na mangga
    AGFOC-eat 1st UNFOC mango
    ‘I’m eating the mango’

In (17) the focus is on the agent, and it would be an appropriate response to the query ‘What are you doing?’ In (18) the focus is on the patient, and it would be an appropriate response to ‘What are you eating?’

(18) ka-kan-en ko -i mangga
    PATFOC-eat-IMPERF 1st FOC mango
    ‘I’m eating the mango’

In (19) we have a causative, so though we have patient focus, the focus is actually on the causee, not the patient.

(19) pa-pa-kan-en ko -i anak ko
    CAUS.-PATFOC-eat-IMPERF> 1st FOC child 1st
    ‘I’m feeding my child’

The dual category in Pangasinan can be interpreted as a question of person rather than number. That is, it can be seen as a combination of first and second person, where there is only one of each: ‘we two’.
(20) man-tongtong ta nabwas

IMPERF talk 1st dual tomorrow

'We'll talk tomorrow'

The statement in (20) can be distinguished from (21), where first and second person are indicated by separate pronouns.

(21) mi-tongtong ak ed sika nabwas

AGFOC-talk 1st OBLIQUE 2nd tomorrow

'I'll talk to you tomorrow'

Note that second person in (21) is indicated by an independent pronoun, preceded by an oblique marker.

Figure 4 below is a paradigm of the independent pronouns. They are, for the most part, naturally predicted by the following formula: si + the appropriate dependent pronominal element from Figure 6. However, absolute regularity is not maintained. For instance, third person singular would be 'si' (si + 0) under the above formula, where in fact it is 'sikato'.
<table>
<thead>
<tr>
<th></th>
<th>SINGULAR</th>
<th>DUAL</th>
<th>INCLUSIVE</th>
<th>EXCLUSIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ST</td>
<td>siak</td>
<td>sikitandwa</td>
<td>sikatayo</td>
<td>sikami</td>
</tr>
<tr>
<td>2ND</td>
<td>sika</td>
<td>siyandwa</td>
<td>sikayo *</td>
<td>sikayora*</td>
</tr>
<tr>
<td>3RD</td>
<td>sikato</td>
<td>sikaranwa</td>
<td>sikara</td>
<td>sikara</td>
</tr>
</tbody>
</table>

Figure 7
PANGASINAN INDEPENDENT PRONOUNS

Pangasinan is an example of a language where focus (which can be seen as a type of case marking) is expressed by suppletion. Indo-European languages use suppletion for case marking on pronouns, as do Semitic languages. In Turkish, however, the agglutinative principle leads to a different result. Pronominal stems, marked for person, are then affixed with number and case markers. Case marking is almost invariant as to person, save for euphonic rules of vowel harmony.

<table>
<thead>
<tr>
<th>SINGULAR</th>
<th>FIRST</th>
<th>SECOND</th>
<th>THIRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS.</td>
<td>ben</td>
<td>sen</td>
<td>o</td>
</tr>
<tr>
<td>ACC.</td>
<td>beni</td>
<td>seni</td>
<td>onu</td>
</tr>
<tr>
<td>GEN.</td>
<td>benim</td>
<td>senin</td>
<td>onun</td>
</tr>
<tr>
<td>DAT.</td>
<td>bana</td>
<td>sana</td>
<td>ona</td>
</tr>
<tr>
<td>LOC.</td>
<td>bende</td>
<td>sende</td>
<td>onda</td>
</tr>
<tr>
<td>ABL.</td>
<td>benden</td>
<td>senden</td>
<td>ondan</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PLURAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS.</td>
</tr>
<tr>
<td>ACC.</td>
</tr>
<tr>
<td>GEN.</td>
</tr>
<tr>
<td>DAT.</td>
</tr>
<tr>
<td>LOC.</td>
</tr>
<tr>
<td>ABL.</td>
</tr>
</tbody>
</table>

(Lewis 1967:67)
Notice, however, that there is a greater constituency parallelism between first and second person in Turkish than there is in third. First and second person could be analyzed as consisting of a root consonant, \( b \) for first, \( s \) for second, and of number markers, \( en \) for singular, \( iz \) for plural. In contrast, third person divides into \( o(n) \) for the root, and \( lar \) for plural. Singular is zero marked. It is no coincidence that third person bears the plural marking which is productive throughout the language, while first and second are idiosyncratic.

First and second are crosslinguistically the more highly marked persons. In those languages that mark only two persons, they are always first and second, not third.

Dyirbal, spoken in Australia, is such a language. While there are ways to refer to third person, such markers are in fact nouns and do not fit into the pronominal system. The pronouns divide into these categories: (1) speaker, (2) speaker and one other person (3) speaker and more than one other person (4) addressee (5) two persons, at least one an addressee (neither the speaker) (6) more than two persons, at least one an addressee (none the speaker) and, in one dialect, (7) speaker and spouse.

(Dixon 1972.50)

In Semitic, the roots of first and second person subject pronouns come from a common source, which can be reconstructed as \(^*an(t)\). (Moscati 1964.102). Third person is totally unrelated.

The otherness of third person, its later development and its idiosyncratic features, seem to be fairly consistent crosslinguistically. This may properly be attributed to the deictic origins and cognitive underpinnings of the pronominal system.

Langacker (1987:127) in his Cognitive Grammar, lists pronouns as a subcategory of deixis:
There are several distinct types of deictic expressions. One type consists of expressions that specifically designate a ground element. Primary examples are I and you, which designate speech-act participants; others include here and now (when used as nouns -- e.g. The best place is right here, Now is a good time), which refer to the place and time of speaking. Here and now are more commonly used adverbially, in which case the same ground element participates in a relational predication.

So close are deixis and predication that languages tend to use deixis as a source for assertions of existence. So for instance, in Biblical Hebrew, possessive pronominal clitics, affixed to a deictic meaning roughly 'here' were used first to indicate spatial and temporal location, and later to express existence or as auxiliaries with participles. (This is not the primary mode of expressing existence in Hebrew, and this is not the copular morpheme that we will study in detail in later chapters.

(22) Δje-ka? Hinē-ni.  Здесь? тут
where-2ndsg Here-1stsg
‘Where are you? Here I am.’

(23) Hinani ha-melex.  Здесь король.
here-1stsg the-king
‘I’m the king now.’

(24) Hinani muxan.  Здесь готов.
here-1stsg ready
‘I’m ready.’
Gildea (1993) describes how pronominal deixis in Panare developed into tensed forms of the copula. Proximal deixis became present or immediate future, while distal became past.

But how do we know whether the pronouns or the deictics come first? Grammaticalization theory in general suggests that we derive the abstract from the concrete. However, deictic function and its close relatives occupy a place in our cognitive systems that is difficult to differentiate with respect to degrees of abstraction. †

What we do know is that eventually distal deixis can lead to third person pronouns which in turn become copulas. The semantic path taken is as follows:

\[ \text{---\rightarrow DEFINITE ARTICLES \rightarrow SUBORDINATORS} \]
\[ \text{DEIXIS \rightarrow DEMONSTRATIVES \rightarrow} \]
\[ \text{-------\rightarrow PRONOUNS \rightarrow COPULAS} \]

**Figure 9**

† A working hypothesis as to the origin of the pronominal system would be to link deixis directly to pronouns. There are three degrees of deixis commonly found in many of the world's languages (and only two in many others.) Where there are two degrees of deixis, we might hypothesize that the result would be a pronominal system in which proximal is first person and distal is second, leaving third person unrepresented. If there are three degrees, then first person derives from proximal, second from medial and third from distal.

1. HERE [HITHER] \rightarrow THIS \rightarrow (I/THOU?)
2. THERE [THITHER] \rightarrow THAT \rightarrow (YOU?)
3. YON(DER) \rightarrow YON \rightarrow THAT ONE \rightarrow HE \rightarrow IS
The introduction of a distal participant, possibly out of sight, implies a predication. We are introducing a participant outside the speech act, which in Langacker’s terms means that we are making a maximally objective statement. To say ‘that’ implies ‘is’. The issue of existence or identity does not normally come up in first and second person, because it is demonstrable situationally as exophora.

A copula is “maximally non-specific and helps indicate a state that endures over time.” A pronoun is maximally non-specific and indicates a participant, whose existence is presumed, but may not be apparent. When in need of a new demonstrative, speakers may recruit a form of the copula. When searching for a new copula, speakers may recruit a demonstrative. Sometimes the meanings and functions of the two are so close that they are difficult to tell apart. This is a situation that leads to reanalysis. It also offers some difficulty for the meta-linguistic analyst.

§2.4 Distinguishing Pronouns from Copulas

If pronouns and copulas are so close in their meanings and functions, what allows us to label a morpheme as one and not the other?

The heuristic that I adopt here is that to the extent that a deictic/stative morpheme patterns with the nominals in a language, it is a pronoun or demonstrative. To the extent that it patterns with the verbs, it is a copula. The patterning to which I refer can be morphological, syntactic or morphosyntactic.

So in English, we know that he is a pronoun and is is a form of the copula by their syntactic patterns of use.
(25)

(a) He is a mailman.
(b) The mailman is tall.
(c) Is he coming?
(d) He called.

Because *is* patterns with verbs in (25) (a) and (b) as SVO constructions, and in (25) (c) as a VSO question, we deem it to be a copula. Because *he* patterns with nouns in an SVO interpretation in (25) (a) and (d), and with VS in (c), we deem it to be a pronoun.

Not every case will be as clearcut. We will address the issue again as it arises in the following chapters.

§2.5 Conclusion

When a copula develops from a third person pronoun or demonstrative, it usually does so from a distal variant, not a proximal or medial. This holds true for the instances provided in this dissertation: Modern Chinese, Modern Hebrew and Finnish (as deduced from Finno-Ugric morphology.)

When a demonstrative is derived from a copula, the resulting form is used for the distal variant, as in Turkish and Biblical Hebrew (infra).

The historical connection between distal demonstratives and copulas is grounded in the cognitive link between third person and objectivity.

First and second person, referring as they do to speech act participants, are maximally contextual. Third person, referring to other, possibly unseen participants, is less contextually dependent. It therefore signifies a greater degree of abstraction.
Historically, third person marking develops later in the grammaticalization cycle, and from a different source, than the marking of first and second person. There are languages without third person pronouns or third person pronominal morphology. There are none without first and second person. Zero anaphora in third person is a crosslinguistically common phenomenon.

The copula is, like third person, often a zero marked construction. We deduce the presence of the copula by the absence of anything else. Semantically, it represents the barest relationship between participants.

We can have a zero marked relation between participants and we can have a zero marked participant, but it is difficult to have both at the same time and still carry the functional load of a predicative statement.

Copulas and third person pronouns code the same basic concept (‘existence in time and space’) under different grammatical guises, the one more verbally, the other more nominally. When the surrounding linguistic material regroups, the same morpheme that coded ‘a being’ may be used to express the act of ‘being’. And vice versa.
Chapter 3. Classical Demonstrative to Modern Copula: Chinese shi.

§3.1 Introduction

In this chapter we will trace the development of the Modern Chinese copula from a third person demonstrative pronoun.

Here I am relying heavily on the work of scholars who have charted this reanalysis in Chinese (Chen 1994, 1995, Li 1977, Sian Yen 1986). I make no claim to originality in my treatment of this well studied phenomenon. The reason for the inclusion of the history of shi in this dissertation is that the material is on point, and the Chinese example provides an interesting addition to the catalog of known instances of this phenomenon.

Typologically Chinese is very different from the languages examined in the following chapters. One of the chief surprises in dealing with Oracle Bone and Classical texts was, for me, the discovery that early Chinese syntax patterns with languages that have a much richer morphology. Like Biblical Hebrew, Classical Chinese begins with a zero copula and then annexes a pronoun to serve as an overt equative linker. The evolution from Classical Chinese to Mandarin parallels changes from Biblical Hebrew to Modern: word order becomes less flexible, and sentences grow longer, both in number of units and number of syllables. Another significant development is what I interpret as increasing preference for nominal coding of stative or descriptive events.

The Chinese example allows us to study the relation between overall grammatical structure and the evolution from pronoun to copula.
§3.1.1 The Guises of shì

The particle shì, represented by the same character and presumably having the same phonological realization, was historically used as a demonstrative pronoun, an emphatic particle, and an affirmative expression. In Modern Mandarin, the emphatic and affirmative meanings remain, while the demonstrative use has been replaced by an equative or assertive copular function. (Chen 1995).

When used as a modern copula, shì can express identity and class membership. Existence, possession, and location employ different morphology. (See Chapter Two for a brief discussion.) In addition to its assertive or declarative function, modern shì is used for emphasis, reiteration, and even agreement or affirmation.

While the assertive function of the Modern Chinese 'A shì B' was carried out by the template 'A B ye' in ancient times, the contrastive or emphatic function of Modern shì was expressed by the particle wei. Thus wei meant 'is so' to fei's 'is not' (or 'true' versus 'false'). (Dobson 1962:100) In Modern Mandarin, the opposite of fei is characteristically shì. (Chen 1995)

Not all statements of class membership can be coded by shì in Modern Chinese. Predicate adjectives normally require no copula, being highly verbal in themselves. Only a nominalized predicate adjective will take shì as a matter of course.

In the following sections, we will plot the evolution of copular shì from its earlier demonstrative function. As we do so, we will also consider its emphatic and affirmative uses, and will discuss their place in the semantics of assertion.

But before we turn to the subject of the copula, we will take a brief look at overall trends in the development of Modern Mandarin from its ancient precursor.
§3.2 Classical Chinese to Modern Mandarin: General Trends

Pre-Classical Chinese displays a remarkably flexible syntax, despite the fact that it has very little morphological marking to fix morphosyntactic roles and grammatical category. Pragmatics disambiguates many utterances, and the nominal and verbal nature of participants is determined by discourse context more than by word order. As we move from the Classical language to Modern Chinese, we find that the number of morphemes in an utterance increases, that order is relatively fixed for main verbs, and that what was previously expressed verbally is coded by means of a nominal and a verbal element, which sometimes virtually lexicalize into a disyllabic verb.

In (1) below, (a) and (b) come from Oracle Bone inscriptions, (c) is Classical Chinese, while (d), (e), (f) and (g) are Modern Mandarin. (Some of the determinations of noun verb status are my own, and it is possible to argue for a different label.)

(1)
(a) (OBC): VERB
yu?
rain
‘Will it rain?’
(b) (OBC): PARTICLE VERB
qi yu?
PRT rain
‘Will it rain?’
(c) (CC): VERB PARTICLE
   yu hu?
   rain PRT
   'Will it rain?'

(d) (MM): VERB NOUN PARTICLE
   xia yu ma?
   fall rain PRT
   'Will it rain?'

(e) (MM): VERB VERB NOUN PARTICLE
   hui xia yu ma
   can fall rain PRT
   'Will it rain?'

(f) (MM): VERB NOUN NEGATION VERB NOUN
   xia yu bu xia yu
   fall rain not fall rain
   'Will it rain?'

(g) (MM) VERB NEGATION VERB VERB NOUN
   hui bu hui xia yu
   can not can fall rain
   'Will it rain?'

(Chen 1995)
Verbal affiliation is not lexically determined in Chinese. Just as in English, 'rain' can be a verb or a noun. In (2) (a) below we recognize rains as a verb, because of its third person agreement with the pronoun, and we know this to be a question because of the inverted VS order. In 2(b) the auxiliary carries the agreement and rain is the main verb, uninflected. In 2(c) we know rain to be a noun, because it is preceded by an article and also because it fills the nominal slot in the VS question template.

(2)

(a) Rains it (in Spain)? (Archaic)

VERB(3rd sg) SUBJ(3rd sg)

(b) Does it rain (in Spain)?

AUX SUBJ(3rd sg) VERB (inf.)

(c) Does the rain in Spain fall mainly in the plain?

AUX ART NOUN PREP NOUN VERB (inf) ADVERB PREP ART NOUN

But while in English verbal or nominal status is determined by a combination of morphological and syntactic cues, in pre-Classical and Classical Chinese there was scarcely any morphology and comparatively little syntax to make the determination.

For instance, while SVO is said to be the word order of Chinese from pre-Classical times onward, the understanding of SUBJECT for such an interpretation requires more emphasis on the TOPIC component of the SUBJECT category and less on the AGENT element. So, in 1 (d) above, we have a verb followed by a noun. Under strict SVO interpretation we might suppose it to be a VO construction. Instead, the rain may be seen as the agent of the falling action, not its
patient. Contrast (1) (d) with 3(a) below, where the rain is in fact the patient in postverbal position.

In 3(a) the king will perhaps not meet the rain. In (3)(b) the king will perhaps not be wet by the rain. Bei is a marker that expresses "that the grammatical subject is the experiencer of the action" (Chen 1995). So in 3(a) the rain is the grammatical object (patient), while in 3(b), it is the demoted subject (agent).

(3)

(a) (OBC): 

NOUN NEGATION VERB NOUN

Wang bu gou yu

king not meet rain

'The king perhaps will not encounter rain?'

(b) (MM): 

NOUN NEGATION AUX ADVERSATIVE NOUN VERB PARTICLE

Wang bu hui bei yu lin ba?

king not would ADVERSATIVE rain wet PRT

'The king perhaps will not encounter rain?'

Later Chinese, such as Mandarin, is distinguishable from the Classical language in the following respects: utterances are longer in comparison to their Classical counterparts, and the role of pragmatics is often taken over by formal syntactic templates and greater use of prepositions and other grammatical markers.

The above cursory description was for the purpose of understanding the environment in which renalaysis of the copula took place in Chinese. Having set the stage, we now turn to examine equative utterances in the Classical language.
§3.2.1 The Particle *ye*

The Classical Chinese equative construction was of the model 'A B ye'. Example (4) below is an instantiation of this pattern.

(4) Changshan, bei yue ye.

Mt. Chang, north mountain PRT-ye.

'Mt. Chang is a/the northern mountain.'

(shuo Yuan, 1st century B.C. in Chen 1995)

The grammatical label that we place upon the particle *ye* is largely influenced by the meta-linguistic context of our inquiry. Chinese grammarians view this as a particle, not a verb. Under such an interpretation, we have a zero copula, with *ye* merely serving to signal the end of the utterance.

Since verbs are not identified morphologically in Chinese, and since most Classical Chinese words are monosyllabic, word order and semantic factors are the best way to determine the grammatical status of this form.

Classical Chinese has two basic sentence templates. The first is 'A VERBAL B.' The two nominal elements in (5), *Sung* and *fu jen*, respectively, are on either side of the verbal element *you*.

(5) Sung you fu jen

*Sung* have rich man

'There was a rich man in Sung'

(Graham 1967:6)

The second template is 'A B PARTICLE'. Note that in (6), the nominal elements are contiguous, followed by the particle.
(7) Yu jen ye.

I man PART.

‘I am a man’

(Graham 1967:8)

Attributes in Classical Chinese are considered to be stative verbs, and are not nominals at all. Graham (1967:12) makes the following crosslinguistic comparison:

Indo-European languages which inflect the adjective decline it like a noun; some other languages, for example Japanese ... conjugate it like a verb ... In Chinese ... words translatable by adjectives count as verbs; they are preceded, for example, by negatives and temporal particles which precede verbs...

We have already seen that if we want to say ‘rich man’ in Classical Chinese, the adjective will precede the noun, or stated more broadly, the attribute will precede the participant. However, when the attribute is not part of the nominal clause, the attribute appears after the nominal it describes.

(8) Jen fu.

man rich

‘The/a men/man is/are rich’

This is not very different from the adjective/noun patterns displayed in languages without an overt copula in present tense, Turkish and Hebrew among
them. In such languages, 'ATTRIBUTE NOMINAL' is a phrase, meaning 'the
such-like thing' but 'NOMINAL ATTRIBUTE' is a complete thought, an assertion of
attribution, meaning 'the thing is such-like.'

If attributes are in fact members of a verb-like category, then we might consider them to be stative or intransitive verbs, which apppear sentence finally. In such a case, transitive sentences in Classical Chinese are primarily of the pattern SVO, while intransitive sentences are of the pattern SV.

In example (9) we examine another sentence of the template 'A B PARTICLE.' This time the particle in question is yi, one of whose functions corresponds to modern perfective aspect.

(9) ...tzu fan erh jen yi ...

self examine CONJ benevolent PART.

'(if) he examines himself and he (turns out to be) benevolent."

(Graham 1967:12)

While yi can be a perfective aspect marker, among other uses, ye may denote assertion of ongoing actions or states.

(10) Wo pi pu jen ye ...

I NEG NEG benevolent PART.

'It must be that I am not benevolent...'

From the above analysis it would seem that the particle ye is not a verb in itself, but that it codes some aspect of verbal semantics, such as would be carried by verbal
morphology in an agglutinative language. Compare (9) above with the Turkish sentence in (11).

(11) (O) Sorumluy-du
(he) responsible + di-past

‘He was responsible.’

Clause final aspect marking which occurs on an attributive verb is typical of SOV languages. Granted, ye is not so firmly established as a grammatical marker as the Turkish affix in (11). Chinese particles often serve to indicate the speaker’s attitude and are intermediate between discourse markers and grammatical morphemes. Still, in its day ye was often used to mark the end of an assertion. While it did not indicate pure tense or aspect as such, it indicated the continuation in space and time of the state asserted.

The typological significance of the above investigation of the particle ye is as follows: the equative relation between two participants, A and B, was expressed in Classical Chinese by the sequence ‘A B’ followed by an assertive particle. Thus the B position in such a construction was a little closer to being verb-like than the A position. In other words, B is the predicate.

All of this has changed drastically by the time we get to Modern Chinese, where the corresponding sequence is ‘A COPULA B’ -- in which both A and B are equally nominal. We tend, however, to feel that A is primary, since in transitive sentences, B is the object. This primacy may be expressed in terms of topic rather than participant hierarchy, in which case it does not tell us a great deal about who is doing what to whom.

Another way to express the difference between the two templates is as follows:
CLASSICAL: A B PARTICLE $\implies$ A Bs (does B) $\implies$ A is a B

example: John lawyer ye $\implies$ John lawyers $\implies$ John is a lawyer

MODERN: A COP B $\implies$ A equals B $\implies$ A is B

example: John shi lawyer $\implies$ John equals lawyer $\implies$ John is a lawyer.

Viewed from this perspective, the Classical Chinese pattern for equative sentences parallels a language-wide preference for expressing simple statements verbally. †

Compare (1) (c) and (d) above, where rain is a verb in the Classical Chinese example, but a noun in Mandarin.

§3.2.2 The Demonstrative shì

Classical Chinese had a prodigious number of demonstrative pronouns. On the other hand, it had no third person pronoun that was not a demonstrative. While first and second person pronouns were also abundant, their inventory was a great deal more limited, and their affiliation as indicators of person was well established. The third person demonstratives were deictics that coded additional semantic functions, but were not firmly placed within any sort of pronominal paradigm.

This state of affairs seems to imply an earlier stage where third person was the default and did not require overt pronominal marking. However, we should not place too much emphasis on this point, since the inventory of pronominal candidates in Classical Chinese varied from century to century, and text to text. That is to say, not all pronominal morphemes listed below were used by speakers of all stages of the language. (Chen, personal communication.)

†That the modern 'A COP B' is a sentence template wherein the copular morpheme plays a more verbal role is evidenced by the placement of adverbs and negations immediately preceding the copula, just as they would be placed before a verb.
What we can infer from the abundance of pronominal candidates is that this was a system in the process of forming. We can also gather from the relatively smaller number of first and second person candidates that the speaker and hearer positions were more conventionalized than third person. That is to say, they had become more grammaticalized earlier. Each of the third person demonstratives was vying for a favored position as a bleached third person pronoun, but for the time being none had attained ascendancy. Figure 1 below shows the inventory of pronominal contenders.

---

**Figure 1**

(Chen 1994)
In example (13) below, we see 使 used as a demonstrative in Classical Chinese.

(13) Shi xin zu yi wang yi.

DEM heart sufficient to/as king

Literal: ‘That kind of heart is sufficient to be a king.’

(Chen 1995)

There was no apparent limitation on the syntactic position that 使 could occupy. In example (14) it appears in the final position, but is not used as a verb; rather, it is treated as a nominal, even though the gloss seems adverbial (manner).

(14) Dang ru shi.

should resemble DEM

‘Should act like this.’

(Chen 1995)

In many clauses, 使 co-occurred with ye, with the latter performing the equative or assertive function and the former serving as a demonstrative. Observe (15) below.

(15) ... shi ke ren ye...

DEM can bear PRT

‘[if]...this is bearable...’

(Chen 1995)
Syntactically, in Classical Chinese, *shì* could appear anywhere, but the syntactic function of *ye* was far more specialized. We might say that *ye* was more grammatical, being so devoid of lexical meaning that it served only a syntactic function and being so conventionalized in its use that its position in the clause was inflexible. The demonstrative, *shì*, was far more robust, having now a deictic, now a pronominal meaning, with full freedom to appear in many different syntactic positions.

§3.3 The Modern Chinese Copula: *shì*

Example (16) shows the modern reflex of the demonstrative described above, here functioning as a copula, linking two nominals in an equation of ‘A *shì* B’.

(16) Heng shan jiu shì bei yue

Heng Mt. EMPH COP north mountain

‘Mt. Heng is the northern mountain.’

(Chen 1995)

The Modern Chinese *shì* is just as rigid syntactically as Classical *ye*. While *ye* had to be placed at the end of a clause, after the nominals, *shì* must be placed in the middle of a clause, between the nominals.

The change in equative templates between Modern and Classical Chinese is just as significant as the change of equative morphemes. And the two changes may be linked.
§3.4 How (and When) a Demonstrative Became a Copula

In this section we will discuss the motivation behind the change. Competing theories will be compared, and a heuristic for the determination of the function of *shi* as nominal or verbal will be considered.

§3.4.1 Earlier Literature

The reanalysis of *shi* has been explained in a number of different ways. Li and Thompson (1977) suggest that this came about via a ‘topic mechanism’: "...the subject pronoun which is coreferential with the topic in the comment of a topic-comment construction is reanalyzed as a copula morpheme in a subject predicate construction." (Li & Thompson 1977:420)

This explanation means, in essence, that we had a sentence of the template ‘A, he B’ which, in time, with the elimination of the pause, was reinterpreted as ‘A is B’. I agree that the topic-comment mechanism is involved. However, I would suggest greater attention to parallel changes taking place throughout the structure of the language, rather than treating this particular template separately.

Yen (1986) has a radically different theory. "*Shi* came to be used as an affirmative particle, because speakers of the language felt that it would be natural to use it as the antonym of the negative particle *fei*. ... Since the negative *fei* was most commonly used before a nominal predicate to negate it, its affirmative counterpart *shi* naturally began to appear before a nominal predicate, just like a copula." (Yen 1986:228).
Yen seems to assume that the affirmative shî is completely separate and distinct from the demonstrative, and that it must perforce predate the copular use.

That a copula is used as an affirmative is not uncommon. In Hebrew, the normal affirmative response to a question about existence is by a repetition of the existential.

(17) (a) jēš maraq?

is soup

'Is there any soup?'

(b) jēš

is

'Yes.' [There is.]

The use of a copula as an affirmative is a natural byproduct of the emphatic function as seen in (18) below.

(18) (MM)

(a) Ta zuofan.

3sg cook

'He cooks.'

(b) Ta shî zuofan.

3sg COP cook

'He does [indeed] cook.'
(c) **Ta shi zuofan de.**

3sg COP cook NOM

'He is a cook.'

(d) **Shi, shi, shi, ta shi zuofan**

COP COP COP 3sg COP cook

'Yes, yes, yes, he does cook.'

Yen (1986:231–232) argues that the dual function of the demonstrative as a copula prior to its replacement by the new third person pronoun *ta* is proof that reanalysis of the type suggested by Li and Thompson could not have taken place. "Long co-existence of the anaphoric use of *shi* with its ‘copular’ use shows that generation after generation of new speakers of the language did in fact learn the anaphoric function of *shi* after the ‘copular’ use of *shi* had come into existence. This would rule out the possibility that the putative change of the demonstrative *shi* into a copula was due to ... misinterpretation of *shi* by some new generation of speakers."

Co-existence of more than one function of the same morpheme does not disprove the evolution of one use from another. On the contrary, it tends to serve as proof of origin. For instance, in English the verb *do* is used as an emphatic with active verbs, an auxiliary in their negations and, further, as an interrogative particle and an affirmative.

(19)

(a) *I did something today.* [CONCRETE]

(b) *I ate an apple.*
(c) You did not. [NEGATION]
(c) I did eat an apple. [EMPHATIC]
(d) Did you? [INTERROGATIVE].
(e) I did. [AFFIRMATIVE]

That these uses coexist with the literal meaning of the verb to do does not negate the fact that some functions of the morpheme do are considerably more grammaticalized than others.

§3.4.2 Determining the Status of a Morpheme as Pronoun or Copula

Yen’s assertion does raise another issue that will be relevant throughout this study: when two uses coexist, how do we determine that they are separate and distinct? How do we know when shì is being a demonstrative and when it is a pronoun?

Li and Thompson seem to think they have a clear-cut answer to that question. They analyze (20) below as an unequivocal use of shì as a copula. (1977:426).

(20) Yu shì suō jia fu-ren zhi fu ye
    I SHI nominalizer marry woman genitive father YE
‘I am the father of the married woman.’

According to Li and Thompson, the use of shì in (20) does “not permit us to analyze it as a demonstrative; its unequivocal function in [this] sentence... is that of a copula.” While they do not directly say why this interpretation is the only one possible, I believe their reasoning probably runs as follows: (a) There is a mismatch
in persons (1st to third) which obviates an interpretation of shì as anaphora for yu, (b) it doesn’t make sense to put two pronouns with the same referent one next to the other and (c) in another version of this sentence, seven centuries earlier, no shì was included. (See (21) below.)

(21) Yu er suo jia fu-ren zhi fu ye
   I your nominalizer marry woman genitive father YE
   ‘I am the father of your married woman.’

While I agree with Li and Thompson’s analysis that shì is in fact serving as a copula in (20), the arguments in favor of this conclusion are not unassailable. To wit: (a) The mismatch in persons would not preclude an interpretation that (20) is following the ancient ‘A B ye’ pattern. Thus, it would mean essentially: ‘I am that [one] who is the father of the married woman.’ The equative function would be filled by the ye template, while shì would be serving as a demonstrative.

The ambiguity of interpretations available during the period when shì still served a dual function as a demonstrative and a copula, and when the particle ye was still used for equative purposes, is precisely what led to the reanalysis. If there had been no ambiguity, there could not have been a reinterpretation.

Fortunately, by the time of Modern Mandarin shì lost its function as a demonstrative and was replaced by ta, as third person pronoun. The ‘A B ye’ pattern is no longer productive for equative functions. Hence, we know the change came about, we know its origin, and we know that the demonstrative and copular uses of shì, even while they co-existed, can be seen as separate and distinct.
This is a relatively easy case. In the next chapter we will deal with the same dilemma under more difficult circumstances. In Modern Hebrew, both the copular and pronominal uses of the third person pronoun are retained to this very day.

§3.4.3 Reanalysis and the Role of Typology

Chen (1995) suggests that the new equative construction came about through a series of related trends in the language, rather than a single source pattern or a sweeping unilateral reanalysis. Rather than plotting a single path for the evolution, such as Yen's (1983) theory on the affirmation or assertion function of *shi*, Chen (1995) considers a convergence of syntactic and morphological changes, leading to a reanalysis in discourse contexts. The progression in (22) below is an illustration of this process of reanalysis.

\[
\begin{align*}
\text{(22) a. NP/action} & \quad \text{DEM, Predicate PRT} & \text{[DEM is Predicate]} \\
\text{NP/action *shi} & \quad \text{Predicate} & \quad \text{ye} \\
\text{NP} & \quad \text{VP} & \quad \text{[NP is Predicate]}
\end{align*}
\]
According to my own interpretation of this view of the change, the clause boundaries have been reanalyzed, leading from an essentially intransitive view of the equative sentence, with the final verbal-aspectual particle marking the end of the clause, to a transitive view of the assertion, with the copula *shi* standing in between two nominal phrases, the first of which is the subject and the second of which is the verb complement.

This change is part of a language-wide regrouping of nominal and verbal forces. Just as Modern Mandarin speakers prefer to say that 'rain (noun) falls', rather '[it] rains' (verb), so they prefer to state that 'John is a lawyer', rather than saying 'John lawyers'.

§3.5 Conclusions

We have seen that the Classical Chinese third person demonstrative morpheme *shi* was the source of the Modern Chinese copula *shi*. In addition to serving as our first example of this phenomenon, namely, grammaticalization of pronouns into copulas, the history of the Chinese copula gives rise to the following observations:

(1) Reanalysis results from typological change: the functions of the elements of a collocation shift to meet the changing expectations for sentence templates;

(2) Zero copulas occur where SV templates are tolerated, as in Classical Chinese;

(3) An overt copula is required where full SVO sentence templates are preferred, as in Modern Chinese;
and

(4) A demonstrative or personal pronoun is recruited to serve as a copula when sentence typology shifts from primary emphasis on clause boundaries, which in Classical Chinese are marked by clause final particles, to emphasis on clause cohesion, with a central verb and nominals succeeding and following it as in Modern Chinese.

While the particular details will change from language to language, we will show in succeeding chapters that the general tendencies as set forth above apply to languages that are not related and are not normally considered to belong to a similar typology.
Chapter 4. Classical Pronoun to Modern Copula: Hebrew נָה [hu].

§4.1 Introduction

In this chapter we will examine the evolution of the Hebrew third person pronouns into present tense copular morphemes.

This usage of pronouns as copulas in Modern Hebrew is, of course, well documented, and I am not the first to discuss it. Berman (1976), Li and Thompson (1977), Devitt (1992), to name a few, have dealt with this issue. It is such an obvious feature of Modern Hebrew that no grammar of the language can fail to note it. I am not particularly indebted to my predecessors in their treatment of this subject, since my analysis was performed independently, before I was aware of their work. On the other hand, neither can I claim any great innovation in noting the phenomenon.

My contribution in this area of the history of Hebrew, and of grammaticalization theory in general, is that I appear to be the first to have identified the pronouns themselves as reflexes of an earlier form of the verb 'to be'. That discovery will be dealt with in detail in Chapter Seven. This chapter is devoted to the synchronic and diachronic aspects of the use of pronouns as copulas in Modern Hebrew.

§4.2 Functional Marking

Because there is an overlap between pronominal and copular use of what are historically the third person pronouns in Modern Hebrew, special attention to functional differentiation is in order.

§4.2.1 Synchronic Co-existence of Copular and Pronominal Uses

In Modern Hebrew the third person pronominal forms also serve as overt present tense copulas, in addition to their ordinary anaphoric function. Since the uses are concurrent, the grammaticalization process has not progressed to the point where the source of the overt copula is opaque. If one asks a native speaker of modern Hebrew to translate the word נָה [hu] into English without providing a context, the speaker
will reply “he”. However, when asked what הוא [hu] means in example (1) (a), the speaker will reply “is”.

(1)
(a) David hu melej jisra’el
    דוד הוא מלך ישראל
David HU king Israel
    ‘David is king of Israel’

(b) hu malax ’al jisra’el
    הוא מלך על ישראל
HU rule-past over Israel
    ‘He ruled over Israel’

In (1) (b) the pronoun is the subject of the sentence. In (1) (a) the subject is an overt noun, and the historical pronoun serves as a copula.

In many sentences, and in most registers of Modern Hebrew, the overt copula in present tense is optional, and the more formal and archaic version of the sentence would leave it out. This is true in particular of attributive sentences, coding class membership. (See chapter 2, Figure 1).

(2)
(a) Jerusalajim ’ir xasuva.
    ירושלים עיר תשובה.
Jerusalem city important
    ‘Jerusalem is an important city.’

(b) Jerusalajim hi ’ir xasuva.
    ירושלים היא עיר תשובה.
Jerusalem HI city important
    Jerusalem is an important city.’
Thus, in (2), which could be used in formal or informal settings, Modern Hebrew speakers have the option of either (a) or (b). Both versions are grammatical.

The vocabulary and general tenor of (3), on the other hand, clearly signal an informal setting. Still, for many speakers, inclusion of the equative linker is optional. For others, it is mandatory.

(3)

(a) Josi baxur nemad  יוסי בוקר נחמד.
    Josi guy nice
    ‘Joe’s a nice guy’

(b) Josi hu baxur nemad  יוסי هو בוקר נחמד.
    Josi HU guy nice
    ‘Joe’s a nice guy.’

Actual usage varies from speaker to speaker, but a general rule of thumb is that the more copulas are included in a discourse, the lower the register and the higher the affect. The trend is toward inclusion. Colloquial Hebrew uses the overt copula with ever increasing frequency, and in some registers its use is mandatory even in attributive sentences.

Moreover, in true equational sentences, coding identity of participants, the inclusion of the appropriate third person pronoun is mandatory for all registers.

(4) David hu ha-melex  דודהואמלך.
    David HU the-king
    ‘David is the king’
If the copula were removed from (4) above, it would be interpreted as a definite noun with an attributive nominal, following the ‘noun-adjective’ template.

(5) David ha-melex ...
    דוד המלך...
    David the-king
    ‘King David’ [David the King]

Nouns and adjectives are treated equally in such collocations. In (6) below, (a) and (b) are statements of class membership, in which use of the copula is optional, while (c) relates an identity. Example (6)(c) would be appropriate in a situation where there is a group of people, one of whom we know to be an Australian. The statement in (6)(c) identifies Lynn to be that specific Australian of whom we had been speaking. On the other hand, (6)(d) is not a statement at all. It is a phrase in which ‘Australian’ modifies ‘Lynn’.

(6) (a) Lynn ‘ostralit
    לינן אוסטרלט.
    Lynn Australian
    ‘Lynn is Australian’

(b) Lynn hi ‘ostralit
    לינן הי אוסטרלט.
    Lynn HI Australian
    ‘Lynn is Australian’

(c) Lynn hi ha-‘ostralit
    לינן הי האוסטרלט.
    Lynn HI the-Australian
    ‘Lynn is the Australian’
(d) Lynn ha-'stralit
Lynn the-Australian
'lyn the Australian' [the Australian Lynn]

Already we can see that the Hebrew reinterpretation of the pronoun as a copula is less complete than in Chinese. The change has not yet run its full course, and a complete split may never happen, but even so, in a large and ever growing portion of Hebrew discourse, the third person pronominal morphemes are obligatorily included to serve a copular function.

#4.2.2 Agreement for Gender and Number -- But Not Person

Unlike Chinese shi, Modern Hebrew וּצְרוּ [hu] inflects for person and number. Hebrew verbal agreement, on the other hand, has an inflectional system with three way distinctions for person, gender and number. Nominal agreement only distinguishes gender and number. Thus, by adhering to the number and gender distinctions, וּצְרוּ [hu] falls into the parameters reserved for a nominal, not a verbal form. This is not so anomalous for Modern Hebrew as a whole, because the same is true for all present tense conjugations, and the pronominal forms function as copulas only in the present tense. The full verbal status in Hebrew is reserved for past and future, which in Biblical Hebrew were perfective and imperfective aspect, respectively. Present tense derives from the Biblical participle, a nominalized form of the verb.

In Figure 1 below, I list all the Modern Hebrew subject pronouns.
Figure 1: MODERN HEBREW PRONOUNS

<table>
<thead>
<tr>
<th>SINGULAR</th>
<th>PLURAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ST</td>
<td>'ani ('anoxi)</td>
</tr>
<tr>
<td>2ND masc.</td>
<td>'ata</td>
</tr>
<tr>
<td>fem.</td>
<td>'at</td>
</tr>
<tr>
<td>3RD masc.</td>
<td>hu</td>
</tr>
<tr>
<td>fem.</td>
<td>hi</td>
</tr>
</tbody>
</table>

The overt copula הָיוֹלָה [hi] agrees with its subject in gender and number, as in (7) below, where the subject is feminine singular.

(7)HALAŠON ha'ivrit hi 'anaf bamiśpaxa gdola.  
the-language the-Hebrew HI(COP) branch in-family big.  
'The Hebrew language is a branch of a large language family.'

On the other hand, the overt copula does not agree in person with its subject, as in (8) below, where the subject is second person, but the copula, as always, is הָיוֹלָה [hu].

(8)'Ata hu ha-Ḥish.  
thou(masc.) HU(COP) the-man  
'You are the man.'

The mismatch in persons is one indication that הָיוֹלָה [hu] and its counterparts for gender and number (but not person) are functioning as present tense copulas, and are not merely serving as deictics. The loss of the category of person is one property which leads us to conclude that the erstwhile pronouns have grammaticalized in such utterances, and are for all intents and purposes full-fledged copulas.
§4.2.3 Agreement for Definiteness

In Hebrew, nominals agree in definiteness with other nominals which they modify. On the other hand, the category of definiteness does not apply to verbal agreement. Thus, attributive adjectives agree in number, gender and definiteness with the noun they modify, but predicate adjectives only agree in gender and number.

(9)

(a) ha-perax ha-jafe...

the-flower the-pretty

'The pretty flower ...'

(b) ha-perax jafe.

the-flower pretty

'The flower is pretty.'

Definiteness is a factor in the mandatory or optional use of an overt present tense copula. If either nominal in a copular statement is definite, the overt copula is mandatory. The copula is mandatory in both (10) (a) and (b)

(10)

(a) David hu ha-roce'ax

David HU the-murderer

'David is the murderer'
(b) ha-roce'ax hu David
the-murderer HU
'The murderer is David'

In (10) we are assuming that a murder has been committed, that we know it was committed by one and only one person, and we are asserting that David and the murderer are one and the same person. Compare (11) below, where the overt copula is optional in some registers.

(11) David (hu) roce'ax
David HU murderer
'David is a murderer'

In (11), there is no reference to any particular murder. This is a general assertion of class membership.

We are dealing with the same issue of identity versus class membership that we discussed before, this time couched in terms of definiteness. That is, the overt copula is mandatory with a definite predicate. But we have already seen in Chapter Two that definiteness is one way in which languages distinguish between statements of identity and class membership. A definite predicate in a copular clause signals identity, an indefinite indicates class membership. Thus the two characterizations of this phenomenon are functionally equivalent.

For historical purposes, however, examining the issue from the standpoint of definiteness yields some useful results. As a demonstrative pronoun, ꝃ [hu] follows the same dictates as any other nominal: when used as a predicate it is not definitized; when used as an attribute, it is.
(12)

(a) Ha-‘iš. “HU the-man ‘He is the man.’ [That’s the man.]

(b) ha-‘iš ha-hu .... “the-man the-HU ‘That man …’

(c) ha-‘iš hu. “the-man -HU ‘He is the man’ [Archaic or extremely topicalized.]

From this standpoint, looking at the overall pattern of the language, we can infer that in sentences such as 12(a), Ha- ‘nu [hu] is the predicate. In fact, this is true regardless of word order. In 12 (c), where the order is reversed, Ḥa- ‘nu [hu] is still the predicate. The mere fact of its failure to agree in definiteness with the other nominal determines it.

The association of the indefinite form of the demonstrative with predication may have helped to prime speakers to reinterpret the third person pronouns as copulas.

4.3 Review of the Literature

Devitt (1992.10-11) notes that while an overt copula is obligatory for present tense identity statements and optional for attributive assertions, the zero allomorph is mandatory with pronominal subjects in attributive sentences. He goes on to state that in identity constructions, pronominal subjects are used with an overt copula.
This is close, but not precisely accurate. The general rule is that echoing pronouns and copulas are awkward and are avoided, regardless of the identificational or attributive nature of the statement.

(13) (a) *hu hu ha-roce'ax
    HU HU the-murderer
    ‘*He is the murderer.’ [Could mean: ‘He is indeed the murderer.’]

(b) hu ha-roce'ax
    HU the-murderer
    ‘He is the murderer.’

(c) *hi hi rocaxat
    HI HI murderess
    ‘*She is a murderess.’ [Could mean: ‘Indeed she is a murderess.’]

(d) hi rocaxat
    HI murderess
    ‘She is a murderess.’

When echoing pronouns and copulas do occur, they serve the same function as reduplication does in general: they emphasize the assertion and add affect.

This avoidance of repetition does tell us something about the degree to which these morphemes have grammaticalized. The split between pronouns and copulas is not so full and complete that speakers do not recognize them as the same morpheme
when they are placed side by side in the same utterance. For one thing, their outward
form has not changed.

This leaves us with the age old dilemma: how do we know when יָהוּ [hu] is being
a pronoun and when it serves as a copula? Is there a sharp demarcation, or are there
areas of overlap?

Li and Thompson (1977:429–431) offer the following arguments in favor of a
split between the pronoun and the copula in identity statements: “First, the subject
of an equational sentence can be indefinite, but the topic in a topic-comment
sentence cannot... Second, the predicate can be questioned in an equational sentence,
but not in a topic-comment sentence ... A third argument, cited in Faltz (1973), is
that sentences [with an overt copula] can be pronounced with an intonation suitable
for a single, simple clause ...”

All of these arguments are presented from the assumption that the pronoun was
reinterpreted as a copula from a topic-comment sentence in which the subject was
fronted, as in ‘Tom, he’s a soldier.’

While I accept the topic-comment mechanism as one of the means whereby this
change came about, I do not consider it to be the only one. I believe that an overall
shift in the system of the language facilitated the reinterpretation of these pronouns
as copulas. I generally agree with Li and Thomspont’s argumentation in favor of a
split, but I believe that the change was gradual and that there are areas of semantic
overlap at play even now.

Likewise, the issue of mismatch in person is not altogether an overriding
indication of a complete grammaticalization of the pronoun into a copula. While it
exemplifies a rare occurrence, (14) below shows that such a mismatch took place
already as early as Isaiah.
(14) Isaiah 52:6

(a) laxen jed a ʿami šnī, laxen bajom ha-hu

therefore know(imprf) people-my name-my, therefore on-day the-HU

'Therefore on that day my people will know my name,'

(b) ki ʿani hu hamadaber hineni

that I HU the-speaker here-I

'That I am the speaker, here I am.'

"Therefore my people will know my name, for it is I who speak, behold me."

Li and Thompson suggest a topic-comment interpretation for purposes of determining whether the pronoun has grammaticalized into a copula or not. When they apply that test to mismatched persons, the result is a nonsensical sentence: hence, they find that the grammaticalization is complete. If we applied this treatment to (14) (b), we get: 'I, he's the speaker', as in (15)(a). That would support a conclusion that ḧw [hu] has already been grammaticalized as a copula in late Biblical Hebrew. I would suggest a different approach. The less grammaticalized version of this clause does not topicalize the first nominal, it backgrounds the last noun. So the more literal reading of (14)(b) is 'I am he, the speaker', as in 15(b). That makes a little more sense.

(15)

(a) *ʿani, hu ha-madaber

I, HU the-speaker

*I, he is the speaker.'
(b) ‘ani hu, hamadaber  אָנִי הָמוֹדֶבֶר  I HU, the-speaker
‘I am he, the speaker.’

Now, if we can do this with Isaiah, we can do the same thing to modern sentences about murderers. Which means that there is still some flicker at play in the system. And this is not a bad thing for the purpose of our analysis. It shows us how the change came about.

§4.4 Biblical Sources for the Modern Day Construction

The Hebrew independent pronouns are very old and are common not only to Biblical and Modern Hebrew, but to many other Semitic languages, with minor differences. Aramaic, like Hebrew, uses the third person pronouns as present tense copulas. As does Arabic.

Modern Hebrew usage of third person pronouns as present tense copulas is not so much an innovation as an extension of previous usage.

Even in Biblical Hebrew, a statement of identity required the use of a pronominal copula.

(16) Genesis 48:07  בָּדֶּרֶךְ אֵфְרָאָת הַיָּהָ בֵּית לְהָם
  ba-derex 'efrat hi bejt laxem  בָּדֶּרֶךְ אֵפְּרָאָת הַיָּהָ בֵּית לְהָם
  in-way-of Efrat she house-of bread
  'by way of Efrat which is Bethlehem’

In (16), Efrat is another name for Bethlehem. The identity is established by way of a pronoun. Similarly in (17) Jerubaal is another name for Gideon.
(17) Judges 7:1

wa-jaškem jeruba’al hu gid’on

(And-)got-up-early Jerubaal HU(PRN/COP?) Gideon

'And Jerubaal who is Gideon arose early..."

The difference between these identity statements and those in section 4.3, above, is that here the identity is established parenthetically, as an aside to the general purport of the sentence. Hence we might discount the interpretation that the pronoun is a copula and suggest that it is being used to assert an apposition.

The situation is compounded by the fact the Biblical texts do not tend to make genuine statements of identity or definition. The rare instances where a statement of identity appears tend to occur in dependent clauses.

There is a cognitive significance to this fact in addition to its grammatical implications. Biblical texts describe actions, people and physical entities, they proscribe and prescribe various activities, they narrate events, they present songs of joy and lamentation -- but they do not engage in formal logic or give outright definition of terms. There are no occasions in the course of a Biblical text for statements of pure identity. Hence there is no need for an equative copula.

There are, on the other hand, plenty of statements of attribution, asserting class membership. In those sentences, no overt copula is used. But then, none is required. Most adjectives are stative verbs, and in the present tense appear as participles. Many nouns are derived from participles as well. Examine (18) below.
(18) **Numbers 35:16**

wa-im bi-xli barzel hikahu wajamot

and-if with-instrument iron struck-him and-died

roce'ax hu

murderer he

'And if he struck him with an iron instrument and he died, then he is a murderer.'

The above is as near to a definition as the Bible will come. Notice that it requires two clauses, an 'if' clause and a 'then' clause. We are not told 'a murderer is a killer who uses an iron instrument...'; we are told 'if he struck and killed with an iron instrument, then he is a murderer.' But grammatically, the clause 'he is a murderer' could just as easily be interpreted as 'he murders.' 'Murderer' in Hebrew is a participle.

§4.5 **Typological Environment**

The transformation of a copula to a pronoun did not happen in a vacuum. Other changes from Biblical Hebrew to Modern Hebrew that facilitated the reinterpretation of the pronouns as a copular verb are as follows:

(1) Biblical Hebrew had a very flexible word order, whose most common surface manifestation was VSO, but which allowed almost every part of speech to be fronted depending on context. See Chapter 7, infra. Modern Hebrew has a far more rigid word order and its basic configuration is SVO. (Givon 1977). While literary Hebrew
does admit more variations, colloquial Hebrew is less and less willing to accept deviations from the basic word order.

(2) Biblical Hebrew had an aspect system. Modern Hebrew has tense. Perfect has become past, imperfect has become future, and the present participles have become the present tense.

(3) Change (2) has further ramifications for use of nominals as verbal elements. While present and future verbs in Hebrew are inflected to show a three way agreement for person, number and gender, present tense only distinguishes gender and number. When the third person pronouns are reinterpreted as the present copular forms for all persons, they do not stand out as being a defective form, because no other present tense verb has person marking either.

All of the above changes conspired to ease the reinterpretation of the pronoun as a copula. There was no sharp transition, no noticeable departure from previous usage. The change was part of an overall regrouping of linguistic forces.

§4.6 Possible Source Constructions

Possible source constructions for the reinterpretation of the third person pronouns as copulas include:

(I) A, HU B \[ \Rightarrow \] A HU B

(II) A HU, B \[ \Rightarrow \] A HU B
(III) HU (SUBJECT) PARTICIPLE (VERB) $\Rightarrow$ HU (COPULA) PARTICIPLE (SUBJECT)

(IV) PARTICIPLE (VERB) HU(SUBJECT) $\Rightarrow$ PARTICIPLE (SUBJECT) HU (COPULA)

All of these patterns involve a flicker in sentence templates. In the case of (I) and (II), the elimination of a pause, signalling clause boundaries, is involved. This is the sort of explanation put forth by Li and Thompson (1977). It is an attractive mechanism, because all it really requires is that speakers be a little hurried, or a little too inclined to emphatic constructions. We can well imagine such a reanalysis happening synchronically, during a conversation, in which the speaker meant the phrase one way, while the hearer took it differently. Such a change could have come about during a relatively short period, under appropriate linguistic stress. It would be the sort of explanation that would resolve the problem if this change had occurred in recent history, with the revival of Hebrew.

But since both Aramaic and Arabic use their copulas in similar fashion, I am inclined to believe that this change was germinating even before those languages split, and that it was slow and ponderous, rather than swift and cataclysmic.

I would suggest that (I) and (II) probably were involved in the reanalysis of נָתַן [nu], but I do not think that they were as instrumental as the overall flicker in verbal and nominal interpretations represented by (III) and (IV).

In a language with a zero copula in which subject and predicate can switch positions, appearing initially or finally in a two word clause, there is lots of
room for reinterpretation. When we add to this the fact that participles are interpreted sometimes as verbs and sometimes as nouns, the language leaves a great deal of ambiguity with regard to grammatical category in a clause.

(19)
roe'ax hu
murderer he

‘VERB (murders) NOMINAL (he)’ \(\rightarrow\) ‘a murderer (NOMINAL) COPULA (is [he])’

It is the flicker of the verbal and nominal categories that allows the reanalysis of a pronoun as a copula to take place. The same flicker will be discussed again in Chapter Seven, this time in reference to a much earlier grammaticalization cycle in Semitic.

§4.7 Conclusions

The reinterpretation of the pronoun as a copula in Modern Hebrew goes hand in hand with two typological changes: (1) Conversion from flexible VSO word order to SVO and (2) reinterpretation of participles as (a) nouns and adjectives and (b) as present tense verbs. The instability in nominal and verbal categories is behind the flicker between pronoun and copula.
Chapter 5. The cycles of Finno-Ugric on

§5.1 Preview

This chapter presents an example of grammaticalization of a third person pronoun into a copula in yet another language family. It will be seen that a common root for this pronominal form exists in both Hungarian and Finnish, and that an earlier Finnish pronoun on (currently han), possibly derived from the ol root, was later reinterpreted as the third person singular form of the verb 'to be'.

The Finno-Ugric example examined in this chapter is not surprising in its scope and directionality, given the Modern Chinese and Modern Hebrew examples discussed in Chapters 3 and 4, respectively. However, in this case, the discovery is new. There does not seem to be much literature on the subject of the evolution of the Finnish third person copula. Thus, the material in this chapter can be seen as an original contribution to the field in the following ways:

(1) As an addition to historical work on Finno-Ugric,

(2) As another example in the growing body of work on grammaticalization,

and

(3) As a specific instance of the cognitive principles at play in the speaker’s perception of pronoun and copula, as discussed in Chapter 2.

§5.2 A Sidetrip into Russian.

Since many Finnish speakers have been exposed to Russian over a large span of time, and since the third person form of the verb 'to be' in Finnish is remarkably similar to the Russian third person pronoun, I would like to take a moment to explore syntactic and evolutionary parallels in the Russian copula. This exploration will show cognitive correspondences, but it is not intended to imply borrowing or
even areal influence from Russian. As further discussion of other Finno-Ugric reflexes of Finnish on will show, it is not necessary to look outside the language family for a source. Nevertheless, the following discussion is illuminating for the progression of pronoun-to-copula internal reanalysis, even in a language where full reanalysis has not yet taken place and may never actually be realized.

Russian does not use an overt copula in the present tense. Instead, syntactic juxtaposition, much as in Biblical Hebrew and in Turkish, is used to code the relation. † In Russian, Ivan bolen means ‘Ivan is sick’ (1). A pattern of left dislocation, for the purpose of topic marking has developed in colloquial Russian. Thus Ivan, on bolen often occurs in discourse. See example (2).

In addition, a subgroup of less educated Russians tend to say Ivan on bolen—‘Ivan he sick’, without any sort of pause between the noun and the pronoun. That is, what was originally a device for topicalization is becoming entrenched and rigid. (See Li and Thompson’s analysis of topic-comment structures (1977).) Grammar teachers correct such usage, but it persists. (Albina Nikolaevna Hill 1992, personal communication.) Historically, grammar teachers very often fight a losing battle. The speaker of such an utterance is treating the pronoun as an equative or conjunctive particle. In the case of Russian, this entrenchment has not made its way into acceptable grammatical structure. I am not necessarily suggesting that it ever will. But the possibility is there.

(1) Ivan bolen.

Ivan sick

‘Ivan is sick.’

†1 For a comparison of the Russian copula to that of other languages discussed in this dissertation, see Figure 1 in Chapter 2.
(2) Ivan(,) on bolen.

Ivan  he  sick

'Ivan is sick.'

We would do well to keep the Russian pronoun *on* in mind as we return to an examination of the Finno-Ugric family. This is not because Finnish *on* is related to Russian *on*, but rather because the historical change whose initial stages we see in Russian has in fact taken place in Finnish.

5.3 The Uses of Finnish *on*

We now turn to the Finnish in sentence (5).

(5) Taivas on sininen.

sky  is blue

'The sky is blue.'

(Lehtinen 1963.610)

The word *on* that we find intervening between 'sky' and 'blue' is not synchronically a pronoun -- it is a verb. Our previous discussion may well have prepared us to deduce that *on* is a the third person singular pronoun, if not in this particular construction, then at least in some contexts of usage. But it is not: at least not in Modern Finnish. It is, in fact, the third person singular conjugation of the verb 'to be'. The stem of the Finnish verb 'to be' is *ole-*. 
The Finnish verb ‘ole-’ in the present tense (with corresponding pronouns)

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st person</td>
<td>(mina) ole-n</td>
<td>(me) ole-mme</td>
</tr>
<tr>
<td>2nd person</td>
<td>(sina) ole-t</td>
<td>(te) ole-tte</td>
</tr>
<tr>
<td>3rd person</td>
<td>han, se on</td>
<td>he, ne o-vat</td>
</tr>
</tbody>
</table>

(Lehtinen 1963:33)

An examination of the paradigm reveals that there is suppletion at play here. Hakulinen (1961:169) notes: "...evidence of other Finno-Ugric languages indicates that l has always been part of the stem, so it is not possible to regard it as a derivational suffix attached to an original "o stem." This is even more apparent when we look at the present tense of typical Finnish verbs.

The general rule is that when a stem is bisyllabic and ends in a vowel, the third person singular ending is a reduplication of that vowel. An example is presented in (7) below. The verb puhu means ‘to speak’.

(7) The Finnish verb ‘puhu’ in the present tense (with corresponding pronouns)

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st person</td>
<td>(mina) puhu-n</td>
<td>(me) puhu-mme</td>
</tr>
<tr>
<td>2nd person</td>
<td>(sina) puhu-t</td>
<td>(te) puhu-tte</td>
</tr>
<tr>
<td>3rd person</td>
<td>han, se puhu-u</td>
<td>he, ne puhu-vat</td>
</tr>
</tbody>
</table>

Monsyllabic Finnish verbs take a null third person singular ending. In (8) below we see the monosyllabic root juo conjugated, with a null third person ending.
(8) The Finnish verb 'juo' in the present tense
(with corresponding pronouns)

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st person</td>
<td>(mina)</td>
<td>(me) juo-mme</td>
</tr>
<tr>
<td>2nd person</td>
<td>(sina)</td>
<td>(te) juo-tte</td>
</tr>
<tr>
<td>3rd person</td>
<td>han, se</td>
<td>he, ne juo-vat</td>
</tr>
</tbody>
</table>

Since ole- is a stem ending in a vowel, the regular form of the third person singular would be *ole-e (Hakulinen 1961:169). But, even if for some reason the stem were reanalyzed as monosyllabic, the result would be *ol, not on. In any event, a stem ending in a consonant does not fit into the normal syllable structure for Finnish verbs. Hakulinen (1961.69) notes: "...we may therefore be dealing with an old noun quite unrelated to the ole- stem, probably an old substantive that became attached to the ole paradigm." The candidate preferred by Hakulinen is the possessive adjective oma, meaning 'own'. I would suggest a different source.

The variables that obtain for the Finnish 'to be' appear to be linked with the scarcity of third person marking in the Finnish paradigm. The nominal most likely to be recruited to fill this slot would be none other than the missing third person marker.

With this in mind, let us examine the obligatory occurrence of overt personal pronouns in present tense sentences -- which is limited to the third person.

In first and second person, pronominal marking is the only morphological indication of person required by the language. Independent pronouns may be used, but they are not required, and they are often omitted. In formal written and spoken Finnish, omission of these pronouns is almost mandatory. Colloquial and urban speech includes them more often. (Lehtinen 1963:34).
But the third person pronouns are obligatory. They may not be left out. This is significant in tracing the evolution of the Finnish word order, as well as the overt copula.

If we assume that at an earlier stage of the language no overt copula was required for third person, and at the same time, the person marking was in the form of independent pronouns, as opposed to inflection, we might find that: (1) the language was verb initial followed by the subject noun or pronoun and (2) the third person singular pronoun was on.

Example (9) below represents the hypothesis we are now considering. The only difference between (9) and (6), above, is that we have redrawn the morphological boundaries.

(9) The Finnish verb ‘ole-’ in the present tense
(with internally reconstructed independent pronouns derived from morphology)

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st person</td>
<td>* ole en</td>
<td>* ole emme</td>
</tr>
<tr>
<td>2nd person</td>
<td>* ole te</td>
<td>* ole ette</td>
</tr>
<tr>
<td>3rd person</td>
<td>* on</td>
<td>* ole vat</td>
</tr>
</tbody>
</table>

The hypothesis represented in (9) is that there was a zero copular morpheme in third person singular, a crosslinguistically attested occurrence. Insofar as the placement of pronouns is concerned, the pattern in (9) would apply to all verbs in that stage of the language. The zero root morpheme for the copular third person would have been unique to ole.

So far, the above is somewhat speculative. But now we turn to the Hungarian branch of Finno-Ugric to examine both its pronouns and its copula in search of cognates.
5.4 The Hungarian Reflex of On

Below is the full complement of subject pronouns in nineteenth century Hungarian.

(10) The Hungarian nominative pronouns:

<table>
<thead>
<tr>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>en</td>
</tr>
<tr>
<td>2nd familiar</td>
<td>te</td>
</tr>
<tr>
<td>2nd formal</td>
<td>ön</td>
</tr>
<tr>
<td>3rd</td>
<td>ö</td>
</tr>
</tbody>
</table>

(Whitney 1977.15)

In this case, the singular third person form is Ó and the addition of the plural ending k does not reveal an underlying n. But, when we turn to the formal second person pronouns, we find that the n has been preserved. The explanation for this is that the formal address is derived from the third person pronouns. Hungarian formal second person forms generally behave like third person (Payne 1987.15). Since it represents a more conservative pattern, the formal address retained the n, long after it was dropped in the original usage of the root for third person pronouns.

Thus we see that both Finnish and Hungarian had at one time a third person singular pronoun on. The relationship between Hungarian and Finnish is well established. So that when we turn to the curious paradigm of the verb to be in Finnish, we would be well advised to consider the possibility that the third person conjugation was formerly the third person pronoun.

The course of this development may be plotted in tandem with typological changes to be found among the three languages. Finnish word order is basically SVO. Use of third person pronouns is mandatory. The copula is required. The Finnish verbal agreement would tend to indicate that at some earlier stage, Finnish was primarily VSO. Hungarian seems to find itself in the middle of the progression,
meeting neither the old (VSO) nor the new pattern (SVO). The Hungarian word order is so flexible that some have termed it "free word order." But it is sufficient to say that any number of surface orders can be found in Hungarian sentences. In Hungarian the present tense copula is required in all persons except third. Likewise, pronouns are mandatory, everywhere but in third person.

Examples (11) through (14) show different instances of word order and the use or omission of pronouns in Finnish and Hungarian. Sentence (8) demonstrates SVO word order in (a) and the obligatory use of a copula (b) in Finnish.

(11)
(a) Han nieli sen.
   s/he swallow(past) it(accusative)
   'She/he swallowed it.'

(b) Se on kyna.
   this is(3rd, present) pencil
   'It's a pencil.'

Sentence (12) displays omission of the subject and an (S)VO ordering in Hungarian.

(12)
Eszi az ebedet.
   eat(3rd person sg.present) the dinner
   'He eats the dinner.'

In (13) We see an example of the extremely flexible word ordering in Hungarian.
(13)

(a) Janos téte a konyvet az asztalra.
Janos put the book-acc the table-on
‘Janos put the book on the table.’

(b) Janos a konyvet tette az asztalra’.
Janos the book-acc put the table-on
‘Janos put the book on the table.’

(c) A konyvet az asztalra Janos tette.
the book-acc the table-on Janos put
‘Janos put the book on the table.’

(Kiss 1986.19)

The word orders in (a) (b) and (c) of (13) above are all equally acceptable. Thus we can have SVO, SOV or OSV. In (14) we see that a verb to be is required in Hungarian in all persons but the third. Third person singular and plural do not require an explicit copula.

(14)

(a) (En) vagyok gazdag.
be(1st sg.) rich(sg.)
‘I am rich.’

(b) (te) vagy gazdag.
thou be(2nd sg.) rich(sg.)
‘You are rich.’
(c) Ő gazdag.
he rich(sg.)
‘He is rich.’

(d) (Mi) vagyunk gazdagok.
We be(1st pl.) rich(pl.)
‘We are rich.’

(e) (Tí) vagytok gazdagok.
You be(2nd pl.) rich(pl.)
‘You are rich.’

(f) Ők gazdagok.
They rich(pl.)
‘They are rich.’

In Hungarian, where the copula is obligatory, the pronouns are optional (first and second person). But where the copula is deleted, the pronouns are required (third person). In Finnish, the first and second person pronouns are optional, but the third is required. The copula in Finnish is always mandatory.

The curious negative correlation between the obligatory use of pronouns and of the copula in Finnish as opposed to Hungarian is worth a closer look. Figure 1 below tabulates the constraints on zero anaphora in Finnish and Hungarian.
<table>
<thead>
<tr>
<th>REQUIRED COPULA</th>
<th>FINNISH</th>
<th>HUNGARIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ST</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>2ND</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>3RD</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>PL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ST</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>2ND</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>3RD</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>REQUIRED PRONOUN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ST</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>2ND</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>3RD</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>PL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ST</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>2ND</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>3RD</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

Figure 1.
COMPARISON OF FINNISH AND HUNGARIAN ZERO ANAPHORA AND NULL COPULA
We find that while obligatory use of the copula and of the pronoun, respectively, is in complementary distribution in Hungarian, the same is not true for Finnish. That is, where a pronoun must be used in Hungarian, it is because the copula is omitted. When the overt copula is obligatory in Hungarian, then the pronoun is optional. But in Finnish, use of an overt copula is always required, while the pronouns can be omitted in first and second person. But in Finnish third person the pronouns are required, following the pattern of distribution in Hungarian.

The reason for this is as follows: third person in Finnish has minimal person marking morphology. This is true for all Finnish verbs, not just the copula. Hungarian shows us a stage of development in which the pronoun appears where the copula does not. It is precisely this state of affairs that facilitates a reanalysis of the pronoun as a copula.

When the copula is missing in an equative statement with a pronominal subject, the full force of of the linking function may be attributed to the only available morpheme: the the third person pronoun. Does it mean 'he' or 'is'? Speakers resolve the issue differently depending on their expectations. If first and second person pronouns can be omitted, but a copular morpheme is obligatory, speakers may come to feel that the third person pronoun has likewise been omitted, and that the included morpheme is in fact a third person form of the copula.

We may assume that reanalysis of the former pronoun *on* in Finnish probably went through the stages set down in (15). (Unattested forms will henceforth be marked with a %. )
(15)

(a) STAGE ONE  (a')
 % Taivas sinienen. % On sinienen.
   sky  blue       s/he blue
 'The sky is blue.' 'It's blue.'

(b) STAGE TWO

 % Taivas (on) sinienen.
   sky (s/he) blue
 'The sky is blue.'

At this stage, the pronoun was optional and used for emphasis.

(c) STAGE THREE

Taivas on sinienen.
   sky  is blue
 'The sky is blue.'

(Lehtinen 1963.610)

This represents the current state. On is now a verb and its use is obligatory.

§5.5 Mechanisms of Reanalysis

The reanalysis suggested in (15) above might have been fueled by more than one mechanism. I will list a couple below:

(I) Topic-Comment to Subject-Predicate  for three element utterances (Li & Thompson 1977):  A, ON B ==⇒ A ON B
(II) Predicate-Subject to Subject-Predicate for two element utterances (as I suggested for Hebrew in Chapter Three):

\[ A \text{ ON (VS)} \implies A \text{ ON (SV)} \implies B \text{ ON A (SVS)} \]

The first explanation is that propounded by Li and Thompson (1977). We discussed it earlier in the chapters on Chinese and Hebrew. While I do not question its validity, I believe that the second mechanism can give a fuller picture of the overall linguistic environment that would promote such a change. Speakers routinize oft repeated patterns, especially those involving pronouns. Even though the linguistic environment was shifting in the direction of SVO syntax, it may be that older templates were repeated when it came to the pronouns. Thus 'A ON' may have been the acceptable way to say 'THIRD PERSON PRONOUN-ZERO COPULA--A', long after most other two element clauses had reversed their order, with the nominal coming first. Under such circumstances, later speakers may have reinterpreted the same two word utterance as consisting of a nominal followed by a copular verb: 'A COPULA'.

The above in itself would serve as a satisfying explanation for the third person conjugation of the verb to be in Finnish and its connection to general typological developments in that language. But this opens up whole new vistas for exploration. We now know that on was a pronoun before it came to be a copula in Finnish. We may well ask: what was on before it became a pronoun? Is the similarity between the root ole- and on a complete coincidence?

We do not have enough information to pursue this question further in Finno-Ugric at present, but it would seem quite plausible that at an even earlier
stage than the one we postulate, the pronoun *on may have originated from the verb
ole-.

For an example of just such an occurrence, we turn to a discussion of the Turkish
pronoun in Chapter 6. No claim is being made for (or against) genetic relation
between Turkic and Finno-Ugric. The analogy is purely developmental.

§5.6 Conclusions: Typological Dating and Grammaticalization

It has been claimed that degree of grammaticalization is directly linked to the
dependent or independent, complex or simplex, characterization of a gram. Some
have suggested that a multimorphemic gram must be more grammaticalized than
one that is isolated and monomorphemic. (Bybee 1991:44–45)

Finnish *on is monomorphemic and isolating. It is also an innovation, in
comparison to the first and second person conjugations of ole. On has undergone
more cycles of change than olet, and yet it is simplex. In fact, the very overuse of this
gram may have contributed to its unanalyzable monolithic character. At an earlier
stage, long before Proto-Finno-Ugric, *on may have consisted of a root and a person
marking.

Grammaticalization moves in cycles. Perhaps we can determine which form is an
innovation within a particular paradigm simply by observing its complexity, but we
cannot determine which form has undergone more cycles of change using this
method. We must be very careful to distinguish between these two claims.
Chapter 6. The Turkish Third Person Pronoun

§6.1 Preview

In this chapter I will present evidence that the Turkish third person singular pronoun \( o(n) \) evolved from the third person singular present of the verb ‘to be’.

This discovery is a departure from the grammaticalization pattern examined in previous chapters. In Chapter Three we watched the Modern Chinese copula, \( s\text{hi} \), evolve from a third person demonstrative pronoun. In Chapter Four we saw that Modern Hebrew has a present tense copula that developed from third person pronouns. In Chapter Five, we observed that the Finnish third person singular conjugation of the verb ‘to be’ came from an archaic pronoun. Now, we will examine a case of the opposite effect: the evolution of a pronoun from a copula. †

§6.2 The Turkish Third Person Singular Pronoun

The Turkish third person singular pronoun in the absolute case is \( o \). The full paradigm is displayed below in (1). (Lewis 1967.67)

(1) The Turkish third person pronouns:

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>abs.</td>
<td>( o )</td>
<td>onlar</td>
</tr>
<tr>
<td>acc.</td>
<td>onu</td>
<td>onlari</td>
</tr>
<tr>
<td>gen.</td>
<td>onun</td>
<td>onlarin</td>
</tr>
<tr>
<td>dat.</td>
<td>ona</td>
<td>onlara</td>
</tr>
<tr>
<td>loc.</td>
<td>onda</td>
<td>onlardar</td>
</tr>
<tr>
<td>abl.</td>
<td>ondan</td>
<td>onlardan</td>
</tr>
</tbody>
</table>

† This dissertation makes no claim for or against any genetic relationship between Turkish and Finno-Ugric. That is, I make no claim that the third person pronoun \( on \) in Finno-Ugric is (or is not) related to the third person \( o(n) \) in Turkish. Likewise, I make no claim that Finnish \( ole \) ‘to be’ is (or is not) related to Turkish \( olmak \) ‘to exist’.

The establishment of genetic relations between language families is a long and difficult project and involves the exhaustive comparison of phonological, morphological, syntactic and lexico-semantic features. Such an undertaking is entirely outside the scope of this dissertation, which deals with a specific instance of grammaticalization in each language.
While the absolute form of the singular pronoun is o, a cursory examination of the paradigm will reveal that the underlying form is on and that the missing n appears everywhere except in the absolutive singular.

In pre-nineteenth century texts, the usual form of the third person singular absolutive was ol. (Lewis 1967:67). There is an ol root in Turkish: it is the verb olmak ‘to exist’.

§6.2.1 Historical Evidence for the Use of the Pronoun As Copula

Clauson (1972:123), in a dictionary entry based on pre-thirteenth century Turkish texts, sheds a great deal of light on the usage of ol. There ol is listed as:

“(1) an indeclinable Demonstrative Adj.: ‘that’, as opposed to bu ‘this’.

“(2) a declinable Demonstrative Pron. ‘he, she, it’”

In describing the functions of ol, Clauson states: "As a Pron. it was also used in the early period as the equivalent of the copula ‘is, are’. In this last case there are two usages: (a) subject (sometimes implied, not stated), predicate, ol, meaning that the subject is the same as, or has the quality described by, the predicate; (b) subject , ol, meaning that the subject ‘is that’, often followed by a subordinate clause...” (Clauson 1972:123.)

That the Modern Turkish form o derives from ol is also evidenced in the earlier texts, where sometimes o and ol are used interchangeably. In one dialect, Southwest Turkmen, o is the informal variant of ol. (Clauson 1972:123).

That o derived from ol (and not vice versa) is evident from the following facts and general linguistic rules: earlier manuscripts almost invariably used ol, and o appears rather late; of the two ol has more phonetic material and the general rule of historic
change is in favor of loss; and lastly, colloquial forms are crosslinguistically more innovating, while the archaic are considered more formal.

The Modern Turkish verb for 'to be, to become, to ripen' is olmak. Since mak is merely the infinitival ending, it clearly follows that the early demonstrative pronoun and the copula share a common root: ol.

Now that we have identified the common source of the pronoun and the copula, it remains for us to determine which came first. I would like to suggest that ol is derived from the root of the verb olmak, and that the current paradigm for the pronoun developed from the third person conjugation of that verb.

\[6.3\] The Spurious Nasal in the Paradigm

If o came from ol, how is it then that we seem to have an underlying nasal in the Modern Turkish paradigm?

My original hypothesis ran as follows: nasalization of the alveolar consonant may have been a result of dissimilation in the environment of the plural marker. Thus ol + lar might have been confusing when it resulted in ollar and the first l may have been nasalized to compensate. Later the nasalization may have spread throughout the paradigm. The plural form in pre-nineteenth century Turkish is already nasalized, lending weight to this assumption.

However, historical and comparative evidence provides a simpler explanation: oblique cases of ol came from a different root an, and on represents a blend. So, for instance, in Uighur, a Turkic literary language, also known as Ancient Turkish, the pronominal third person paradigm ran as follows (Menges 1968:123).
(2) The Uighur third person pronouns:

<table>
<thead>
<tr>
<th>Case</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>nom.</td>
<td>ol</td>
<td>olar</td>
</tr>
<tr>
<td>gen.</td>
<td>anyng</td>
<td>olarnyng</td>
</tr>
<tr>
<td>dat</td>
<td>anga (angar)</td>
<td>olarga</td>
</tr>
<tr>
<td>acc.</td>
<td>any</td>
<td>olarny</td>
</tr>
<tr>
<td>loc./abl.</td>
<td>anta, anda; antada</td>
<td>olarda</td>
</tr>
<tr>
<td>abl.</td>
<td>antyn</td>
<td>olardyn</td>
</tr>
<tr>
<td>instr.</td>
<td>anyyn</td>
<td>---</td>
</tr>
<tr>
<td>comparat.</td>
<td>antaq, andaq; any-tag, antag</td>
<td></td>
</tr>
<tr>
<td>aequ.</td>
<td>anca</td>
<td></td>
</tr>
<tr>
<td>aequ./loc.</td>
<td>anacata</td>
<td></td>
</tr>
</tbody>
</table>

The pattern in the Uighur paradigm indicates that the oblique singulars come from a completely different root, while the nominative singular and all the plurals come from the ol root.

Interestingly enough, in Ancient Turkish, the first person demonstrative pronoun is bu, (which is the proximate third person demonstrative in Modern Turkish), and it follows a very similar pattern to that of the third person pronoun (Menges 1968:121):

(2) The Ancient first person pronouns:

<table>
<thead>
<tr>
<th>Case</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>nom.</td>
<td>bu, po</td>
<td>bular</td>
</tr>
<tr>
<td>gen.</td>
<td>munung</td>
<td>bularnyng</td>
</tr>
<tr>
<td>dat</td>
<td>munga (mungar)</td>
<td>bularga</td>
</tr>
<tr>
<td>acc.</td>
<td>muny, buny</td>
<td>bularny</td>
</tr>
<tr>
<td>loc./abl.</td>
<td>munta, muntada...bunda</td>
<td>bularda</td>
</tr>
<tr>
<td>abl.</td>
<td>mantyn</td>
<td>bulardyn</td>
</tr>
<tr>
<td>comparat.</td>
<td>muntay, munday; muny-tag</td>
<td></td>
</tr>
<tr>
<td>aeqat.</td>
<td>munca, bunca</td>
<td></td>
</tr>
</tbody>
</table>
That the proximal demonstrative was formerly used as first person pronoun is rather gratifying, since it confirms our earlier speculation on the cognitive connection between the persons and the deictics. But there is another more important issue here: Clauson (1972:331) states that an earlier form of the verbal root ol- was bol-. "Bol-, originally ‘to become’, implying a change of state ... From fairly early, however, bol- began to lose its distinctive character and verge, at any rate sometimes, towards ‘to be’... For unknown reasons the b- was elided in some Western languages at an early date, probably through an intermediate form with w-." Today, ol- is the only form in many dialects.

This might suggest that not only the distal but also the proximal demonstrative had copular origins. The evolution of the proximal may have been from bol to bu.

However, returning to our main subject, the pronoun ol/o clearly originated as a derivation from the root ol. Post-nineteenth century analogical processes resolved the split paradigm in favor of a nasal over an alveolar. This does not in any way reduce the connection between the demonstrative third person pronouns in Turkish with their original form as nominalizations from the root ol-.

§6.4 Zero Morphemes

We have seen that the pronoun ol served a copular function in early Turkish. Why then do we assume that the pronoun was derived from the verb ‘to be’ and not the other way around? Two main points of evidence support this:

(1) It is only in the very earliest texts that the pronoun serves this copular function. (Clauson 1972:123). That would indicate that the copular function came earlier, and that the use of the ol as a pronoun was the spreading innovation.

(2) Early Turkish oral texts make sparing use of independent pronouns. (See
Modern Turkish uses them amply. Again, the use of *ol* as a pronoun appears to be a spreading innovation.

Clauson (1972:123) has indicated that in early times, two sentence templates made use of *ol* (the demonstrative pronoun) as a copula:

(I) A B OL $\implies$ 'A is B'

Example: Bilge Toňukuk aniq ol. 'Counselor Toňukuk is evil'

Counselor Toňukuk evil IS

and

(II) A OL + Subordinate Clause $\implies$ 'A is that ...'

Example: Umid ol ... 'My hope is that ...'

Hope IS

How is it that a copula could be reinterpreted as a pronoun?

The key to understanding the phasing in and out of the pronoun/copula in Turkish is the use of zero morphemes. The copula is indicated by lack of a verb, where a verb would be found in any other sentence. Person marking and tense are affixed at the end of the sentence, which is the place the verb normally occupies.

Sentence (3) is an example of a present tense third person singular equative statement.

(3) Gok mavi.

sky blue

'The sky is blue.'

(Erguvenli 1984:9)
Example (4) demonstrates how the assertive statement ‘the sky is blue’ differs from the phrase ‘the blue sky’ in Turkish.

(4) mavi gok...
  blue   sky
  ‘the blue sky...’
(Erguvanli 1984.9)

Just as the null verb is the copula, so the null person is third person and the null number is singular.

In example (5) we see a full sentence in Turkish with no nouns or pronouns. Note that while the second and third persons display personal marking, the third person singular form is unmarked for person or number.

(5)
(a) gel-di-m
  come-past-1st sg.
  ‘I came.’
(b) gel-di-sin
  come-past-2nd sg.
  ‘You came.’
(c) gel-di
  come-past-3rd sg.
  ‘He came.’
It is widely held by scholars in the field of grammaticalization that person marking affixed to a verb derives from earlier independent pronouns. Following this line of reasoning, we might conclude that at an earlier stage, Turkish had second and third person pronouns, but it had no third person pronoun.

I do not rely on the above principle exclusively in positing an earlier stage without a third person pronoun. As it happens, even a synchronic analysis of Turkish leads to the conclusion that a third person pronoun is not essential, and that the need for it is so minimal that it could have arisen rather recently, in relative terms.

§6.5 The Origin of *on* in Turkish

Turkish, as an agglutinative language, has very little real need for either independent pronouns or the verb *to be*. The explanation for this is the idea of zero anaphora and the default form. Absence can be as linguistically significant as presence, since failure to morphologically mark a grammatical category will trigger the default category. In Turkish, the default person is third. The default number is singular. And the default verbal root is the verb *to be*. For purposes of this examination, it is sufficient to take note of the personal endings as tabulated in (6) and the tense markers outlined in (7). (Lewis:1967.136).

(6) The Turkish personal endings:

<table>
<thead>
<tr>
<th></th>
<th>SINGULAR</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>-im</td>
<td>-m</td>
<td>-eyim</td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>-sin</td>
<td>-n</td>
<td>-esin</td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>(-dir)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLURAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>-iz</td>
<td>-k</td>
<td>-elim</td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>-siniz</td>
<td>-niz</td>
<td>-esiniz</td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>(-dir)ler</td>
<td>-ler</td>
<td>-eler</td>
<td></td>
</tr>
</tbody>
</table>
In a sentence that contains no nominal subjects, pronominal subjects are expressed by affixing the pronominal endings to the verb or adjective in question, as in (6) above.

(7) The Turkish tense markers:

PRESENT FUTURE AORIST MIS-PAST NEC. DI-PAST COND. SUBJ.
-iyor -cek -ir -mis -meli -di -se -e

(Lewis 1967:136)

The tense markers are suffixed to the verbal root before any person markers. When tense marker appears without a verbal root, the default verb ‘to be’ comes into play.

(8) Present tense statements about being Turkish.

SINGULAR
1st Turkum ‘I am Turkish’
2nd Turksun ‘You are Turkish’
3rd Turk ‘He is Turkish’

PLURAL
1st Turkuz ‘We are Turkish’
2nd Turksunuz ‘You are Turkish’
3rd Turkler ‘They are Turkish’
In order to form the past tense of the verb *to be*, the past marker is affixed, as in (9) and (10).

(9) Turk-tu.
Turkish + di-past
‘He was Turkish.’

(10) Sorumlu-yu-nuz
responsible + di-past+ 2nd. pl. affix
‘You were responsible.’

Given that pronominal information is so neatly catalogued affixes in Turkish, we can easily envision a time in the history of the language when independent pronouns did not exist. They did not exist because all of the deictic functions which pronouns fill were expressed more verbally. Thus in (11) below, *ol* was originally seen as a verb, but later perceived as a nominal.

(11) Kayu ol?
What OL?
‘What is that?’

You can parse (11) as either ‘What COPULA NULL-3RD?’ or ‘What NULL-COPULA 3RD-DEMONSTRATIVE?’ The same morpheme can serve either purpose, but not both.
However, with a zero morpheme for the verb \textit{to be} and another zero morpheme for the third person singular, we are presented with the following dilemma: The Turkish for ‘He is’ would be null. And it is from this idiosyncracy that the present third person singular pronoun in Turkish arose.

While Turkish does not use a copula for equative sentences, it nevertheless does have a verb ‘to be’ which can be fully conjugated. That verb is \textit{olmak}. Its root is \textit{ol}. It means roughly ‘to be, to become, to happen, to mature’. Examples (12) and (13) illustrate the distinction between the use of the tense and person markings alone and the use of the root verb \textit{ol}. (Lewis.1967.141-142)

(12) Bir zelzeleyi-di.
\hspace{1cm} \text{one earthquake + di-past}
\hspace{1cm} ‘It was an earthquake.’

(13) Bir zelzele ol-du.
\hspace{1cm} \text{one earthquake ol+ di-past}
\hspace{1cm} ‘An earthquake occurred.’ (‘There was an earthquake’)

From this we can see that the \textit{ol} version of \textit{to be} is more lexical than the zero copula. It carries more semantic significance than mere equativeness. And yet \textit{ol} is the only variant available for exhortation and general discussions of “being”, such as in (14).

(14)
(a) Olunuz!
\hspace{1cm} \text{ol+ 2nd pl.}
\hspace{1cm} ‘Be!’
(b) Olasi.

\[ ol+ \text{cond.} \]

'May he be.'

(c) Olamam.

\[ ol+\text{neg+ 1st person sg.} \]

'I cannot be.'

With the verbal root \textit{ol} in mind, we turn once more to the question of the earlier situation in Turkish, when there were no third person pronouns. How would one say \textit{He is}? The possibility that comes readily to mind would be \textit{ol}, consisting of the verbal root together with the null third person singular ending. Lewis (1967.67) confirms this when he notes that "in pre-nineteenth century texts" the usual form of the third person singular absolute was \textit{ol}.

The earlier pronoun's identity with the equative verb is a clear indication that the verb came first, while the pronoun is an innovation.

\textbf{§6.6 Reanalysis and Grammaticalization of \textit{on}}

By internal reconstruction we may posit the development of the verbal root into an independent pronoun in Turkish that must have followed the pattern indicated in (15) below. (Note that \textit{on} is the Turkish distal demonstrative as well as a pronoun. We may assume that in Turkish, as in many other languages, the path followed was from demonstrative to pronoun. This will be discussed in more detail in Chapter 8.)
(15) Answers in response to the question “Who is Turkish?"

(a)
Turkish: O1.

NATIVE SPEAKER’S ANALYSIS: [o1+ null 3rd sg. marker .]

English gloss: ‘He[that] is.’

(b)
Turkish: O1.

NATIVE SPEAKER’S ANALYSIS: [3rd sg. pronoun + null verb to be]

English gloss: ‘He[that] is.’

The reanalysis described above is something that must have taken place very early in the pre-history of the Turkish language.

§6.7 From Archaic to Current Usage of the Pronouns

In Modern Turkish, many sentences still make no use of independent pronouns. But the reliance on the pronouns has grown. In (17), below we contrast situations in which a pronoun is the correct answer to the question ‘Who is happy’ and those in which a verb is the answer to leading question to the same effect.

(17) (a) Q: Kim mutlu?

who happy

‘Who is happy?’
(18)

(a)

(i) Ben mutluyum
    I happy-1st
    ‘I am happy.’

(ii) Ben.
    I
    ‘I am.’

(iii)*Mutlyum [Could answer: mutlusun mo?]
    happy-1st [Could answer: happy-2nd interrog.]
    *‘I’m happy.’[Could answer: ‘Are you happy?’]

(b)

(i) Sen mutlusun
    Thou happy-2nd
    ‘You’re happy’

(ii) Sen.
    Thou
    ‘You are.’

(iii)*Mutlsun [Could answer: mutluyum mo?]
    happy-2nd [Could answer: happy-1st interrog]
    *‘You are happy.’[Could answer: ‘Am I happy?’]

(c)

(i) O mutlu.
    He happy
(17) (c)(ii) O.

He.

'He is.'

(iii)*Mutlu. [Could answer: mutlu mo?]

happy [Could answer: happy interrog.]

*'He's happy.' [Could answer: 'Is he happy?']

Notice that the response without the pronoun is acceptable as a mere affirmation of the question: an alternate to saying 'yes'. This is the same mechanism that we saw in Chapter Three which allowed the Chinese copula shi to be used as an affirmative particle. In these affirmative responses both in Chinese and in Turkish, the verbal element was repeated. The difference in the Turkish example is that the descriptive word is the verb. O is the pronoun.

The Turkish word order is SOV. The final element is the verb. Hence in a phrase such as (17)(c)(i), there is no question that o is functioning as the subject of the sentence and as a pronoun. We might ask ourselves what was the mechanism that allowed o, originally a verb appearing sentence finally, to make the transition to a nominal, appearing sentence initially.

The answer is to be found in the variant represented by (17)(c)(ii). Under this discourse context, namely the question 'Who is happy?', the answer o, or earlier ol, means 'he is.' But this is a portmanteau for both the verbal concept of 'being' and the nominal concept of 'a being' or 'that one there'.

The flicker for speakers in the interpretation of o under this context is between seeing it as a variant for naming the happy one (as in 'John is') and asserting the continuation of the happy state 'that one is being happy'.
We have already seen above (Clauson 1972:123) that the word order of attributes and verbals was different in the earlier language:

(18) Bilge Toňukuk aňig ol.
Counselor Tonukuk evil IS
‘Counselor Tonukuk is evil’

In response to the question ‘Is Counselor Tonukuk evil?’, the affirmative answer would have been Ol: ‘he is.’ The modern response would be aňig: ‘he’s evil’.

The phasing in and out occurs when we have less than a completely filled in sentence, which is the normal state of events in a discourse context. Then a monosyllabic word meaning essentially ‘he is’ can very easily flicker between a reading of ‘is’ and a reading of ‘he’. We can have a null copula or a null third person pronoun, but not both in the same phrase.

§6.8 Implications for Grammaticalization Theory

The analysis presented above extends the accepted notions of directionality of grammaticalization. If a copula can develop from a pronoun and a pronoun from a copula in different languages, then there is nothing in principle that would prevent the same reanalysis from occurring in each direction at different points in the history of a single language, with the result of the changes forming a cyclical pattern. We may find that the directionality of grammaticalization is inextricably intertwined with developments in language typology. If that is the case, it may be that current notions of grammaticalization are based on the historical study of languages that are in the
process of changing to SVO word order. Perhaps directionality of grammaticalization of individual morphemes is closely related to the present state of a language and its currently active tendencies toward change.

The fact that a process of grammaticalization from pronoun to copula is observed in one language family, while in another language family we see grammaticalization in the opposite direction is enough to establish that this type of change is not unidirectional. We do not have to find the full cycle within the same language. If we determine that there is a correlation between language typology and direction of change, we may even suggest that at an earlier stage of the same language, a similar transformation took place in the opposite direction, if the typology and tendency of the language created a situation that was ripe for such a change.

The fact that an old independent monomorphemic lexical item has been recycled as another independent lexeme is perhaps surprising in view of the linear, unidirectional expectations for the process of grammaticalization. We would expect to see a great deal more semantic bleaching and phonological reduction taking place here. Likewise, the form is not only independent, it can be stressed for emphasis. This would lead one to assume that o has lexemic status. On the other hand, it does seem that in the process of re-categorialization o has been somewhat bleached. The pronoun has less semantic content than its source verb. Some of the other, substantive meanings of ol-, such as 'to happen, to become, to mature' have been discarded. All that remains is the relational meaning that all pronouns and copulas convey. In that sense, the abstraction that we come to expect as a product of grammaticalization has been effected.

The modern verb olmak retains the morpheme ol- as a bound form. Modern
Turkish does not use this form in the present (where it would have no affix), so it is usually found with an affix, such as *oldu* 'he/she/it was' or *oldusunuz* 'you were'. But the pronoun *o* is a free form, appearing by itself. True, it can carry case marking, but it often appears in the nominative with no affix at all. That is something the verbal *ol*- does not do.

Are independent grams less grammaticized than bound forms? Some (Bybee 1991:44-45) have made that claim. In the case of Turkish, its bound morphemes, like the root *ol*-, are older than its independent forms, like the pronoun *o*. But the older form is not more grammaticized, it is less so, for it has undergone fewer cycles of change.

Another way of looking at the issue is to say that the pronoun is more lexical than the verb. But there is no contradiction between lexical status and the fact that a morpheme has undergone grammaticalization. The verb *olmak*, and each of its inflected forms, are just as much lexical as the pronoun *o*. The lexical meaning of *olmak* is ‘to be’. The lexical meaning of *o* is ‘he/she/it/that’. By the functor/contentive test, *o* is a little more functional than *ol*. According to the dichotomy of closed and open sets, we probably would find that *o* has a higher occurrence rate than any form of *olmak*. But we rejected all of these test in Chapter One. All that interests us here is which of these forms has undergone more cycles of grammaticalization. The answer is: the pronoun.

And having been grammaticalized to such an extreme degree, *o* bears the scars of the experience. It is so small and reduced that in its most ordinary manifestation it consists of a mere vowel. And yet it represents the highest degree of isolation to be found in an extremely synthetic language. It is a free form.

I would suggest that this is not an unusual outcome. It is the normal result of the
grammaticalization process. For every pronoun that becomes an affix under the stress of language change, there is another pronoun that breaks free and becomes independent. Free forms come from bound forms and bound forms come from free forms. It's an endless cycle.
Chapter 7

Biblical Hebrew Third Person Pronouns and Verb ‘to be’

In this chapter I will show that the evolution of the third person singular pronouns from pre-Proto-Semitic to Biblical Hebrew took the path described in Figure 1, below:

\[
\begin{align*}
&x-j/x-w \ (xaja/xawa) \ 'live' \\
&x-j.x-w \ (xaja/xawa) \ 'live' \quad h-j/h-w \ (haja/hawa) \ 'be' \\
&\quad \vdots \quad h-j/h-w \ (haja/hawa) \ 'be' \quad hu/hí \ 'he/she'
\end{align*}
\]

The argument will consist of the following comparisons:

(1) form and function of the third person singular pronouns and the verb ‘to be’ in Biblical Hebrew

(2) form and function of the verb ‘to be’ and ‘to live’ in Biblical Hebrew

(3) form and function of ‘to live’, ‘to be’ and the third person pronouns in other Semitic languages.
The material in this chapter is a contribution to knowledge in the following ways:

(1) by establishing the genetic relation between the ancient Hebrew pronouns and the verb 'to be'

(2) by pointing out the link between the Semitic verb 'to live' and the shared Semitic third person pronouns

(3) as a study of yet another situation where a copula becomes a pronoun

(4) as constituting, together with Chapter Four, supra, a study of a complete cycle from verb to pronoun back to verb once more, within the various stages of a single language as it develops.

The arguments that follow rely in some measure on a phenomenon already discussed in Chapter Four: the flicker between a nominal and a verbal interpretation of the copular morpheme. We will even be discussing some of the same clauses. The difference will be one of directionality. Before, I argued that the instability in word order, verbal and nominal interpretations of participles, and fluidity of subject/predicate versus predicate/subject templates, were some of the factors that led to the reinterpretation of a pronoun as a copula. These very same factors, working in the opposite direction, were involved in the reinterpretation of a form of the copula as a pronoun. The main difference is that the trend in pre-Biblical times was from an overt to a null copula. The morphology formerly attributed to the overt copula was reinterpreted as a pronoun. The change from Biblical Hebrew to Modern was from a null to an overt copula. The material recruited is that very copular morpheme, now serving as a pronoun. The same morphology, only somewhat altered, is involved in both cases. It fills a similar function in both cases. But its categorization as nominal or verbal is determined in relation to the shifting grammatical framework of the linguistic system of which it is a part.
§7.1 The Hebrew Third Person Singular Pronouns and the Verb ‘to be’ are Related

In this chapter we will see that the Hebrew verb ‘to be’ רוחה/רוחיא [hawa/haja] is related to the third person singular pronouns רוחה/רוחיא [hu/hi] and that the root from which the pronouns and the verb ‘to be’ synchronically derive consists of the following suppletive bi-radical pairs \( \text{ḥ-}
\)\( \text{ḥ-} \)-. When the root appears, the second radical is sometimes ‘w’, sometimes ‘j’. It will be seen that verb רוחה/רוחיא [hawa/haja], ‘to be’, developed historically from the verb רוחה/רוחיא [xawa/xaja] ‘to live’. Hence the progression is from a semantically stronger verb, ‘to live’, to a more bleached meaning, ‘to be’, to a pronominal interpretation of a form of that verb as a third person pronoun.

§7.1.1 Internal Reconstruction

There are two traditional methods for proving historical developments: internal reconstruction and comparative reconstruction. I will begin with an internal reconstruction, which is the most common method in grammaticalization studies.

Once the internal analysis is completed, I will also apply the comparative method as a means of double checking the results. Any discrepancies between the results will be examined and resolved.

§7.1.1.1 In Hebrew, Roots Consist of Consonants

In section 7.1 I referred to the third person pronouns as deriving synchronically from the same root as the verb ‘to be’, while I said that the verb ‘to be’ developed historically from the verb ‘to live’. The use of the term ‘synchronously’ and ‘historically’ was intentional and significant.
A synchronic analysis of Hebrew, whether it be Modern, Mishnaic or Biblical, reveals three levels of word formations: bare roots, derived words and inflected words. Derivation takes us from the root, consisting only of consonants and a broad semantic range, to a word with a grammatical category and specific lexical meaning. Inflection takes the derived word and puts it through its paces in the grammatical category in which it belongs, conjugating or declining, as the case may be.

<table>
<thead>
<tr>
<th>ROOT ==&gt; consonants e.g. '-m-d 'stand'</th>
</tr>
</thead>
<tbody>
<tr>
<td>DERIVATION</td>
</tr>
<tr>
<td>LEXEME ==&gt; consonants+vowels</td>
</tr>
<tr>
<td>e.g. a-o or e-a</td>
</tr>
<tr>
<td>+</td>
</tr>
<tr>
<td>'-m-d</td>
</tr>
<tr>
<td>'+'</td>
</tr>
<tr>
<td>'-md-</td>
</tr>
<tr>
<td>'amod</td>
</tr>
<tr>
<td>'emda</td>
</tr>
<tr>
<td>'stand'(verb) 'stance, position'(noun)</td>
</tr>
<tr>
<td>[active, inf abs] [feminine, sg., abs.]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INFLECTED FORM ==&gt; w/ affixes, vowel &amp; euphonic changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g. a-a-ti or e-o-tai</td>
</tr>
<tr>
<td>+</td>
</tr>
<tr>
<td>'-m-d or '-m-d</td>
</tr>
<tr>
<td>'amadti</td>
</tr>
<tr>
<td>'emdotali</td>
</tr>
<tr>
<td>'I stood' 'my positions'</td>
</tr>
<tr>
<td>[perfect, 1st, sg] [contract, pl w/ 1st sg poss]</td>
</tr>
</tbody>
</table>

FIGURE 2
There are different schools of thought on the status of roots as synchronically viable linguistic units. Lambdin (1971.18) takes the following position: "Note that the root is a grammatical abstraction from the given words and not \textit{vice versa}; that is, because a root has no existence apart from its incorporation into words, it leads to misunderstanding the nature of language to say that words are derived from the root."

My own approach to the subject is more in accordance with traditional Hebrew grammar. While it is certainly true that roots are abstractions that have no existence outside derived words, derived words are likewise abstractions that have no existence outside inflected words, inflected words are abstractions that have no existence outside an utterance, and utterances are abstractions that do not exist outside a discourse context. Language is full of such abstractions. As linguists it is our job to study them.

In Hebrew, the derivational level is very productive. It is accessible to speakers. A speaker of Hebrew can identify the root of most words in his vocabulary, he can point out other words that come from the same root, and he is capable of deriving new lexemes from the roots in his radical lexicon.

Admittedly, even in a language with an active derivational system, the process of inflecting occurs more often than deriving. Cognitive scientists speculate that many derived words in the active vocabulary of a speaker are stored already derived. But even the process of actually inflecting a word is rarer than a purely formal approach would assume. Frequently used words are stored already inflected. These cognitive strategies for processing language in no way negate the usefulness or validity of describing the grammar of a language into terms of derivation from roots or inflection of derived forms.

The above discussion is particularly relevant to a work on grammaticalization,
since here we are trying bridge the gap between synchronic and diachronic linguistic processes. A synchronic derivation may be a product of historical change, but it is not opaque to the speaker, and the speaker is capable of independently replicating the historical process, which to him is nothing but a grammatical rule.

The difference then, between synchronic derivation and historical development from roots is as follows: a synchronic derivation does not violate the phonemic rules of the language. A historical development admits the application of sound change rules, including mergers and splits.

Hence, on the consonantal and radical level, הראזרא [hu/hi] derives from the same root as הינריר [hawa/haja] and such derivation is synchronic, since both use the consonants /h-j/-/h-w/ as the root. On the other hand, הינריר [hawa/haja] can be said to have developed from חנעריר [xawa/xaja] ‘to live’, only through a historical process, since it requires a phonemic split, as in Figure 3, below.

```
  /x/
 /x/  /h/
```

**Figure 3**

On the vocalic and grammatical level, even the derivation of pronouns הראזרא [hu/hi] is not synchronic, since the template into which the root is placed is not readily identifiable. (For instance, participles from the root would be howe, howa or haj, haja, according to the existent patterns. There is no synchronic pattern that would produce the exact result of hu or hi.)

For this reason, speakers of Hebrew do not recognize the connections between the
pronouns and the verb ‘to be’, and the pronouns cannot be said to synchronically derive from the verb ‘to be’. That derivation is diachronic. On the other hand, the root from which the pronouns and the verb derive is synchronically composed of the same phonemes. Hence the pronouns do synchronically derive from the same root as the verb ‘to be’, but the process by which this derivation took place is historical. Figure 4 illustrates these distinctions.

\[
\begin{align*}
\text{RADICAL LEVEL} & \quad /h-w/-/h-j/ \\
\text{ derivation fuzzy} & \quad \quad \text{ derivation regular} \\
\text{LEXICAL LEVEL} & \quad [hu/hi] \quad [hawa/haja]
\end{align*}
\]

Figure 4

7.1.1.2 The Root of the Verb ‘to be’ Has Two Alternates: /h-j/ and /h-w/

In Northwest Semitic, an alternation of /j/ and /w/ is not altogether unheard of. For instance, according to Ginsberg (1971.102), Proto-Semitic \(w\) in initial position is represented in Northwest Semitic by \(j\). So warada in Arabic meaning ‘he approached’ corresponds to Hebrew jarad ‘he went down’, and Akkadian \(w-s-b\) corresponds to Hebrew \(j-s-b\) ‘sit’.

Within Hebrew itself, a few roots manifest this alternation. In (1) below we see such an example:
(1)

\[ \text{walad\ 'jeled} \]

(a)

\text{Genesis, 11:30}

\[ \text{watashi Sarai 'aqara, 'ejn la walad} \]
and was Sarah sterile not to her child

\'(And) Sarah was barren; she had no child.'

(b)

\text{Isaiah 9:5}

\[ \text{ki 'jeled julad lanu, ben nitan lanu} \]
for child born (caus.) to us, son given to us

For a child was birthed for us, a son was given to us.

However, within the body of Hebrew proper, it is unusual that there should be such an alternation. It is difficult to find examples of this alternation, other than the one in (1), the root \[ /w-d^-/-/j-d^-/ \] ‘to know,’ and that shared by the third person singular pronouns, the verb ‘to be’ and the verb ‘to live’, and other words derived from those roots. It is a notable idiosyncracy.

Figure 5, below, shows the alternation at play within the simple active conjugation of the verb ‘to be’. For purposes of emphasis, I have placed the root consonants in caps.
<table>
<thead>
<tr>
<th></th>
<th>IMPERFECT</th>
<th>PERFECT</th>
<th>IMPERATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ST</td>
<td>1st sg</td>
<td>'eH:Je</td>
<td>HaJiti</td>
</tr>
<tr>
<td></td>
<td>masc</td>
<td>tiH:Je</td>
<td>HaJita</td>
</tr>
<tr>
<td></td>
<td>2nd sg</td>
<td>tiH:Ji</td>
<td>HaJit</td>
</tr>
<tr>
<td>2ND</td>
<td>fem</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2nd sg</td>
<td>jih:Je</td>
<td>HaJa</td>
</tr>
<tr>
<td>3RD</td>
<td>masc</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3rd sg</td>
<td>tiH:Je</td>
<td>HaJ:ta</td>
</tr>
<tr>
<td>1ST</td>
<td>1st pl</td>
<td>niH:Je</td>
<td>HaJinu</td>
</tr>
<tr>
<td>2ND</td>
<td>masc</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2nd pl</td>
<td>tiH:Ju</td>
<td>HeJitem</td>
</tr>
<tr>
<td></td>
<td>fem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2ND</td>
<td>2nd pl</td>
<td>tiH:Jena</td>
<td>HeJiten</td>
</tr>
<tr>
<td>3RD</td>
<td>masc</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3rd pl</td>
<td>jih:Ju</td>
<td>HaJu</td>
</tr>
<tr>
<td>4RD</td>
<td>fem</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3rd pl</td>
<td>tiH:Jena</td>
<td>HaJu</td>
</tr>
</tbody>
</table>

**FIGURE 5**

Regularized Paradigm of *haja*
(Standardized Biblical)

**Note:** For purposes of this very cursory view, ' stands for a schwa, while *xatafs* are not distinguished from their corresponding full vowels.
7.1.1.3 The Root of Third Person Pronoun has Two Alternates: /h-j-/ and /h-w/

In the pronouns, the alternating radicals waw \w\ and yod \j\ that we have seen in the paradigm for the verb 'to be' take on their vocalic value of [u] and [i] respectively. This vocalization of glides is a common phenomenon. In (2) and (3) below we see the root affiliation of two unrelated words with vocalic and consonantal values of these two radicals.

(2) root \m-w-t\ מות
mawet/mot/mut
dead(absolute)/death-of(construct)/die!(imperative)

(3) root \d-j-n\ דין
dajan/din
judge/justice

The masculine third person pronoun רה [hu], whose second radical is a waw \w\, has the vocalic value of its second root consonant. The feminine third person pronoun רה [hi], whose second radical is a yod \j\, has the vocalic value of its second root consonant.

In other Semitic languages, because the waw and yod are intervocalic in the corresponding forms, they take on their consonantal value, as in the case of the Arabic pronouns huwa and hija, 'he' and 'she'.
§7.1.1.4 The Orthography of the Feminine and Masculine Pronouns is Identical in Some Parts of the Bible.

For any who might suppose that the third person singular feminine and masculine pronouns derive from separate roots, I provide the following example of a spelling of the feminine pronoun with a waw, just like its masculine counterpart. The pointing‡, however, agrees with the normal pronunciation of [hi].

(4) Genesis 4:22

וְכִּלָּה גָּמֶּה יִתְיָדַּלָּה

and-Cila also-she bore

And Zillah, she, too, gave birth...

It is interesting to note that in many of the cases where the orthography for the feminine third person pronoun is identical to that of the masculine (and by extension, to the spelling of the copula, as seen in (4), above) the pronoun is in fact being used as a surrogate copula. See (5) below.

7.1.1.5 Some Alternate Orthographic Renditions of the Verb ‘to be’

The orthography of the third person pronouns in Hebrew gives the third letter of each as an alef, yet that is not actually part of the root, but is merely a mater lectionis, just as is the hey at end of הָאָדָם /hām/hāma/. The mater helps a reader of an unpointed text determine that the word ends in a vowel. It is not intended to represent a consonant and hence is not part of the root.

In fact, in a few unusual but persistent Biblical passages, the verb ‘to be’ is

‡ As a mater, yod can help with the vocalization of an [i] sound and waw can help with the vocalization of an [u] sound, but not vice versa. And yet this form of hi is pointed with a ziriq [i], while its mater is waw.
rendered with a final *alef*, rather than a *hey*. There are passages in the Old Testament where the consonantal orthography for pronoun and verb is entirely identical, with only the pointing distinguishing the two. See (5) below.

(5) **Job 37:6**

![Image of text from the page](image-url)

For to snow he says "be on earth". (Fall to earth)

Example (6) is even more intriguing, since it displays a form that, save for the third person imperfect prefix *[je]*, looks and sounds exactly like the third person pronoun. In other words, the *waw* is vocalic here, as in the normal pronominal form, but contrary to most verbal forms of the root.

(6) **Lamentations 11:3**

![Image of text from the page](image-url)

‘Wherever the tree falls there it will be.’

We have seen that neither the alternation between vocalic and consonantal values of the glides nor the use of the differing *maters* distinguishes the root of the verb ‘to be’ from the root of the third person pronouns. What is more, we have seen that the pronouns and the verb ‘to be’ share many of the same idiosyncracies.

**7.1.1.6 Biblical Hebrew Pure Identity Statements with Pronoun as Linker**

Not only are the pronouns and the copula similar and sometimes
identical in form, but their function is likewise very similar. Often, a pronoun is used as a linker between two participants.

In example (7) an identity is established between Efrat and Bethlehem by way of a feminine third person pronoun which happens to be spelled just like the masculine.

(7) Genesis 48:07

ba-derex 'efrat hi bejt laxem 7
in-way-of Efrat she house-of bread
'by way of Efrat which is Bethlehem'

In the following example, which we discussed earlier in Chapter Four, we see class membership or definition established with the help of a pronoun.

(8) Numbers 35:16

wa-'im bi-xli barzel hikahu wajamot
and-if with-instrument iron struck-him and-died
roce'ax hu
murderer he

'And if he struck him with an iron instrument and he died, then he is a murderer.'

If we take a look at the second (and simpler) clause in (8) we get roce'ax hu,
meaning, ‘murderer he’ or ‘he is a murderer’. Now, within this clause, which is our subject and which is the predicate? The clause is amenable to two readings, one in which the word for murderer is given its more verbal participial meaning and the pronoun is seen as the subject, as in gloss (9) (a), below. But, if we see murderer as a focused nominal, we are likely to decide that *hu* is the predicate, as in gloss (9) (b).

(9)  rocę'ax *hu*

murderer he

Gloss:

(a) Predicate Subject

murders he ==> He murders.

(b) Subject Predicate

a murderer he/is ==> (He) is a murderer.

Given the context of the full sentence in (8), (9)(b) is the more appropriate interpretation of the clause. The pronoun is being used for predication.

§7.1.1.7 In Both Form and Meaning the Verb ‘to be’ and the Third Person Singular Pronouns of Hebrew are Related.

The root is identical, consisting of the same root consonants, in the same order, with the same idiosyncrasies; the semantics is similar. Therefore, the third person pronouns of Hebrew and the copula are related.

We have seen above that:

(1) the pronouns and copula share the same root consonants
(2) idiosyncracies of the root are also shared:

(a) suppletion of /j-w/ radical

(b) vocalic/consonantal interplay /w-/u/

(3) both the pronoun and the verb 'to be' are used for predication and equative linking.

Having established that there is an unmistakeable link between the third person pronouns and the copula, we will now turn to a functional reconstruction of this relation, with especial attention to the development of the third person singular pronouns from the root /h-w-/h-j/.

§7.1.2 Functional Development

In this section we will see that the Hebrew third person pronoun arose out of a situation where the usual forms for the aspectless copula and the third person masculine pronoun were null. When the need for a deictic demonstrative arose, the material recruited was a nominalized form of the verb 'to be'.

In order to determine the path that development of the third person singular pronouns took from the root of the verb 'to be', a closer look at the morpho-syntactic system is in order, with special attention to default forms.

The default word order of Biblical Hebrew is considered to be VSO. A canonical VSO sentence is provided in (10) below.

(10) Kings II, 1:5

\begin{verbatim}
wa-jašuvu ha-mal'axim eš-l-aw

and-returned the-messengers to-him

'The messengers returned to him'
\end{verbatim}
Surface orders vary considerably, with emphasized participants occupying the initial slot. Sometimes the emphasis is a question of contrasting two or more participants, as in (11) below:

8. וַיִּנְבָּא הַנָּחָה לְלֹא הַנָּחָה לְאָלָּפִים

(11) Samuel I, 18:8 נִנְבָּא הַנָּחָה לְלֹא הַנָּחָה לְאָלָּפִים

נתָּנָה לוֹ דָּוִד רְבִּיָּו וַּיִּנְוָא הַנָּחָה לְאָלָּפִים

gave(3rd,pl) to-David ten-thousands and-to-me gave(3rd,pl) the-thousands

They gave David tens of thousands and me they gave [only] thousands.'

Here the contrastive emphasis is given in the second clause, where the indirect object is placed at the beginning of the clause, followed by the verb and the direct object. In (12), below, we see an example of contrastive emphasis between two objects, each of which is placed at the beginning of its clause.

(12) Judges 5:25 ישפיחו היה

מַעֲזַּל וַיִּנְּלָס נָתָנָה

water ask(3rd,m,sg,prf) milk gave(3rd,f,sg,prf)

He asked for water; she gave milk.

In other cases, the emphasis is not contrastive, but is based on newness or general discourse salience. So in (13), below, the objects are seen as more salient than either the actors or the actions. It is not a question of contrast, however, because the actions performed on each of the objects are parallel. The overall message is one of destruction, with each instance being an illustration of the same
policy.

(13) Kings II, 3:25

(a) וָהָאֲרֵ֣ים יָהָ֔רוּשָׁיִ֖ים
and-the-cities destroyed(3rd,pl,imprf)

and they destroyed the cities

(b) וַ֖א-גָּלְקַ֣ה חוֹבֶ֑ה יֵשְׁפַֽיָּהְ֖ו אָֽיִית אָֽבָנָֽיִית

and each plot good threw man his-stone and fill(3rd,pl,imprf)-her

and each man would throw stones into every good plot of land and fill it

(c) וַ֖א-גוֹלַ֑ה מַזֶּ֖ה מָזוֹן יָסַֽהְמוּ
and each spring water plug

and they plugged up every (water) spring.

(d) וַ֖א-גוֹלַ֑ה יִטוֹפְּלֶ֖ם יִפְטִֽיסָהּ
and each tree good fell((3rd,pl,imprf)

and they felled every good tree.

In the following example, the subject, which is placed at the beginning, is new information and represents a change of topic. In this case, it also serves to introduce a new participant into the discourse.

(14) Kings II, 4:1

וָּלְשָׁא אַמְתֵּי מִשָּׁה מִי-מְנֵבּוֹיִם לְצַעְקָה אֶלּ-רָֽיִשּׁא

and-woman one from-women-of sons-of the prophets shouted to Elisha

And one woman from the wives of the sons of the prophets called out to Elisha.
However, not every fronted subject involves a new participant. The following sentence appears in a passage that describes a conflict with Moab. The subject, while not mentioned by name in the previous sentence, was nevertheless a topic of the preceding discourse.

(15) **Kings II 3:21**

\[
\text{ופֵּשֵׁת-מִצְצֶה מִן-יוֹרֵעָה מַסִּים לְהַלְחִיא בַּמָּם}
\]

wa-xol Mo‘av šam‘u ki ‘alu ha-malaxim lohilaxem bam

and-all Moab heard that ascended the-kings to fight against-them

And all Moab heard that the kings arose to fight them...

We have seen that the initial slot can be occupied by subject, object, or verb. While verb-initial sentences are prevalent, the word order is very flexible. In Biblical Hebrew SVO or OSV word orders are no less grammatical than VSO. It is simply that the fronted item, whatever its part of speech, is seen as more salient or emphasized. Since verbs (or other predicates) often occupy initial position, we might conclude that information coded in the predicate is more often deemed by speakers to be especially salient. However, there is a device available to place special stress on the verb, beyond merely placing it in initial position. An infinitival form can precede the finite verb for added emphasis.

(16) **Genesis 37:33**

\[
\text{בראשית ופִּים}
\]

tarof taraf Josef

devour(inf) devour(3rd,m,sf,prf,pass) Joseph

Joseph has [indeed] been killed by a wild animal
The relatively free word order of Biblical Hebrew extends to clauses where the predicate is a nominal. That is, even a predicate adjective or a participle can appear either before or after the subject.

(17) Numbers 14:7

cemarur yir. 7

tova ha-arec me'od me'od

good the-land very very

The land is very, very good.

In (17), above, the predicate adjective appears initially. In (18) the subject is followed by the predicate adjective, in each of the two clauses.

(18) Jeremiah 24:3

ireha ber. 3

(a) ha-ta'enim ha-tovot tovot ma'od

the-figs the-good good very

The good figs are very good

(b) wa-ha-ra'ot ra'ot ma'od

and-the-bad bad very

and the bad [ones] are very bad.

In example (19) and (20) we see examples of variable placement of the participle. In (17), the participle ‘judging’ precedes the subject pronoun ‘I’.

(19) Samuel I, 3:13

samueli: 3, 13

wa-higadi lo ki sofet 'ami 'et bejt-o

and-tell(1st, sg, converted prf) to-him that judge(prtcp) I.o. house-his

And I will tell him that I'm judging his house.
Contrast the clause ‘judging I his house’ in (19) with ‘Jotham ...judging the people’ in (20). The first is VSO, the second SVO.

(20) Kings II, 15:5

וַיַּהוֹלֵם הִבְּנֵו מִלֶּכֶת הַבְּיִטֵּת שָׁפֵטֵת וְאָמַר הַאָרֶץ רְשֵׁי

wa-jotam ben ha-melex 'al ha-bajit šofet 'et 'am-ha'arec.

And Jotham, the king's son, was in charge of the house judging the people of the land.

In both (19) and (20) the participle has a largely verb-like usage. But the identical word can be used as a noun, meaning ‘a judge’, as in ‘one who judges’ or ‘a judge’. In (21), we see an instance of such a usage.

(21) Psalms 7:12

אֲלָהָם שָׁפֵטֵת צְרֵיךְ

God(pl) judge just

God is a just judge.

The identification of the participle as a nominal becomes even more vague when we add a pronoun to the clause when we already have an overt noun.

(22) Psalms 50:6

כִּי אֲלָהָם שָׁפֵטֵת הָא

for god(pl) judge he

For God is a judge.
Under contextually appropriate circumstances, example (22) could be glossed into English as 'for God is judging' or 'for God, he is a judge'. Figure 6 illustrates the various readings.

<table>
<thead>
<tr>
<th>GOD</th>
<th>JUDGER/JUDGES</th>
<th>HE/IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOUN IN APPPOSITION</td>
<td>NOMINAL A</td>
<td>NOMINAL B</td>
</tr>
<tr>
<td>NOUN IN APPPOSITION</td>
<td>VERB</td>
<td>SUBJECT</td>
</tr>
<tr>
<td>NOMINAL A (SUBJECT)</td>
<td>NOMINAL B</td>
<td>VERB (COPULA)</td>
</tr>
</tbody>
</table>

**FIGURE 6**

While there is no question that 'for God is a judge' is a good English gloss, the Hebrew words are resonant with the other connotations. That is, the interpretation of 'judger' and 'judges' and 'judging' coexists within the Hebrew grammar. A speaker doesn't have to decide that a particular participle is only nominal or only verbal. That distinction is not expressed morphologically, is syntactically indeterminate and need not be made semantically.
Because the Biblical Hebrew word order is so free, and because participles and adjectives can serve as either subject or predicate, Hebrew syntax leaves a great deal of free play in the area of grammatical category and perceived word order.

§7.1.2.1 Unmarked Time/Aspect was Present/Continuing

Biblical Hebrew has two full verb conjugations: perfect and imperfect. These two conjugations can be reversed by the addition of a waw conversive. An imperfect form with a waw conversive is a perfect, and a perfect with a waw conversive is an imperfect. These distinctions are useful in narratives, to give information on sequence of actions.

<table>
<thead>
<tr>
<th>Perfect: jelex</th>
<th>Imperfect: wa-jelex</th>
</tr>
</thead>
<tbody>
<tr>
<td>'will go'</td>
<td>'went'</td>
</tr>
<tr>
<td>Imperfect: halax</td>
<td>Perfect: we-halax</td>
</tr>
<tr>
<td>'went'</td>
<td>'will go, would go'</td>
</tr>
</tbody>
</table>

FIGURE 7

Verb *halax* in perfect, imperfect and inverses

Perfect forms indicate perfected action. They are often used to describe past occurrences. Imperfects indicate unperfected action, which includes present, future, habitual, and irrealis.
The full conjugations, perfect and imperfect, distinguish person, gender and number. Participles distinguish only gender and number, like all other nominals. Infinitives indicate only the verbal derivation (active, extremely transitive, passive, causative, passive-causative, and reflexive/reciprocal/middle). When no aspect is indicated, it may be inferred from discourse context. If context does not indicate otherwise, the clause is deemed to refer to imperfect aspect (or present time). Clauses that need have no overt aspect marking are of the following types:

(1) infinitival
(2) participial
and
(3) equative

Infinitival clauses are nominal forms, where the agent of the action is in a construct relation † with the infinitive. The infinitive is the first element of the construct chain, and the agent is the second element.

(23) Judges 5:2

בְּפַרְעַה בְּיִשְׂרָאֵל

bi-fr'ah pra'ot ba-yisra'el
in-riot(inf-cnstr) riots in-Israel

In rioting of riots in Israel $\Rightarrow$ When riots riot(ed) in Israel

† When two nominals are in construct, $A B$, the relation is one of possession, $A$ of $B$. The form of $A$ often changes slightly, and it is said to be in the construct, while $B$ is in the absolute and does not change. A construct chain may consist of more than two nominals, such as $A B C$. The last member of the chain is in the absolute, while all the others are in the construct ($A$ of $B$ of $C$).
In (23), the time attributable to the event is determined by the overall context of
the narrative, and not from morphology or syntax.

The item in construct with the infinitive can be a possessive personal suffix.
These suffixes are distinct from the independent pronouns and represent
grammaticalized versions of an earlier set of pronouns.

(24) **Psalms, 51:6**

... la-ma'an ticdaq ba-dovre-xa tizke be-softe-xa

in-order be-right(2nd,s,m,imprf) in-your-speaking win(2nd, m,s,imprf) in-your judging

'In order that you be right when you speak, be vindicated when you judge.'

Aspect (or time) can be indicated by adding a form of the verb 'to be' to the
clause, as in (25).

(25) **Ruth 1:1**

wajhi † b-imjm sofot ha-sofim ....

(and)-be(cvn-imprf) in-days-of judge(inf-const) the judges

It was in the days of the judging of the judges ==> In the days when the judges
judged.

Participles can be seen as nominals as well, but they follow a different pattern
from infinitives, since the agent of the action is in the absolute nominative, being
neither grammatically subordinate nor superordinate to the participle. Participial

† Converted perfect, from ḫle."
clauses can be seen as a special case of equative clauses.

(26) Proverbs 14:21
bash-la-re'ehu xote
cow-nir-neh omoa
scorn(prtp) to-friend-his sin(prtcp)
He who scorns his neighbor sins (or is a sinner.)

In equative sentences that are not marked for aspect, no verb whatever is required. These are the sorts of sentences that we discussed in Figure 1 of Chapter 2.

(27) Genesis 28:20
ba-de'ox ha-ze aser anoxi holex
on-road the-this which I walk
On this road that I walk (am walking)

§7.1.2.2 The Default Verb Was the Copula

The reason that no verb is required in equative sentences is that the verb ‘to be’ is the default verb. It must be expressed overtly only where it carries aspect marking. Otherwise, in the absence of a verb, the verb that is understood is the copula. When converting a verbless clause to either perfect or imperfect, the verb ‘to be’ in the appropriate conjugation must be inserted, as seen in (25) above.
§7.1.2.3 From Suffixed Verbal Morphology, Third Person Was Once the Default Person

In the perfect, which marks person by affixing person marking endings, the third person singular masculine is the base form, with no suffix added. Standard grammaticalization theory suggests that suffixes derive from previously independent words. If we assume that the suffixes were earlier pronouns, we may infer that third person was once the default person in an antecedent of Biblical Hebrew. Stated otherwise, this would mean that at some point in the pre-history of the language, there were independent second and third person pronouns, but there was no independent pronoun for third person singular masculine.

Zero anaphora for third person is a common crosslinguistic phenomenon. It has been explained by the fact that third person subjects are often designated by overt nouns, and when the noun is deleted, its discourse salience is so high that it need not be mentioned at all. On the other hand, people seldom refer to themselves by name and only rarely address their interlocutor by his sobriquet.

It has likewise been claimed that the reason for this phenomenon is that first and second person are particularly salient to speakers, since, barring a soliloquy, there must be both a speaker and a hearer, although they need not necessarily refer to anyone or anything other than each other.

Both of the above explanations, while containing some modicum of truth, leave much to be desired. Fortunately, we do not need to explain the etiology of the phenomenon to note that it exists.
§7.1.2.4 The Need for a Third Person Pronoun Developed in Unmarked Aspect

In previous sections we have established the following:

(1) Biblical Hebrew word order is very free
   (a) While VSO is said to be the default order
   (b) the actual order can be SVO, OSV or VOS as well
       and
   (c) many lexemes can be used as either subject or predicate
       so that
   (d) the actual word order found can be ambiguous and even
       semantically irrelevant.

(2) The default verb is the verb ‘to be’
   (a) equative clauses without express aspect marking are verbless
   (b) when verbless clauses are converted to a specific aspect, the verb
       ‘to be’ is used to carry the aspect marking

(3) At one point in the pre-history of the language, third person singular
    masculine was the default person

Just as in the Turkish example in Chapter 6, we have a situation where the third
person pronoun is null, the verb ‘to be’ in aspectless clauses is null, and an occasion
might arise for the need to make an equative or aspectless statement about a third
person singular participant that is not designated by an overt noun.

The lexical material that comes readily to hand is the root of the verb ‘to
be’, in either an infinitival (howo infinitive construct with third person singular possessive affix) or a participial form (howe, sg, m). Either howo or howe can have become reduced to hu under normal phonetic processes. However, I am not necessarily suggesting that at the relevant time in the prehistory of Hebrew, (which may in fact precede the formation of the separate Semitic languages) the infinitival or participial forms were constituted exactly as they are in Biblical Hebrew. Of the two forms, I favor the infinitive howo as a likely candidate for the source, because it is phonetically very close to hu. But all we can really suggest with any certainty is that the root recruited was the root of the verb ‘to be’, and the form was nominal.

The initial use of hu was probably as a demonstrative pronoun, with strong deictic characteristics. In (28) below, we see a context under which such a use might arise.

(28) Answers in response to the question “Who is the murderer?”

(a) Hebrew: [hu]  וה

NATIVE SPEAKER’S ANALYSIS: [‘to be’ root + 3rd sg. masc. marking.]

English gloss: ‘He is.’

(b) Hebrew: [hu]  וה

NATIVE SPEAKER’S ANALYSIS: [3rd sg. pronoun + null verb to be]

English gloss: ‘He is.’

The reanalysis of the nominalized verb as a demonstrative, and then a personal pronoun, was undoubtedly a gradual process. Because grammatical category in early Biblical Hebrew is extremely fluid, it is possible for the interpretation to have phased in and out for speakers, within their own speech, depending on specific context. With time, the form was phonetically reduced in its pronominal usage, and as it became opaque, it became dissociated from the verbal root that was its source. A split had been effected, and the independent third person pronouns attained their full status as grammatical units and part of the pronominal paradigm.
§7.2 Comparative Reconstruction

Hebrew is not an isolate. It is part of the Semitic language family. In testing the validity of the above analysis, we would do well to examine similar developments in sister languages.

A cursory examination of the pronouns in the other Semitic languages reveals that the third person forms are related. Reflexes of a Proto-Semitic verb ‘to be’ are somewhat more rare. This might initially lead to the conclusion that the verb ‘to be’ developed from the pronouns, and not vice-versa. However, when we consider that the verb ‘to be’ is a bleached form of the verb ‘to live’, the development that seems most likely would be an evolution from ‘live’ to ‘be’ to ‘he’, as is suggested in Figure 8, below.

<table>
<thead>
<tr>
<th>Ge'ez</th>
<th>Arabic</th>
<th>Hebrew</th>
<th>Aramaic</th>
<th>Akkadian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live</td>
<td>b-j</td>
<td>x-j-j</td>
<td>x-w/x-j</td>
<td>x-w/x-j</td>
</tr>
<tr>
<td>Be</td>
<td>h-l-l</td>
<td>k-w-n</td>
<td>h-w/h-j</td>
<td>h-w/h-j</td>
</tr>
<tr>
<td>He/She</td>
<td>we'etu/je'eti</td>
<td>huwa/hija</td>
<td>hu/hi</td>
<td>hu/hi</td>
</tr>
</tbody>
</table>

Figure 8

TABLE OF SEMITIC CORRESPONDENCES: 'LIVE' TO 'BE' TO 'HE'
§7.2.1. Comparative Evidence that the Verb ‘to be’ Preceded the Pronouns.

The following is evidence to the effect that the verb ‘to be’ derived from the root ‘to live’, in Hebrew, and by extension in the other Semitic languages.

(1) The root ‘to be’ is weaker, more bleached version of the verb ‘to live’ in both form and meaning.

(2) The alternation of a /w/ and a /j/ is a shared idiosyncracy.

In Hebrew, words from the root ‘to live’ display alternates in /w/ to the more normal /j/ pattern. Thus the word for encampment, being a place where one lives, is xawa, as well as the name ‘Eve’, who is so called because ‘she is the mother of all life’.

(26) Genesis 3:20

בראשית ל, 20

גְּקַקָּה קַנֵּקָה שֶׁ֔מֶּשׁ אָחָשָׁ֖יו מִי קֵמָֽוָ֖ה אֲשֶׁר כִּֽיָּ֑צָּֽה

wajiqra ha’adam sem isto xawa ki hi hajta ‘em kol-xaj

and called the-man name wife-his XAWA because she was mother of all-live

‘And Adam called the name of his wife Eve because she was the mother of all living things.’

Notice that in the above verse, reflexes of the proto-Semitic root for ‘life, being’ occur in four different words:

(1) xawa -- ‘Eve’, in which the /w/ alternative is realized, in both orthography and pronunciation,

(2) hi -- ‘she’, in which the orthography is /w/, but the pronunciation is
indicated as the vowel [i], (the pointing and the consonants are at odds here, as they are in all cases of the third person feminine independent pronoun in Genesis.)

(3) *haj̪a* -- ‘was’, in which the */j/* alternative is realized, and

(4) *xaš* -- ‘live’, an adjective, in which the */j/* alternative is realized.

Turning to other Semitic languages for reflexes of ‘to live’, Arabic, and Ge’ez display the */j/* alternate, Aramaic shows both alternations, and Akkadian has no attested cognate.

When we examine reflexes of ‘to be’, Arabic and Ge’ez have no attested cognate, Hebrew and Aramaic have both */j/* and */w/* forms, and the Akkadian reflex has a */w/*.

When it comes to the pronouns, however, all six languages have both a */w/* and a */j/* version, the first being associated with the masculine and the second with the feminine.

The plentitude of attested forms for pronouns, and the scarcity of forms of the verb ‘to be’ from this root might lead us to believe that the pronouns preceded the verb ‘to be’. I would argue against this conclusion as follows:

(1) The process of semantic bleaching leads us from ‘live’ to ‘be’. It seems very unrealistic to suppose the opposite progression from ‘be’ to ‘live’. Likewise, it is difficult to imagine ‘he’ being derived directly from ‘live’. A more abstract intermediate stage of ‘be’ seems called for.

(2) Assuming that all of this progression from ‘live’ to ‘be’ to ‘he’ took place before the languages split up, the older forms would be more likely to disappear and be replaced, not the more recently evolved independent pronouns.

(3) The first and second person pronouns all seem to derive from a proto-Semitic
root of *₇-n-(t), (Moscati 1964:102), which is entirely unrelated to the third person pronouns. Third person is thus a privileged category, supplied from a common, but different source.

(4) The perfect conjugation shows a default or null third person singular, indicating that at the VS stage of person marking, there was no third person singular pronoun employed.

(5) The verb 'to live' has a full paradigm for verbal conjugation in all the languages in which it is attested. Likewise, 'to be', where it occurs. The more opaque form of the third person singular pronouns, which does not undergo normal inflection, is an indication that it is a newer form which has undergone more grammaticalization, and is more reduced. As a newer form, it has not had the opportunity to gather other lexemes to become affixed to it and grammaticalized to it, such as inflectional morphology. So we find that in all the attested languages, the object pronouns are formed from a different root. This is entirely consistent with our observation, in Chapter Two, that the copula tends to conjoin two nominatives, and that neither participant in the equative clause is subordinated to the other.

The grammaticalization cycle as we now describe it has effected the following change:

'To live' --> 'to be' --> 'his being' --> 'he'

When we look at our end product, it appears reduced, opaque, independent and very abstract as any end product of a long process of grammaticalization must be. It is so old that it has been polished of all its chinks and its creases, like a pebble in a brook, whose roundness and shine might induce the unwary passerby to suppose it to
be brand spanking new. But it is the centuries of wear that have brought it to its present state. It is so old, it might as well be new. And it is ready for another cycle.

§7.3 Conclusion

In this chapter we have seen that from pre-Proto-Semitic to Biblical Hebrew

(1) the verb ‘to be’ developed from the verb ‘to live’
(2) a form of the verb ‘to be’ developed into the third person singular pronouns

and

(3) the following factors contributed to reanalysis of a nominal form of the copula as a pronoun:
   (a) fluid word order
   (b) dual nature of participles as both nominals and verbal elements
   and
   (c) the special status of the verb ‘to be’ as present by inference in all verbless clauses.

When viewed in conjunction with the material in Chapter Four, the findings here constitute one half of a cycle: the evolution from copula to pronoun to copula.

As I stated in Chapter One, it is not necessary to find the full cycle within a single language in order to refute claims of uniformity of result. The fact that Chinese has a pronoun developing into a copula, and that Turkish has a copula developing into a pronoun is sufficient to show that this sort of progression does not always go in the
same direction. Furthermore, when the tendency to change form copula to pronoun or vice versa is linked with typological attributes and developmental trends, we can then predict that the same language might apply the opposite strategy to that which we have documented, under appropriate typological pressures.

However, the argument is even more compelling, when we have an example of the a uniform process of unidirectional grammaticalization resulting in an apparent reversal of effect in different periods within the same language. Hebrew gives us a full cycle.
Chapter 8

Typological and Theoretical Implications of the Findings

§8.1 Introduction

We have now examined the data which are the subject of this dissertation. We have seen the cognitive and crosslinguistic metaphors that underlie the use of the copula. We have seen instances of grammaticalization in five different languages involving the evolution of a copula from a pronoun and vice versa. We have even examined a full cycle from copula to pronoun to copula in the evolution from Pre-Biblical to Biblical to Modern Hebrew.

There are two areas I will now briefly explore:

(1) Typological traits of languages that undergo the change

(a) from copula to pronoun

and

(b) from pronoun to copula.

And

(2) The implications for grammaticalization theory from these findings.

§8.2 Typological Survey

In the course of this dissertation, we have observed the change from a pronoun to a copula in the following languages:
(1) Classical Chinese to Modern Chinese
(2) Pre-attested Finnish to Modern Finnish
and
(2) Biblical Hebrew to Modern Hebrew.

We have seen a copula change into a third person pronoun in the following languages:

(1) Ancient to Modern Turkish
and
(2) Pre-Proto-Semitic to Biblical Hebrew.

If we are dealing with a crosslinguistic tendency, perhaps it would be useful to survey the typological characteristics that accompany a change with one of the two from copula to pronoun, or vice versa.

Admittedly, this is a very small sampling of such occurrences, and it may be entirely speculative to assume that any preliminary conclusion we may draw from the observation of five such changes will hold true for other languages in which the pronoun and the copula are intermittently reanalyzed each as taking the place of the other.

However, it may be entirely appropriate to attempt to formulate a well-reasoned hypothesis on this subject, with the full understanding that later findings may prove it false.

With this in mind, let us then turn to the smaller sampling first: the change from
copula to pronoun. I will give the characteristics of the target languages (Modern Turkish and Biblical Hebrew), not of the source languages (Ancient Turkish and Pre-Proto-Semitic), since the source languages here are unattested and shrouded in antiquity.

§8.2.1 Copula to Pronoun: Turkish and Biblical Hebrew

Modern Turkish has:
(1) a flexible SOV word order,
(2) sparing use of independent pronouns,
(3) a strong agglutinative system,
and
(4) a null present tense copula.

Biblical Hebrew has:
(1) a flexible VSO word order,
(2) sparing use of independent pronouns,
(3) a strong inflectional system,
and
(4) a null present (or aspectless) copula.

The significance of the word orders is not so much in their actual configuration, but rather in the flexibility afforded by the structure. Case and role are not determined by word order. The morphological system takes care of that. The prevalence of zero anaphora and sparing use of independent pronouns may be a
vestige of an earlier era when independent pronouns were non-existent, or when the current verbal person marking was still interpreted as a set of pronouns. The null copula seems to be an essential part of the linguistic environment where a pronoun evolves into a copula.

Turkish and Biblical Hebrew differ in the sort of case marking that accompanies third person pronouns. In Biblical Hebrew, only the third person subject pronouns come from the 'to be' root, while the object pronouns have the direct object marker as their root, and all other obliques are simply prepositions with possessive pronominal suffixes. See Figure 1.

<table>
<thead>
<tr>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
</tr>
<tr>
<td>'ani/'anoxi</td>
<td>'otix</td>
<td>li</td>
</tr>
<tr>
<td>'otx</td>
<td>lexa</td>
<td>bexa</td>
</tr>
<tr>
<td>'otax</td>
<td>lax</td>
<td>bax</td>
</tr>
<tr>
<td>m</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>'atu</td>
<td>'at</td>
<td></td>
</tr>
<tr>
<td>'otx</td>
<td>lax</td>
<td>bax</td>
</tr>
<tr>
<td>m</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>'ut</td>
<td>'atu</td>
<td></td>
</tr>
<tr>
<td>'oto</td>
<td>lo</td>
<td>bo</td>
</tr>
<tr>
<td>m</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>'hu</td>
<td>'hi</td>
<td></td>
</tr>
<tr>
<td>'oto</td>
<td>la</td>
<td>ba</td>
</tr>
</tbody>
</table>

Figure 1

On the other hand, the Turkish third person pronouns are all built on the 'to be' root. See Figure 2.
<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>abs.</td>
<td>o</td>
<td>onlar</td>
</tr>
<tr>
<td>acc.</td>
<td>onu</td>
<td>onlari</td>
</tr>
<tr>
<td>gen.</td>
<td>onun</td>
<td>onlarin</td>
</tr>
<tr>
<td>dat.</td>
<td>ona</td>
<td>onlara</td>
</tr>
<tr>
<td>loc.</td>
<td>onda</td>
<td>onlarda</td>
</tr>
<tr>
<td>abl.</td>
<td>ondan</td>
<td>onlardan</td>
</tr>
</tbody>
</table>

**Figure 2**

(Lewis 1967:67)

The reason for this difference can be found in the overall structure of the languages themselves. Turkish has case marking suffixes. Hebrew has a preposed direct object marker and prepositions. Originally, the Turkish absolute case of the third person pronoun was suppletive, while the oblique cases all came from a different root. (See Chapter Six, example (2)). But because of the strong motivation in Turkish for agglutinative regularity, the paradigm was levelled. *Ol* in the oblique cases became *on*, merging with the *an* morpheme previously used.

An area where both languages display a similar pattern is in the third person pronoun as demonstrative. Both Turkish and Hebrew probably began to use the third
person pronouns(s) as deictic demonstratives first, and then generalized the demonstratives as personal pronouns.

Hebrew has a two way opposition between 'this' and 'that'. Turkish marks a three way difference. (Lewis 1967.71) Note that both Hebrew and Turkish developed their third person pronoun from the distal demonstrative, not one of the proximals. See Figure 3.

<table>
<thead>
<tr>
<th>this (proximal)</th>
<th>this/that (medial)</th>
<th>that (distal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TURKISH</td>
<td>bu</td>
<td>o[n]</td>
</tr>
<tr>
<td></td>
<td>su</td>
<td></td>
</tr>
<tr>
<td>HEBREW</td>
<td>ze</td>
<td>hu</td>
</tr>
<tr>
<td></td>
<td>zot</td>
<td>hi</td>
</tr>
</tbody>
</table>

Figure 3

The significance of this similarity is not entirely clear. Does a distal demonstrative seem appropriate for third person, because a third person participant is often absent?

It is interesting to note that Modern Hebrew has begun to mark animacy by means of the proximal/distal distinction, so that use of the proximal demonstrative denotes an inanimate subject, while use of the distal indicates an animate (and possibly
human) participant.

Putting aside the differences and regarding the similarities, I would venture to say that a copula is more likely to develop into a pronoun where:

1. Word order is flexible
2. Independent pronouns are scarce
3. The present/aspectless copula is null.

and

4. The path the pronoun will take will be:

from copula to distal demonstrative to pronoun.

§8.2.2 Pronoun to Copula: Chinese, Finnish and Modern Hebrew

In the case of Hebrew and Chinese, we have attestation on both the source and target languages. For Finnish we have only the target language.

The changes from Classical Chinese to Modern display the following patterns:

1. Flexible word order to more rigid,
2. Multiple demonstratives to limited personal pronoun set,
3. Fluid categoriality (noun/verb/preposition) to more fixed patterns,

and

4. Null copula to obligatory overt copula.

The changes from Biblical to Modern Hebrew display the following patterns:

1. Flexible VSO to rigid SVO,
(2) Inflectional patterns to more isolating formations,
(3) Aspect system to tense system,
and
(3) Null aspectless copula to obligatory present tense copula.

In the case of Finnish, our target language displays the following characteristics:

(1) Rigid SVO word order,
(2) Obligatory copula,
and
(3) obligatory third person pronoun.

Generalizing from the above, I would venture to say that an environment where a copula is likely to develop from a pronoun is one in which sentence patterns and word order grow more rigid, where omission of either a copula or a pronoun is no longer countenanced, and where there is a tendency toward greater reliance on syntax to determine case and role.

I would say that a language which is ripe for this sort of reanalysis of a pronoun as copula would be a language which is tending toward isolating patterns. Such a language need not actually be isolating: Modern Finnish has a strong morphological case system and Modern Hebrew still has a robust derivational and inflectional morphology. It is merely that in relation to its antecedents, Modern Hebrew is more isolating. What a Biblical Hebrew sentence would say in one word, a Modern Hebrew sentence requires three words to say:
(1)

(a) **BIBLICAL:**

hikitixa

I beat you.

(b) **MODERN:**

'ani hiketi, 'ani tatzvūx

I beat you.

The same sort of multiplication of verbiage can be found in the difference between Classical and Modern Chinese:

(2) (a) **CLASSICAL:**

er yu Wu wang wo hu?

2sg want Wu King 1sg PRT-Q

Do you want to make me king of Wu Country?

(b) **MODERN:**

ni yao shi wo cheng wei Wu wang ma?

Do you want to make me king of Wu Country?

In the case of the above Chinese example, it is a proliferation of verbal forms that is inflating the size of sentence (2)(b). And while Classical Chinese was an isolating language to begin with, the overall result of the change is that what was previously packaged into a small number of morphemes is being spread out into a much larger one. Each new word makes overt some small portion of the meaning previously
attributed to a single word in the Classical sentence.

One could argue that the Hebrew in example (1) is in the process of de-inflecting, with each of the morphemes on the one word being stripped away and made independent. But in fact, if we do a syllable count, we see that example (1) is longer, not just in words, but also in overall duration. It has more syllables (seven instead of four) and it takes longer to say.

By the same token, the Classical Chinese sentence has six syllables; the Modern has nine. One could argue that Modern Chinese is in the process of creating inflection: but again, the overall tendency is to lengthen the utterance, not to make it more compact.

While we cannot make a similar observation about Finnish, we can compare it to its more conservative relation: Hungarian. Hungarian sentences are more flexible in their word order. Hungarian copulas can be omitted in third person. They are required in Finnish.

Needless to say, after a language allows its sentences to expand, the next step is for them to contract once more. We need not expect Hebrew, Chinese and Finnish sentences to become increasingly longer without end. Practical limitations will prevent that.

It may be that it is during the expansive stage of language change that a copula will be recruited from within the ranks of third person pronouns. I call the expansive stage a ‘tendency’ toward isolation, because it attempts to package more and more semantic elements into separate lexemic compartments.
§8.2.3 Typological Conclusions

From our limited sampling, it would appear that languages develop a pronoun from a copula during their periods of dense packaging, with variable word order, and relatively few syllables per proposition. On the other hand, a copula is developed from a pronoun during periods of sentence expansion, when speakers tend to express more and more semantic elements lexically in overt, separate linguistic packages.

The above conclusion is highly speculative, based on a very small data sampling, and is certainly open to falsification. It is not, in any way, essential to the main point of this dissertation. But it is worth putting forth this hypothesis, in the hopes that future studies will either confirm or refute its claims.

§8.3 Theoretical Conclusions

The above discussion of 'expansive' versus 'compacting' phases of language change opens the way for a theoretical issue that has been not far below the surface of this dissertation from the start. Namely, can something that appears to be 'lexicalization' be termed an instance of 'grammaticalization'?

My response is a resounding 'Yes!'

If we recall that grammaticalization is a process, not a result, we can see that this process "whereby independent linguistic units are recruited into grammatical paradigms or into systems that are [then]... forming" (Chapter 1) may lead to the reinterpretation of a multi-morphemic lexeme as monomorphemic.

Grammaticalization is grammar in the process of forming. And grammar is not limited to morphological affixes.

The same process of reanalysis that leads an independent word to be seen as an
affix can cause a speaker to reinterpret a multimorphemic form of a verb as a deictic, then a demonstrative, then a pronoun, and after that, as an equative linker. With each succeeding stage the item becomes more bleached, reduced and conventional. These are the hallmarks of grammaticalization. What was once a meaningful derived form becomes an unanalyzable unit, whose obligatory use is automated in the grammar.

The tendency toward isolation, like the tendency toward synthesis, is driven by the constant reinterpretation of patterns in language. In both cases, speakers reanalyze what were previously free forms into conventional little packages. In the case of synthesis, syntax is reinterpreted as morphology. In the case of sentence expansion, free lexemes are reinterpreted as grammatical linkers in an ever hardening sentence structure. But while the results may seem like polar opposites, the process is the same: from free to bound, from optional to obligatory, from motivated to conventional.

§8.3.1 Unidirectionality and Hierarchies

All too often, our views of language genesis, development and typology are colored by assumptions that are so deeply ingrained that we do not even realize that they exist. Many naive interpretations of grammaticalization have at their base a hierarchical view of the various language typologies.

Heine et al. (1991.6) quote Von Humboldt with regard to his proposed four stage evolution of language:

*Stage I (which he calls "the lowest stage"): idioms, phrases and clauses;*
Stage II: fixed word order and words vacillating between “matter and form meaning”

Stage III: “analogs of forms,” which are “pure expressions of relations”;

Stage IV (“the highest stage”): “true forms, inflection, and purely grammatical words” (Humboldt 1825:66).

At the base of some views of grammaticalization we can find a similar model, which is both hierarchical and finite. That is, the lowest stage is seen as primitive, while the highest is considered to be both the most refined and also the last and final stage.

Language does not end, unless it dies. So long as a language is alive, there is no final stage. We are all aware that relatively isolating daughter languages can develop from more synthetic or inflectional mother languages. Thus, French is more isolating than Latin and English is more isolating than Proto-Germanic.

Did French and English undergo a process of de-grammaticalization or lexicalization?

We could call it that, but such a label would merely describe the result. The process whereby the grammar of a language becomes more isolating is the same process that leads another language, or even an earlier version of the same language, to become primarily inflectional.

Even as I write this dissertation, my colleagues in grammaticalization are charting the changes taking place in the English language. Joyce Tang Boyland is studying the development of coulda/woulda/shoulda from could have/would have/should have (MS). This is an example of a lexemically complex form being
reinterpreted as opaquely monolexemic. While it loses grammatical structure, it creates a new paradigm.

Would anybody argue that the recruitment of the various forms of to do in English as auxiliary verbs was not an example of grammaticalization?

In (3) (a) there is only one instance of to do, and it is used as a lexical item with a relatively literal meaning. In (3) (c), the literal use is retained in the nonfinite verb, while the finite don't serves as a grammaticalized part of the negation. But in 3 (b), the intermediate stage, do appears twice, once as a finite helping verb and the second time as a nonfinite main verb.

(3)

(a) I do it not.

(b) I do not do it.

(c) I don't do it.

Sentence (3) (b) is a typical instance of the expansive stage of language change. Just as in our Chinese and Hebrew examples, the sentence has gotten longer, both in number of words and number of syllables, without any gain in content. The same unidirectional process that will later cause the negation and the auxiliary to contract into a single unit, brought about the expansion from an optional emphatic to an obligatory auxiliary. We call that process grammaticalization.
§8.3.2 Unidirectionality and Cyclicity

Unidirectionality of change is not inconsistent with cyclicity of result. Linguistic material is limited. We have the same articulatory system, the same processing capacity, storage space, short term memory and the same attention span from one generation to the next.

Language is unstable, so it constantly changes. But even as it changes, its basic mechanisms remain the same: ordering of our articulatory output into phonological, morphological, lexical, syntactic and discourse sized units.

Givon's (1979:209) characterization of the process as one of cyclic waves is apt: Discourse --> Syntax --> Morphology --> Morphophonemics --> Zero.

But languages do not run out of material. Just as one form is on its way out, another is recruited. Because the process is unidirectional, speakers always tend to take whatever independent units are ready at hand and recruit them for more conventional usage. It is precisely because all the independent pronouns have become affixed that a vacuum is created requiring new independent pronouns. It is precisely because the present copula has been recruited as our new independent pronoun, that we suddenly have a null copula. It is precisely because of our sudden need for a new copula that we recruit our third person pronoun as an equative linker. In terms of result, we may have a circle. In terms of changes, it is the same process moving relentlessly forward in the same direction.

§8.4 Conclusion

We have seen that in the languages studied here, a change from a copula to a third person pronoun occurs during a more compacting or synthetic phase of language
development (Turkish, Biblical Hebrew), while a change from third person pronoun to copula (Chinese, Finnish, Modern Hebrew) occurs during a more expansive or isolating period of language development.

Since the tendency toward synthesis and isolation varies during the longterm history of a language, it is possible for a copula to develop into a pronoun and then back to a copula, as is demonstrated in the Hebrew examples. Such a cycle is the direct result of a unidirectional process of grammaticalization over a very long period of time.
Chapter 9
Conclusions

This thesis distinguishes between the process of grammaticalization and its results. Under the process of grammaticalization linguistic units change from:

(1) relatively free to relatively bound,
(2) more motivated to less motivated,
(3) optional to obligatory,
(4) lexical to grammatical,

The process of grammaticalization is the transition from a less 'grammatical' state to one that is more so. But the results of the process, if we follow the history of any given linguistic unit as it undergoes grammaticalization over and over again, may involve a return to a state similar to one that was in effect at an earlier stage of its history.

Unidirectionality of change is not inconsistent with cyclicity of result. If we set out on a trip on this planet and consistently maintain the same heading, we will eventually return to our starting point. If we ask a computer to continuously increase a byte by one, starting with '00', after it reaches 'FF', it will return to '00'. This dissertation does not seek to negate the unidirectionality of change: it merely clarifies what we can or cannot infer from the fact of unidirectionality. We cannot infer, for instance, that because a unit presents as isolating at one stage in the language, it is less grammaticalized than one that presents as affixed: it could have undergone more grammaticalization.

The particular phenomenon explored in this dissertation is the
grammaticalization of pronouns into copulas and copulas into pronouns.

A crosslinguistic examination of copulas in ten languages, Chinese, English, Finnish, French, Hebrew, Hungarian, Korean, Russian, Turkish, and Vietnamese, reveals the following cognitive tendencies:

1. one morpheme is used for possession and existence,

and

2. another morpheme is used for identity and class membership.

Of the ten languages, only English did not fit into the above pattern, using *is* for existence, identity and class membership, but *have* for possession.

Significantly, in all ten of the languages, identity and class membership constructions expressed the two participants (the token and the class, or the two items whose identity is being asserted) as grammatically coequal. That is, in languages with morphological case, they were both in the nominative. In languages without overt case, examination of sentence patterns reveals that neither participant was subordinated to the other.

It is from these equative copulas that third person pronouns are developed, and by the same token it is from subject pronouns that we get newly formed copulas.

In this dissertation the following instances of grammaticalization of pronoun and copula are demonstrated:

1. Chinese (Classical to Modern) pronoun to copula
2. Hebrew (Biblical to Modern) pronoun to copula
3. Finnish (pre-attested to Modern) pronoun to copula
4. Turkish (pre-attested to modern) copula to pronoun
5. Hebrew (pre-proto-Semitic to Biblical) copula to pronoun
The Hebrew example, if we combine (2) and (5) above, provides us with a full cycle, from copula to pronoun to copula.

A pronoun is no less grammatical than a copula, nor is a copula less grammatical than a pronoun. How, then can a transformation from one to the other be deemed an example of grammaticalization?

The key point is that in judging whether something is an example of grammaticalization we must look to the process, not the results. At each point along the continuum of change, the unit in question was becoming more opaque, dependent, bleached and grammatical. In the case of Hebrew, the path followed went like this:

\[
xaja/xawa \rightarrow haja/hawa (\rightarrow) \text{ howo} \rightarrow \text{ hu} \rightarrow \text{ hu}
\]

\[
\begin{array}{cccc}
\text{A} & B & C \\
\text{‘To live’} & \rightarrow & \text{‘to be’ (e.g. ->)} & \text{‘his being’} \rightarrow \text{‘he’} \rightarrow \text{‘be’}
\end{array}
\]

The changes as they are diagramed above are each an example of grammaticalization, because they are examples of the following processes:

A. Semantic and morphological bleaching

B. The change from an independent lexical form with grammatical morphology, to a member of a closed set of grammatical morphemes (deictics)

C. The change from a member of a closed set of grammatical morphemes to membership in a paradigm for the conjugation of a highly bleached, syntactically privileged verb.
At each step of the way, the linguistic unit underwent a further degree of abstraction. In step A, the semantic bleaching changed from a concrete description to an abstract relation, in step B, from a fully parsable derived form to a member of a deictic pronoun set, and in step C from a deictic, referring to an entity, to a member of a verbal paradigm referring to the most bleached relation expressed in the language.

Why is this phenomenon not merely an example of lexicalization? After all, we have a new lexical item created at each step in the progression. The answer is that this is an example of lexicalization, but not merely lexicalization.

Lexicalization is the formation of new lexemes. Grammaticalization is the formation of new grammar. In this case, we have both co-occurring.

It is possible to have new lexemes formed without grammaticalization: e.g. hussy from housewife. Here the only thing being recoded is on the lexico-semantic plane. The grammatical system is unaffected.

It is possible to have grammaticalization without lexicalization: for instance, when English developed an inflexible SVO word order, preverbal position for nouns in assertive clauses became grammaticalized as subject.

It is by no means unusual to have both grammaticalization and lexicalization taking place simultaneously. That is, most grammaticalization involving independent morphology is also lexicalization of said morphemes into new grammatical categories.

This position is consonant with the heuristic given in Chapter 1, supra. It is also in accord with Traugott (1994:1481): “Grammaticalization is the linguistic process whereby grammatical categories ... are organized and coded ... The study of
grammaticalization therefore highlights the tension between relatively unconstrained lexical expression and more constrained morphosyntactic coding, and points to the relative indeterminacy in language and to the basic non-discreteness of categories."

The relationship between lexicalization and grammaticalization is analogous to that between phonetic change and phonemic change. Phonetic change does not affect the contrast of segments in the phonemic system. If one phone is replaced by another in a particular environment, that is mere phonetic change, so long as the integrity of the phoneme of which it is an allophone is not jeopardized. By the same token, if one lexeme is replaced by another, as could have happened if an earlier pronoun in Hebrew had been replaced by hu without change of category, then this is mere lexicalization. But what happened in Hebrew, at each step in the progression, was a drastic change in the grammatical system, involving the disappearance of a verbal element in the clause in favor of a nominal or vice versa. These changes went hand in hand with language wide restructuring of verbal and nominal categories. The grammatical system was reshaped and hence this is an instance of grammaticalization.

Is this a counterexample to unidirectionality? No. The direction of the change was always the same: from the specific to the general, from the highly particularized to the abstract. But the result of the series of changes examined was circular.

From this we can infer that while grammaticalization progresses along a unidirectional cline from concrete to abstract, the history of a particular linguistic unit may reveal that it has travelled the same path more than once.

The implications for grammaticalization theory from the above discussion are as follows:
(a) an independent linguistic unit is not necessarily less grammaticalized than one that is dependent,

and

(b) an isolating language has not necessarily undergone fewer cycles of grammaticalization than one that is synthetic.

The process of grammaticalization is a constant force that drives language onward, and it can account for all change that moves from iconicity to formalism. The study of grammaticalization can encompass anything from nonce formation to paradigm building to syntactic change to typological evolution. The same processes that we can see taking place over a few stretches of discourse can also be traced back as operating over thousands of years. But the linguistic resources upon which the process operates are limited. This calls for a certain amount of recycling. Most cycles are so long and convoluted that returning to the point of departure in any recognizable shape is entirely out of the question. But for those items that are already highly grammaticalized, the radius of the cycle is small, and so the closing of the circle is recognizable, even over a considerable stretch of time.
Appendix -- Index of Cited Authors

Berman 85
Bybee 14, 15, 19, 24, 26, 117, 135
Chen 63, 64, 66, 68, 75, 76
Claudi 27 (footnote)
Clauson 119, 122, 123, 133
Copeland 44
Delancey 22, 23
Devitt 85, 93
Dixon 57
Dobson 64
Erguvanli 123
Fouche 20
Gildea 13, 59
Graham 69, 70, 71
Grevisse 21
Ginsberg 143
Givon 99, 185
Hakulinen 106, 107
Heine 1, 2, 3, 13, 25, 27 (footnote), 182
Hockett 9, 19
Hopper 4
Hunnemeyer 27 (footnote)

Kiss 111
Li 63, 77, 80, 85, 95, 96, 101, 115, 116
Labov 22
Lambdin 141
Langacker 48, 51, 57
Lehtinen 105, 106, 107
Lewis 56, 118, 125, 126, 128, 175, 176
Menges 120, 121
Moscati 57, 168
Thompson 77, 80, 85, 95, 96, 101, 115, 116
Traugott 1, 2, 3, 4, 12, 190
Whitney 109
Yen 63, 77, 79
Bibliography


Devitt, Dan. 1992. 'To be' or no 'to be': The Evolution of Zero From Copulas. MS.


Traugott, Elizabeth Closs. MS. Constructions in Grammaticalization.


